DECLARATION OF JUNE ANN MUNFORD

1. My name is June Ann Munford. I am over the age of 18, have personal knowledge of the facts set forth herein, and am competent to testify to the same.

2. I earned a Master of Library and Information Science (MLIS) from the University of Wisconsin-Milwaukee in 2009. I have over ten years of experience in the library/information science field. Beginning in 2004, I have served in various positions in the public library sector including Assistant Librarian, Youth Services Librarian and Library Director. I have attached my Curriculum Vitae as Appendix CV.

3. During my career in the library profession, I have been responsible for materials acquisition for multiple libraries. In that position, I have cataloged, purchased and processed incoming library works. That includes purchasing materials directly from vendors, recording publishing data from the material in question, creating detailed material records for library catalogs and physically preparing that material for circulation. In addition to my experience in acquisitions, I was also responsible for analyzing large collections of library materials, tailoring library records for optimal catalog

search performance and creating lending agreements between libraries during my time as a Library Director.

4. I am familiar with the Internet Archive, a digital library accessible to the general public. Among other services that the Internet Archive makes available to the general public is the Wayback Machine, an online archive of the web. The Internet Archive's Wayback Machine service archives webpages as of a certain capture date to track changes in the web over time. The Internet Archive has been in operation as a nonprofit library since 1996 and has hosted the Wayback Machine service since its inception in 2001. During my time as a librarian, I frequently used the Internet Archive's Wayback Machine for research and instruction purposes. This includes teaching instructional classes on using the Wayback Machine to library patrons and using the Wayback Machine to research reference inquiries that require hard-to-find online resources. I consider the Internet Archive's recordskeeping to be as rigorous and detailed as other formal library recordskeeping practices such as MARC records, OCLC records and Dublin Core.

5. I have reviewed Exhibit 1006, a document entitled "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in

Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics.

6. Attached hereto as Appendix PHILLIPS01 is a screen capture of the Internet Archive Wayback Machine entry for http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015. I secured these screen captures myself from https://web.archive.org/web/20160803043658/http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015. This record was captured as of August 3, 2016.

7. Appendix PHILLIPS01 is an Internet Archive capture of a web page entitled SOTICS 2015 as hosted on the ThinkMind website. ThinkMind is an open-source repository of academic materials, allowing visitors to the ThinkMind website to download conference proceedings and papers. Appendix PHILLIPS01 contains a download link to "Extracting Social Structure from Darkweb Forums," visible on page 3. As the Internet Archive first captured this page as of August 3, 2016, it is my determination that "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media

Technologies, Communication and Informatics was first made available to the public as of August 3, 2016 via the ThinkMind website.

8. Attached hereto as Appendix PHILLIPS02 is a screen capture of a webpage entitled Extracting Social Structure from Darkweb Forums as hosted on the Oxford University Research Archive website. I secured these screen captures myself from https://ora.ox.ac.uk/objects/uuid:40118c74-bcca-4120-896b-8995b14c2693.

9. Appendix PHILLIPS02 is a web page describing "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics. The 'Deposit Date' field present on page 1 of Appendix PHILLIPS02 indicates this paper was first deposited in the Oxford University Research Archive as of December 21, 2015.

10. Appendix PHILLIPS02 also contains a link to a true and complete copy of "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics. This PDF

copy can be found on pages 3-8 of Appendix PHILLIPS02, downloaded directly from the link. In comparing Exhibit 1006 and Appendix PHILLIPS02, it is my determination that Exhibit 1006 is a true and correct copy of "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics.

11. The 'Deposit date' field present on page 1 of Appendix PHILLIPS02 indicates "Extracting Social Structure from Darkweb Forums" was deposited into the Oxford University Research Archive as of December 21, 2015.

12. Attached hereto as Appendix PHILLIPS03 is a screen capture of the Internet Archive Wayback Machine entry for http://ora.ox.ac.uk:80/information/termsOfUse. I secured these screen captures myself from https://web.archive.org/web/20151121124431/http://ora.ox.ac.uk:80/information/termsOfUse. This record was captured as of November 21, 2015.

13. Appendix PHILLIPS03 is an Internet Archive capture of a web page entitled Terms of Use as hosted on the Oxford University Research Archive website. This webpage, describing the access policies of the archive, includes the

following passage on page 2: ‘Unless otherwise indicated in the relevant permissions notice on the item record, users may download and/or print one copy of any item in ORA…’. In reviewing Appendix PHILLIPS02, the permissions associated with “Extracting Social Structure from Darkweb Forums” indicate that any user of the Oxford University Research Archive would have been able to access a copy of the Phillips paper upon being deposited into the archive. In fact, PHILLIPS02 includes a “Views and Downloads” metrics bar graph tracking the number of views and downloads for the Phillips paper.

14. Attached hereto as Appendix PHILLIPS04 is a screen capture of the Internet Archive Wayback Machine entry for http://openaccess.ox.ac.uk/next-ref/. I secured these screen captures myself from https://web.archive.org/web/20151209231304/http://openaccess.ox.ac.uk/next-ref/. This record was captured as of December 9, 2015.

15. Appendix PHILLIPS04 is an Internet Archive capture of a web page entitled Act of Acceptance: Open Access and the next REF as hosted on the Open Access Oxford website. This web page describes the process of depositing material into the archive for authors. The following passage on page 2 of Appendix PHILLIPS04 describes the internal processes of Oxford archival

staff: ‘The Bodleian team will check copyright and licensing conditions, create an ORA record and make the full-text available…’. Appendix PHILLIPS04 indicates that authors depositing their work into the Oxford University Research Archive should expect the full text of their work to be accessible under an Open Access license. Some other key points of the Act on Acceptance policy require papers to be deposited in an open access repository “within three months of acceptance for publication” such that “readers and search engines must be able to find out about the article straight away.”

16. Attached hereto as Appendix PHILLIPS05 is a screen capture of the Internet Archive Wayback Machine entry for http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/. I secured these screen captures myself from https://web.archive.org/web/20151023045941/http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/. This record was captured as of October 23, 2015.

17. Appendix PHILLIPS05 is an Internet Archive capture of a web page entitled What is Open Access? as hosted on the Open Access Oxford website. This web page describes the mission statement of the Oxford University Research Archive on page 2 of Appendix PHILLIPS05 as follows: ‘The Oxford

Research Archive was established in 2007 as a permanent and secure online archive of research materials produced by members of the University of Oxford. We are enhancing the coverage and features in ORA to provide a single point of public access to electronic copies of peer-reviewed journal articles, conference proceedings by Oxford authors and Oxford research theses…This enables authors to publish in any journal and to self-archive a version of the article for free public use.' This passage indicates that any author submitting material to the Oxford University Research Archive would be doing so knowing that the material would then be made accessible to the public.

18. Appendix PHILLIPS02 indicates "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics was deposited into the Oxford University Research Archive as of December 21, 2015. Appendix PHILLIPS03 indicates any user of the Oxford University Research Archive is able to download full-text copies of works found in the archive as of November 2015, one month before the paper was submitted. Appendix PHILLIPS04 indicates as of December 2015, a paper is indexed and made available to the public soon after being submitted to the Oxford University Research

Archive. Appendix PHILLIPS05 indicates as of October 2015, the Oxford University Research Archive's institutional goal is to make materials included within the archive available to the public. As such, it is my determination that "Extracting Social Structure from Darkweb Forums" by Elizabeth Phillips, et al. as published in Proceedings of the 2015 International Conference on Social Media Technologies, Communication and Informatics was made available to the public via the Oxford University Research Archive soon after December 21, 2015.

19. I have reviewed Exhibit 1013, "Securing Cyberspace: Identifying Key Actors in Hacker Communities" by Victor Benjamin, et al. as published in Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics.

20. Attached hereto as Appendix BENJAMIN01 is a true and correct copy of the MARC record for Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix BENJAMIN01 accurately describes the title, publisher and ISBN number of Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics.

IPR2025-01170
CrowdStrike EX1014 Page 10

21. Attached hereto as Appendix BENJAMIN02 is a true and correct copy of "Securing Cyberspace: Identifying Key Actors in Hacker Communities" as published in Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. This PDF was secured from the library's collection. In comparing Exhibit 1013 to Appendix BENJAMIN02, it is my determination that Exhibit 1013 is a true and correct copy of "Securing Cyberspace: Identifying Key Actors in Hacker Communities" by Victor Benjamin, et al. as published in Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics.

22. The 008 field of the MARC record in Appendix BENJAMIN01 indicates the date of record creation. The 008 field of Appendix BENJAMIN01 indicates the Penn State University Library first acquired Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics as of October 27, 2014. Considering this information, it is my determination that Proceedings of the 2012 IEEE Conference on Intelligence and Security Informatics and therefore "Securing Cyberspace: Identifying Key Actors in Hacker Communities" was made available to the public shortly after October 27, 2014.

IPR2025-01170
CrowdStrike EX1014 Page 11

23. I have reviewed Exhibit 1010, "Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits" by Mehran Bozorgi, et al. as published in Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery.

24. Attached hereto as Appendix BOZORGI01 is a true and correct copy of the MARC record for Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery as held by the Penn State University Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix BOZORGI01 accurately describes the title, publisher and ISBN number of Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery.

25. Attached hereto as Appendix BOZORGI02 is a true and correct copy of "Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits" as published in Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery as held by the Penn State University Library. This PDF was secured from the library's collection. In comparing Exhibit 1010 to Appendix BOZORGI02, it is my determination that Exhibit 1010 is a true and correct copy of "Beyond Heuristics: Learning

IPR2025-01170
CrowdStrike EX1014 Page 12

to Classify Vulnerabilities and Predict Exploits" by Mehran Bozorgi, et al. as published in Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery.

26. The 008 field of the MARC record in Appendix BOZORGI01 indicates the date of record creation. The 008 field of Appendix BOZORGI01 indicates the Penn State University Library first acquired Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery as of February 28, 2014. Considering this information, it is my determination that Proceedings of the 2010 ACM SIGKDD International Conference on Knowledge Discovery and therefore "Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits" was made available to the public shortly after February 28, 2014.

27. I have reviewed Exhibit 1009, "Product Offerings in Malicious Hacker Markets" by Ericsson Marin, et al. as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

28. Attached hereto as Appendix MARIN01 is a true and correct copy of the MARC record for Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. I

IPR2025-01170
CrowdStrike EX1014 Page 13

secured this record myself from the library's public catalog. The MARC record contained within Appendix MARIN01 accurately describes the title, publisher and ISBN number of Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

29. Attached hereto as Appendix MARIN02 is a true and correct copy of "Product Offerings in Malicious Hacker Markets" as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. These scans were secured from the library's collection. In comparing Exhibit 1009 to Appendix MARIN02, it is my determination that Exhibit 1009 is a true and correct copy of "Product Offerings in Malicious Hacker Markets" by Ericsson Marin, et al. as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

30. The 008 field of the MARC record in Appendix MARIN01 indicates the date of record creation. The 008 field of Appendix MARIN01 indicates the Penn State University Library first acquired Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as of November 8, 2016. Considering this information, it is my determination that Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics and

IPR2025-01170
CrowdStrike EX1014 Page 14

therefore “Product Offerings in Malicious Hacker Markets” was made available to the public shortly after November 8, 2016.

31. I have reviewed Exhibit 1007 “Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence” by Eric Nunes, et al. as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

32. Attached hereto as Appendix NUNES01 is a true and correct copy of the MARC record for Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. I secured this record myself from the library’s public catalog. The MARC record contained within Appendix NUNES01 accurately describes the title, publisher and ISBN number of Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

33. Attached hereto as Appendix NUNES02 is a true and correct copy of “Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence” as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as held by the Penn State University Library. These scans were secured from the library’s collection. In

IPR2025-01170
CrowdStrike EX1014 Page 15

comparing Exhibit 1007 to Appendix NUNES02, it is my determination that Exhibit 1007 is a true and correct copy of "Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence" by Eric Nunes, et al. as published in Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics.

34. The 008 field of the MARC record in Appendix NUNES01 indicates the date of record creation. The 008 field of Appendix NUNES01 indicates the Penn State University Library first acquired Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics as of November 8, 2016. Considering this information, it is my determination that Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics and therefore "Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence" was made available to the public shortly after November 8, 2016.

35. I have reviewed Exhibit 1016 "The Risks Posed by Jihadist Hackers" by Chirstopher Heffelfinger as published in CTC Sentinel July 2013.

36. Attached hereto as Appendix HEFFELFINGER01 is a true and correct copy of the MARC record for CTC Sentinel as held by the Penn State University

Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix HEFFELFINGER01 accurately describes the title, publisher and ISSN number of CTC Sentinel. The Showing Results field present on page 3 of this record indicates the library's collection includes the July 2013 issue of CTC Sentinel containing "The Risks Posed by Jihadist Hackers."

37. Attached hereto as Appendix HEFFELFINGER02 is a true and correct copy of "The Risks Posed by Jihadist Hackers" as published in CTC Sentinel as held by the Penn State University Library. These scans were secured from the library's collection. In comparing Exhibit 1016 to Appendix HEFFELFINGER02, it is my determination that Exhibit 1016 is a true and correct copy of "The Risks Posed by Jihadist Hackers" by Chirstopher Heffelfinger as published in CTC Sentinel.

38. The 008 field of the MARC record in Appendix HEFFELFINGER01 indicates the date of record creation. The 008 field of Appendix HEFFELFINGER01 indicates the Penn State University Library first acquired CTC Sentinel as of March 4, 2010 and held the journal in perpetuity until the publication of "The Risks Posed by Jihadist Hackers." Considering this information, it is my determination that CTC Sentinel July

2013 and therefore "The Risks Posed by Jihadist Hackers" was made available to the public shortly after its initial publication in July 2013.

39. I have reviewed Exhibit 1015 "Identifying Digital Threats in a Hacker Web Forum" by Mitch Macdonald, et al. as published in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

40. Attached hereto as Appendix MACDONALD01 is a true and correct copy of the MARC record for Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining as held by the Penn State University Library. I secured this record myself from the library's public catalog. The MARC record contained within Appendix MACDONALD01 accurately describes the title, publisher and ISBN number of Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

41. Attached hereto as Appendix MACDONALD02 is a true and correct copy of "Identifying Digital Threats in a Hacker Web Forum" as published in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining as held by the Penn State

IPR2025-01170
CrowdStrike EX1014 Page 18

University Library. These scans were secured from the library's collection. In comparing Exhibit 1015 to Appendix MACDONALD02, it is my determination that Exhibit 1015 is a true and correct copy of "Identifying Digital Threats in a Hacker Web Forum" by Mitch Macdonald, et al. as published in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

42. The 008 field of the MARC record in Appendix MACDONALD01 indicates the date of record creation. The 008 field of Appendix MACDONALD01 indicates the Penn State University Library first acquired Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining as of January 25, 2016. Considering this information, it is my determination that Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining and therefore "Identifying Digital Threats in a Hacker Web Forum" was made available to the public shortly after January 25, 2016.

43. I have reviewed Exhibit 1025 a paper entitled "Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums" by Mohammed Almukaynizi, et al. as published in the Proceedings of the 2017

IPR2025-01170
CrowdStrike EX1014 Page 19

International Conference of The Computational Social Science Society of the Americas.

44. Attached hereto as Appendix ALMUKAYNIZI01 is a screen capture of a webpage entitled Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums as hosted on the ACM Digital Library website. I secured these screen captures myself from https://dl.acm.org/doi/10.1145/3145574.3145590.

45. Appendix ALMUKAYNIZI01 is a webpage containing a descriptive record of the Almukaynizi paper. Pages 5 – 11 of Appendix ALMUKAYNIZI01 contain a true and correct copy of the Almukaynizi paper I acquired from the ACM Digital Library. In comparing Appendix ALMUKAYNIZI01 to Exhibit 1025 it is my determination that Exhibit 1025 is a true and correct copy of "Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums" by Mohammed Almukaynizi, et al. as published in the Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas.

IPR2025-01170
CrowdStrike EX1014 Page 20

46. I have reviewed Exhibit 1017 a thesis entitled Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach by Vivian Fiona Guetler.

47. Attached hereto as Appendix GUETLER01 is a screen capture of the Internet Archive Wayback Machine entry for https://researchrepository.wvu.edu/etd/11253/. I secured these screen captures myself from https://web.archive.org/web/20220701000000*/https://researchrepository.wvu.edu/etd/11253/. This record was captured as of October 26, 2022.

48. Appendix GUETLER01 is a webpage containing a descriptive record of the Guetler thesis. Pages 4 – 156 of Appendix GUETLER01 contain a true and correct copy of the Guetler thesis I acquired from the Internet Archive by clicking the 'Download' button visible on page 2 of Appendix GUETLER01. In comparing Appendix GUETLER01 to Exhibit 1017 it is my determination that Exhibit 1017 is a true and correct copy of Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach by Vivian Fiona Guetler.

49. I have reviewed Exhibit 1004 a web page entitled "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits" by Carl Sabottke, et al. as published in Proceedings of the 24th USENIX Security Symposium.

50. Attached hereto as Appendix SABOTTKE01 is a screen capture of the Internet Archive Wayback Machine entry for https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke. I secured these screen captures myself from https://web.archive.org/web/20150923100812/https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke. This record was captured as of September 23, 2015.

51. Appendix SABOTTKE01 is a webpage containing a descriptive record of the Sabottke paper. Pages 4 – 22 of Appendix SABOTTKE01 contain a true and correct copy of the Sabottke paper I acquired from the Internet Archive by clicking the 'Sabottke PDF' button visible on page 2 of Appendix SABOTTKE01. In comparing Appendix SABOTTKE01 to Exhibit 1004 it is my determination that Exhibit 1004 is a true and correct copy of "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits" by Carl Sabottke, et al. as published in

IPR2025-01170
CrowdStrike EX1014 Page 22

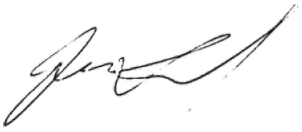
Proceedings of the 24th USENIX Security Symposium. As pages 1 – 4 of Appendix SABOTTKE01 contains a download link to the Sabottke paper dated September 23, 2015 and Pages 5 – 22 of Appendix SABOTTKE01 contains a true and correct copy of the Sabottke paper itself for download as of August 12, 2017, it is my determination that "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits" by Carl Sabottke, et al. as published in Proceedings of the 24th USENIX Security Symposium was likely first available to the public via the USENIX website as of September 23, 2015 and certainly no later August 12, 2017 via the Internet Archive.

52. I have been retained on behalf of the Petitioner to provide assistance in the above-illustrated matter in establishing the authenticity and public availability of the documents discussed in this declaration. I am being compensated for my services in this matter at the rate of $200.00 per hour plus reasonable expenses. My statements are objective, and my compensation does not depend on the outcome of this matter.

53. I declare under penalty of perjury that the foregoing is true and correct. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to

be true; and further that these statements were made the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 7/18/2025



June Ann Munford

IPR2025-01170
CrowdStrike EX1014 Page 24

June A. Munford
Curriculum Vitae

Education

University of Wisconsin-Milwaukee - MS, Library & Information Science, 2009
Milwaukee, WI

- Coursework included cataloging, metadata, data analysis, library systems, management strategies and collection development.
- Specialized in library advocacy, cataloging and public administration.

Grand Valley State University - BA, English Language & Literature, 2008
Allendale, MI

- Coursework included linguistics, documentation and literary analysis.
- Minor in political science with a focus in local-level economics and government.

Professional Experience

Researcher / Expert Witness, October 2017 – present
Freelance ● Pittsburgh, Pennsylvania & Grand Rapids, Michigan

- Material authentication and public accessibility determination. Declarations of authenticity and/or public accessibility provided upon research completion. Experienced with appeals and deposition process.

- Research provided on topics of public library operations, material publication history, digital database services and legacy web resources.

- Past clients include Alston & Bird, Arnold & Porter, Baker Botts, Fish & Richardson, Erise IP, Irell & Manella, O'Melveny & Myers, Perkins-Coie, Pillsbury Winthrop Shaw Pittman and Slayden Grubert Beard.

Library Director, February 2013 - March 2015
Dowagiac District Library ● Dowagiac, Michigan

- Executive administrator of the Dowagiac District Library. Located in Southwest Michigan, this library has a service area of 13,000, an annual operating budget of over $400,000 and total assets of approximately $1,300,000.

● Developed careful budgeting guidelines to produce a 15% surplus during the 2013-2014 & 2014-2015 fiscal years while being audited.

● Using this budget surplus, oversaw significant library investments including the purchase of property for a future building site, demolition of existing buildings and building renovation projects on the current facility.

● Led the organization and digitization of the library's archival records.

● Served as the public representative for the library, developing business relationships with local school, museum and tribal government entities.

● Developed an objective-based analysis system for measuring library services - including a full collection analysis of the library's 50,000+ circulating items and their records.

November 2010 - January 2013
Librarian & Branch Manager, Anchorage Public Library ● Anchorage, Alaska

● Headed the 2013 Anchorage Reads community reading campaign including event planning, staging public performances and creating marketing materials for mass distribution.

● Co-led the social media department of the library's marketing team, drafting social media guidelines, creating original content and instituting long-term planning via content calendars.

● Developed business relationships with The Boys & Girls Club, Anchorage School District and the US Army to establish summer reading programs for children.

June 2004 - September 2005, September 2006 - October 2013
Library Assistant, Hart Area Public Library
Hart, MI

● Responsible for verifying imported MARC records and original MARC cataloging for the local-level collection as well as the Michigan Electronic Library.

● Handled OCLC Worldcat interlibrary loan requests & fulfillment via ongoing communication with lending libraries.

Professional Involvement

Alaska Library Association - Anchorage Chapter

● Treasurer, 2012

Library Of Michigan
- ● Level VII Certification, 2008
- ● Level II Certification, 2013

Michigan Library Association Annual Conference 2014
- ● New Directors Conference Panel Member

Southwest Michigan Library Cooperative
- ● Represented the Dowagiac District Library, 2013-2015

Professional Development

Library Of Michigan Beginning Workshop, May 2008
Petoskey, MI
- ● Received training in cataloging, local history, collection management, children's literacy and reference service.

Public Library Association Intensive Library Management Training, October 2011
Nashville, TN
- ● Attended a five-day workshop focused on strategic planning, staff management, statistical analysis, collections and cataloging theory.

Alaska Library Association Annual Conference 2012 - Fairbanks, February 2012
Fairbanks, AK
- ● Attended seminars on EBSCO advanced search methods, budgeting, cataloging, database usage and marketing.

Depositions

2019 ● Fish & Richardson

Apple v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282 and 39521-00421IP2)

2019 ● Erise IP

Implicit, LLC v. Netscout Systems, Inc (Case No. 2:18-cv-53-JRG)

2019 ● Perkins-Coie

Adobe Inc. v. RAH Color Technologies LLC (Case No. IPR2019-00627, IPR2019-00628, IPR2019-00629 and IPR2019-00646)

2020 ● O'Melveny & Myers

Maxell, Ltd. v. Apple Inc. (Case No. 5:19-cv-00036-RWS)

2021 ● Pillsbury Winthrop Shaw Pittman LLP

Intel v. SRC (Case No. IPR2020-1449)

2022 ● Perkins-Coie

Realtek v. Future Link (Case No. IPR2021-01182)

2023 ● Fish & Richardson

Neuroderm Ltd. v. Abbvie, Inc (Case No. PGR2022-00040)

2023 ● Fish & Richardson

Nearmap US Inc. v. Pictometry International Corp. (Case No. IPR2022-00735)

2023 ● Fish & Richardson

Samsung Electronics v. MemoryWeb LLC (Case No. 39843-0136PS1)

2023 ● Pillsbury Winthrop Shaw Whitman LLP

Gravel Rating Systems v. Costco Wholesale Corp. (Case No. 4:21-cv-149-ALM)

2024 ● Willkie-Farr

Netflix, Inc. v. VideoLabs. Inc. (Case No. IPR2023-00628)

2024 ● Quinn Emanuel Urquhart & Sullivan

Purdue University v. Wolfspeed (Case No. 1:21-CV-840)

2024 ● Fish & Richardson

Dish Network v. Entropic Communications (Case No. IPR2024-00393)

2025 ● Perkins Coie

Amazon.com, Inc. v. Nokia Technologies Oy (Case No. IPR2024-00572)

2025 ● Fish & Richardson

Dish Network v. Entropic Communications (Case No. IPR2024-00373)

Limited Case and Clientele History

Alston & Bird

- Ericsson

  v. Collision Communications (Case No. IPR2022-01233)

- Nokia

  v. Neptune Subsea, Xtera (Case No. 1:17-cv-01876)

- Universal Electronics Inc

  v. Roku Inc (Case No. IPR2022-00818)

Alavi & Anaipakos PLLC

- Stingray Group Inc. and Stingray Music USA, Inc.

  v. Edwin A. Hernandez-Mondragon and Egla Corp. (Case No. 1:24-cv-21226-RAR)

Arnold & Porter

- Ivantis

  v. Glaukos (Case No. 8:18-cv-00620)

- Samsung

  v. Jawbone (Case No. 2:21-cv-00186)

Benesch Friedlander Coplan & Aronoff

- Voyis

  v. Cathx (Case No. 5:21-cv-00077-RWS)

Buchanan, Ingersoll & Rooney PC

- Google LLC, AT&T Service Inc., T Mobile USA Inc., Cellco Partnership, Ericsson Inc. & Nokia of America Corp.

  v. KT Corp & Pegasus Wireless Innovation LLC (Case No. IPR2025-00293)

Deschert LLP

- Smaxtec Animal Care

  v. ST Reproductive Technologies, LLC (Case No. IPR2024-00885)

Duane Morris

- Kangxi Communication Technologies

  v. Skyworks Solutions Inc. (Case No. IPR2025-00373)

- Microsoft

  v. Edge Networking Systems LLC (Case No. 1:24-cv-00215)

Erise I.P.

- Apple

  v. Ericsson Inc. (Case No. IPR2022-00715)

  v. Future Link Systems (Case No. IPRs 6317804, 6622108, 6807505, and 7917680)

  v. INVT (Case No. 20-1881)

  v. Navblazer LLC (Case No. IPR2020-01253)

  v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)

v. Quest Nettech Corp (Case No. 2:19-cv-00118-JRG)

v. Telefonaktiebolaget LM Ericsson (Case No. IPR2022-00275)

v. Theta IP, LLC (Case No. IPR2024-00818)

v. THL Holding Company LLC (Case No. 1:23-cv-00548)

- Fanduel

v. CGT (Case No. 19-1393)

- Garmin

v. Phillips North America LLC (Case No. 2:19-cv-6301-AB-KS)

- Netscout

v. Longhorn HD LLC (Case No. 2:20-cv-00349)

v. Implicit, LLC (Case No. 2:18-cv-53-JRG)

- Sony Interactive Entertainment LLC

v. Bot M8 LLC (Case No. IPR2020-01288)

v. Infernal Technology LLC (Case No. 2:19-CV-00248-JRG)

- Tesla

v. Charge Fusion Technologies LLC (Case No. IPR2025-00032)

- Unified Patents

v. GE Video Compression (Case No. 2:19-cv-248)

Fish & Richardson

- Apple

v. AliveCor (Case No. 3:21-cv-03958)

v. LBS Innovations (Case No. 2:19-cv-00119-JRG-RSP)

v. Koss Corporation (Case No. IPR2021-00305)

v. Masimo (Case No. IPR 50095-0012IP1, 50095-0012IP2, 50095-0013IP1, 50095-0013IP2, 50095-0006IP1, 50095-0135IP1)

v. Neonode (Case No. 21-cv-08872-EMC)

v. Qualcomm (Case No. IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)

v. Resonant Sys. (Case No. 7:23-cv-00077-ADA)

- AsusTek Computer Inc.
  v. Videolabs, Inc. (Case No. 22-CV-00720-ADA)

- Dell
  v. Neo Wireless (Case No. IPR2022-00616)

- Dish Network

  v. Entropic Communications, LLC (Case No. 2:2023-CV-01043)

  v. Entropic Communications LLC (Case No. IPR2024-00393)

  v. Realtime Adaptive Streaming (Case No. 1:17-CV-02097-RBJ)

  v. TQ Delta LLC (Case No. 18-1798)

- Evapco Dry Cooling

  v. SPG Dry Cooling (Case No. IPR2021-00688)

- Genetec

  v. Sensormatic Electronics (Case No. 1:20-CV-00760)

- Huawei

  v. Bell Northern Research LLC (Case No. IPR2019-01174)

- Kianxis

  v. Blue Yonder (Case No. 3:20-cv-03636)

- LG Electronics

  v. Bell Northern Research LLC (Case No. 3:18-cv-2864-CAB-BLM)

- MediaTek

  v. MOSAID (Case No. IPR2024-00718)

- Metaswitch

  v. Sonus Networks (Case No. IPR2018-01719)

- Microsoft

  v. Throughputer Inc (Case No. IPR2022-00757)

- Mom Enterprises

  v. Ddrops Company (Case No. 1:22-cv-00332)

- MLC Intellectual Property

  v. MicronTech (Case No. 3:14-cv-03657-SI)

- Nearmap Inc

  v. EagleView Technologies (Case No. IPR2022-01009)

- Neuroderm Ltd.

  v. Abbvie, Inc (Case No. PGR2022-00040)

- Posco Co Ltd.

  v. Pascal Drillet et al. (Case No. IPR2025-00370)

- Realtek Semiconductor

  v. Future Link (Case No. IPR2021-01182)

- Quectel
    - v. Koninklijke Philips (Case No. 1:20-cv-01710)

- Samsung
    - v. Aire Technology (Case No. IPR2022-00877)
    - v. Bell Northern Research (Case No. 2:19-cv-00286-JRG)
    - v. Communication Technologies Inc (Case No. IPR2022-01221)
    - v. Jawbone Innovations (Case No. IPR2022-00865)
    - v. MemoryWeb LLC (Case No. IPR2022-00885)
    - v. Dodots Licensing Solutions (Case No. IPR2023-00701)
    - v. Telefonaktiebolaget LM Ericsson (Case No. IPR2021-00615)

- Texas Instruments
    - v. Vantage Micro LLC (Case No. IPR2020-01261)

- Xilinx
    - v. Sentient Sensors LCC (Case No. 1:22-cv-00173)

Goodwin Proctor

- Intel
    - v. Proxense, LLC (Case No. 6:24-cv-00283)

- Samsung Electronics
    - v. Mobile Data Technologies (Case No. 2:24-cv-00435)

- Taiwan Semiconductor Manufacturing Company Ltd.
    - v. Advanced Integrated Circuit Process LLC (Case No. 2:24-cv-00623)

Groombridge Wu

- Nearmap

  v. Eagle View Technologies Inc. and Pictometry International Corp. (Case No. 2:21-cv-00283-TS-DAO)

Irell & Manella

- Curium

Latham & Watkins LLP

- Apple

  v. Advanced Coding Research LLC

Leydig, Voit & Mayer, Ltd.

- Public Availability Consultancy

O'Melveny & Myers

- Apple

  v. Maxell (Case No. 5:19-cv-00036-RWS)

- Disney Media & Entertainment Distribution LLC

  v. Digital Media Technology Holdings, LLC (Case No. 2:22-cv-01642)

- Micron Technology Inc.

  v. Besang Inc (Case No. IPR2023-00900)

- Samsung

  v. Daedalus Prime LLC (1335 Investigation)

Perkins-Coie

- Amazon
  - v. Nokia (Case No. IPR2024-00847 and IPR2024-00848)
  - v. Nokia Technologies Oy (Case No. IPR2024-00572)
- Heru Industries
  - v. The UAB Research Foundation (Case No. IPR2022-01148)
- Intel Corporation
  - v. BeSang Inc. (Case No. IPR2023-00991)
  - v. BitMICRO (Case No. 5:23-cv-00625)
  - v. Telefonaktiebolaget L M Ericsson (Case No. IPR2024-00610)
- Realtek Semiconductor
  - v. Future Link (Case No. IPR2021-01182)
- r-pac
  - v. Adasa, Inc. (Case No. 1:2024cv06102)
- Twitter Inc
  - v. VOIP-Pal.com (Case No. 3:20-cv-02397-JD)
- TCL Industries
  - v. Koninklijke Philips NV (Case No. IPR2021-00495, IPR2021-00496 and IPR2021-00497)
- VusionGroup
  - v. Hanshow (Case No. IPR2024-00857)

Pillsbury Winthrop Shaw Pittman

- Intel

  v. FG SRC LLC (Case No. 6:20-cv-00315)

- Gravel Rating Systems

  v. Costco (Case No. 4:21-cv-149)
  v. Lowe's Home Centers (Case No. 4:21-cv-150)
  v. T-Mobile USA (Case No. 4:21-cv-152)
  v. Kohl's Inc. (Case No. 4:21-cv-258)
  v. Under Armor (Case No. 4:21-cv-356)

Quinn Emanuel

- Exact Sciences Corporation

  v. Geneoscopy, Inc. (Case No. IPR2024-00459 and IPR2024-1330)

- ServiceNow, Inc.

  v. InQuisient, Inc. (Case No. 22-900-CJB)

- Wolfspeed

  v. The Trustees of Purdue University (Case No. 1:2021-cv-00840)

Sheppard, Mullin, Richter & Hampton LLP

- Advanced Micro Devices, Inc and Pensando Systems, Inc.

  v. Concurrent Ventures LLC and Xtreamedge, Inc. (Case No. IPR2025-00478)

- Cadence Design Systems Inc.

  v. Semiconductor Design Technologies (Case No. 3:23-cv-01001)

Shook, Hardy & Bacon LLP

- HP Inc.

  v. Universal Connectivity Technologies (Case No. 1:23-cv-01177)

Stern, Kessler, Goldstein & Fox

- Preliminary Research

Wilmer, Cutler, Pickering, Hale and Dorr

- Apple

  v. Shunock (Case No. 1:23-cv-08598)

- Quest Diagnostics Inc., Haystack Oncology Inc. and Haystack Oncology GmbH

  v. Natera, Inc. (Case No. 1:23-cv-08598)

- Roche Diabetes, Inc.

  v. Trividia Health, Inc. (Case No. 1:24-cv-00668)

Willkie, Farr & Gallagher LLP

- Lenovo, Dell, HP

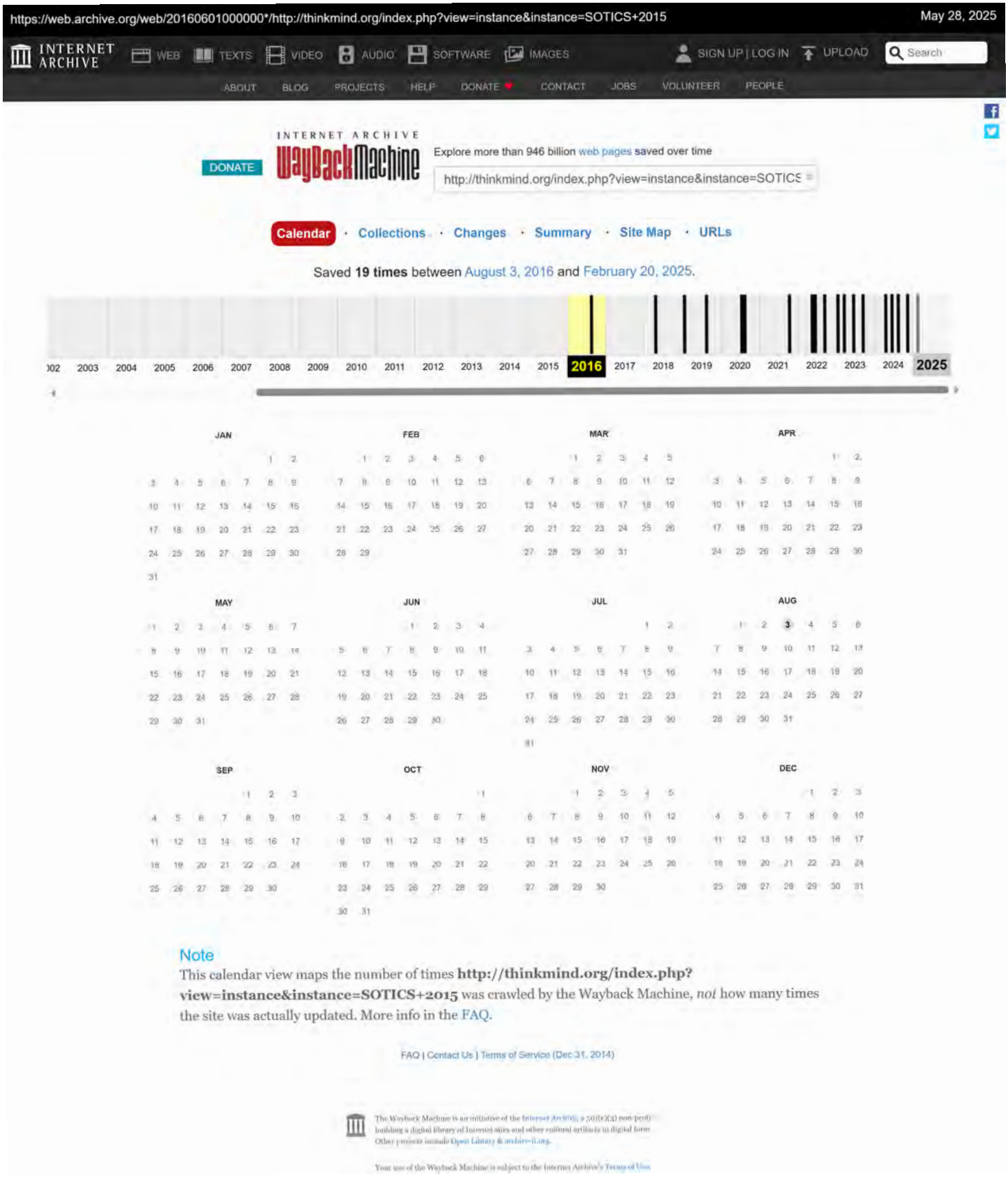
  v. Universal Connectivity Technologies Inc. (Case No. 2:2023cv00449)

- Neurent

  v. The Foundry, LLC (Case No. IPR2024-00669)

- Netflix, Inc.

  v. VideoLabs, Inc. (Case No. IPR2023-00628)


https://web.archive.org/web/20160601000000*/http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015
May 28, 2025
INTERNET ARCHIVE
WEB TEXTS VIDEO AUDIO SOFTWARE IMAGES
SIGN UP | LOG IN UPLOAD
Search
ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE
DONATE
INTERNET ARCHIVE
WayBackMachine
Explore more than 946 billion web pages saved over time
http://thinkmind.org/index.php?view=instance&instance=SOTICS
Calendar · Collections · Changes · Summary · Site Map · URLs
Saved 19 times between August 3, 2016 and February 20, 2025.
002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Note
This calendar view maps the number of times http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015 was crawled by the Wayback Machine, not how many times the site was actually updated. More info in the FAQ.
FAQ | Contact Us | Terms of Service (Dec 31, 2014)
The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.
Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use

https://web.archive.org/web/20160803043658/http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015 May 28, 2025

INTERNET ARCHIVE WayBackMachine http://thinkmind.org/index.php?view=instance&instance=SOTICS+2015 Go JUL AUG APR 03 2015 2016 2018 About this capture
19 captures
3 Aug 2016 - 20 Feb 2025

ThinkMind // SOTICS 2015,The Fifth International Conference on Social Media Technologies, Communication, and Informatics

Type: conference

Download: SOTICS 2015 cover page

Download: SOTICS 2015 foreword page

Download: SOTICS 2015 committee page

Download: SOTICS 2015 table of contents

Download: SOTICS 2015 full proceedings

Notes: IARIA Conference.

Editors:
Carla Merkle Westphall University of Santa Catarina, Brazil
Dumitru Roman, SINTEF, Norway

Event editions: - there are 5 editions.

ISSN: 2326-9294

ISBN: 978-1-61208-443-5

Location: Barcelona, Spain

Dates: from November 15, 2015 to November 20, 2015

Articles: there are 17 articles

Smart Navigation: Using Artificial Intelligent Heuristics in Navigating Multiple Destinations
Hatem Halaoui
keywords: Smart Navigation, Artificial Intelligence, Heuristics, GIS.

What Do People Affected by Cancer Talk about Online? Text analysis of online cancer community usage in Bosnia and Herzegovina
Suncica Hadzidedic Bazdarevic and Alexandra I. Cristea
keywords: online health communities; text analysis; cancer; online health information; Bosnia-Herzegovina

Towards Social Media Platform Integration with an Applied Gaming Ecosystem
Munir Salman, Kam Star, Alexander Nussbaumer, Michael Fuchs, Holger Brocks, Duc Binh Vu, Dominic Heutelbeck, and Matthias Hemmje
keywords: Social Media; Applied Gaming; Digital Ecosystem; Access and Information Integration; Know-how transfer; Social Networking Site

A Real-Time Disaster-Related Information Sharing System Based on the Utilization of Twitter
Osamu Uchida, Masafumi Kosugi, Gaku Endo, Takamitsu Funayama, Keisuke Utsu, Sachi Tajima, Makoto Tomita, Yoshitaka Kajita, and Yoshiro Yamamoto
keywords: Twitter, tweet mapping; disaster-related information; disaster mitigation

Making Sense of Large-Group Discussions using Rhetorically Structured Text
Ana Cristina Bicharra Garcia and Mark Klein
keywords: Rhetorical structure theory; RST; online group deliberation; web forums; argument maps; crowd-computing; social computing

Virtual Activity Ontology Modeling: The Case of Sociocultural Knowledge Sharing
Kaladzavi Guidedi, Diallo Papa Fary, Kolyang Dina Taiwe, and Lo Moussa
keywords: sociocultural knowledge; sharing; Human Activity Theory

Influence of Couple Patterns on Entropy in Multilayer Networks
Xiaoling Nian

Influence of Couple Patterns on Entropy in Multilayer Networks
Xiaoling Nian
keywords: multilayer network; couple pattern; influence; entropy; energy

Social Network Analysis Tools for Career Advancement
Timothy Arndt
keywords: social networks; software tools; social simulation; legal aspect

Mapping Competences of Distance Education Students
Ketia Kellen Araújo da Silva and Patricia Alejandra Behar
keywords: Distance Education; Competences; Distance Education students

Open Data from Social Media as Tool for Better Understanding Complex Territory. Application Through Photos Data in Calabria
Alexandra Middel Silva, Rafael Soares Gonçalves and Carla Cristina Bitencourt Siqueira
keywords: Big Open Data; collective sensing; community; territory.

The Influence of Social TV Strategies and Contents on TV Online Engagement
Angela Fortunato, Michele Gorgoglione, and Umberto Panniello
keywords: Social TV; social networks; viewers' online engagement.

Rumor Detection and Classification for Twitter Data
Sardar Hamidian and Mona Diab
keywords: Rumor Detection and Classification, Supervised Machine Learning, Feature-based model

Helping Hands of Autism Blogger Community in Social Media Platforms
Amit Saha and Nitin Agarwal
keywords: Autism; ASD; Social Support; Twitter; Blogger; Community

Sharing Our Heritage to Shape Our Future
Andrea Nanetti and Anna Simpson
keywords: multi-user sharing platforms; collaborative visioning for the future; heritage science

The Effects of Travel Information Sources on Traveller's Resonance and The Travel Destination Decision-Making Process
Masayuki Kaneko, Lourdes Villaverde, Katsuko Nakahira, and Makiko Okamoto
keywords: Travel-destination information; empathy; motivation; decision-making

Extracting Social Structure from Dark Web Metadata
Elizabeth Phillips, Jason R. C. Nurse, Michael Goldsmith, and Sadie Creese
keywords: Social Network Analysis; Dark Web Forum; Jihadist forums; Social Structure

Towards an Index of Mental Wellbeing in Language - The relationship between time orientation, self-focus and mood during prolonged bed-rest.
Sonja Schmer-Galunder, Peggy Wu, Tammy Ott, and Chris Miller
keywords: Latent Semantic Analysis; Sociolinguistics; Wellness; Psychosocial State Detection; Sentiment Analysis; Data Mining

https://ora.ox.ac.uk/objects/uuid:40118c74-bcca-4120-896b-8995b14c2693 May 28, 2025

ACTIONS
ACCESS DOCUMENT
AUTHORS
FUNDING
BIBLIOGRAPHIC DETAILS
ITEM DESCRIPTION
TERMS OF USE
METRICS

Export

Conference item

Extracting social structure from darkweb forums

Abstract: This paper explores various Social Network Analysis (SNA) techniques in order to identify a range of potentially 'important' members of Islamic Networks within Dark Web Forums. For this experiment, we conducted our investigation on five forums collected in previous work as part of the DarkWeb Forum portal and built upon the tool support created in our previous research in order to visualise and analyse the network. Whilst existing work attempts to identify these structures through state-of-the-art Computational Linguistic techniques, our work relies on the communication metadata alone. Our analysis involved first calculating a range of SNA metrics to better understand the group members, and then apply unsupervised learning in order to create clusters that would help classify the Dark Web Forums users into hierarchical clusters. In order to create our social networks, we investigated the effect of repeated author resolution and various weighting schemes on the ranking of forum members by creating four social networks per forum and evaluating the correlation of the top n users (for n = 10; 20; 30; 40; 50 and 100). Our results identified that varying the weighting schemes created more consistent ranking schemes than varying the repeated author resolution.

Collapse abstract

Publication status: Published
Peer review status: Peer reviewed

Actions

Email Cite Tweet Print

Access Document

Files: Extracting social structure from darkweb forums.pdf (Preview, **Version of record, pdf, 1.0MB**, Terms of use)

Authors

Phillips, E — More by this author
Nurse, J — More by this author
Goldsmith, M — More by this author
Creese, S — More by this author

Funding

Engineering and Physical Sciences Research Council — More from this funder

Bibliographic Details

Publisher: IARIA
Host title: 5th International Conference on Social Media Technologies, Communication, and Informatics (SOTICS '15)
Pages: 97-102
Publication date: 2015-01-01
ISSN: 2326-9294
ISBN: 9781612084435

Item Description

Keywords: social structure, dark web forum, jihadist forums, social network analysis
Pubs id: pubs:580315
UUID: uuid:40118c74-bcca-4120-896b-8995b14c2693
Local pid: pubs:580315
Source identifiers: 580315
Deposit date: 2015-12-21

Terms of use

Copyright holder: IARIA
Copyright date: 2015
Notes: Copyright © IARIA, 2015.

Licence: Terms and Conditions of Use for Oxford University Research Archive

Metrics

Views and Downloads



About views and downloads

If you are the owner of this record, you can report an update to it here: Report update to this record

TO TOP

© Copyright - Bodleian Libraries 2025
Disclaimer | Privacy Policy | Policies | Cookies | Accessibility | Copyright | API | Contact

BODLEIAN LIBRARIES

Extracting Social Structure from DarkWeb Forums

Elizabeth Phillips, Jason R.C. Nurse, Michael Goldsmith, Sadie Creese
Cyber Security Centre,
Department of Computer Science,
University of Oxford, UK
Email: $\{firstname.lastname\}@cybersecurity.ox.ac.uk$

***Abstract*—This paper explores various Social Network Analysis (SNA) techniques in order to identify a range of potentially *'important'* members of Islamic Networks within Dark Web Forums. For this experiment, we conducted our investigation on five forums collected in previous work as part of the Dark Web Forum portal and built upon the tool support created in our previous research in order to visualise and analyse the network. Whilst existing work attempts to identify these structures through state-of-the-art Computational Linguistic techniques, our work relies on the communication metadata alone. Our analysis involved first calculating a range of SNA metrics to better understand the group members, and then apply unsupervised learning in order to create clusters that would help classify the Dark Web Forums users into hierarchical clusters. In order to create our social networks, we investigated the effect of repeated author resolution and various weighting schemes on the ranking of forum members by creating four social networks per forum and evaluating the correlation of the top n users (for $n = 10, 20, 30, 40, 50$ and 100). Our results identified that varying the weighting schemes created more consistent ranking schemes than varying the repeated author resolution.**

Keywords–Social Network Analysis; Dark Web Forum; Jihadist forums; Social Structure

I. INTRODUCTION

With the dawn of the digital revolution, the Internet has transformed the way in which people communicate with each other. One area in which this has been seen is within terrorist networks[1], [2]. The near-instantaneous responses from users now achievable with these developing technologies and the increased audience has meant that previous face-to-face interactions and radical discussions have migrated to online mediums [3], [4]. This migration has led to an increase in the popularity of Dark Web Forums as a means of sharing text based content as well as links to other sites, videos and rich Web 2.0 features [5].

Even before the increased adoption of dark web forums, researchers have used social network analysis to identify key individuals within these groups. Numerous researchers have investigated various techniques in order to retrospectively understand the inner-workings of the 19 terrorists involved in the September 11th attacks in 2001[6]. Early research involved sourcing news articles related to the known suspects after the event and analysing the structure of the group after the event and with known targets [7]. Others have focused on researching how different terrorists organisations work together [8] by performing social network analysis [9].

The widespread adoption of the Dark Web Forums have enabled researchers to analyse the rich source of data [10] and has enabled researchers to establish a greater insight into the inner workings of some of these groups. Whilst existing research has been undertaken into creating a social network topology for Dark web Forums [11], these existing research focus on small communities and networks and rely on sophisticated sentiment analysis techniques, which are difficult to scale when evaluating millions of messages[12].

Since the revelations of metadata collection exposed by Edward Snowden in 2013 [13] , the importance of metadata from emails has gained awareness. In light of these revelations, many individuals have been unsure of their own risk exposure using these metadata techniques alone. Existing work has been effective at establishing hierarchy from the metadata of email communications alone [14], [15]. Such metadata include the post's author and timestamp. In this paper, we set out to investigate whether similar techniques could be applied within Dark Web Forums and assess to what extent we can identify the social structure of the networks.

Our paper is focused on the research question "Using communication metadata alone, can we make reasonable inferences about the structure of terrorist groups". In particular, we set out to trial various Social Network Analysis (SNA) techniques in order to identify a range of potentially *'important'* members of the forums.

A. Social Network Analysis (SNA)

For decades, complex interactivity between entities has been modelled as networks. These include the internet [16], food webs and biochemical networks [17]. For each network, entities (such as computers, routers, animals, etc.) are considered as nodes or vertices that are connected together by links or edges (such as communications between computers, the flow of energy within a food network, etc.)

Link Analysis (LA) is the analysis of information flow within the networks above and has been a topic of study for several decades [18], [19]. A *Social Network* (SN) is defined as the representation of communication networks where the nodes are people and the edges correspond to the relationships between and *Social Network Analysis* (SNA) is defined as the application of Link Analysis to a social network. We can perform SNA on our newly created social network, where the flow of information corresponds to the flow of information on the forums which allows us to perform SNA on our network to determine hidden network structures.

II. METHODOLOGY

In order to assess the extent of the network discovery of our forums, we begin by collecting the forum posts. This in turn

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-443-5
Appendix PHILLIPS02

allowed us to extract the metadata from each forum from which to build our network. Once we had extracted the metadata mentioned above, the next step was to create a social network representation of each forum. We then set out to experiment ways in which we create a social network representation of the communication data [see section II-A]. Figure 1 shows an overview of our process.



Figure 1. An outline of our approach to extract social structure from online communications

From this social network, we calculated the top-ranked SNA metrics from previous work [14]. After calculating the metrics for each node within the network, we then used unsupervised machine learning to identify clusters of interest within our network. Unsupervised machine learning was selected due to the unavailability of ground truth of the seniority of the forum members (other than identifying the top posters). Given the nature of our metrics chosen, the clusters will be grouped based on the similarity of their metrics, which in turn we hypothesise reflects the importance of the individual within the network.

A. Creating the Social Network

Once the metadata for each forum was collected, we then set out to experiment the ways in which to convert the metadata into a Social Network. We convert the metadata into a social network using algorithm 1

Algorithm 1 Algorithm to convert a forum into a social network where w(a, s, m) is the weighted score of the message m from sender a to sender s and is determined based on our design decisions below and SNW(s_1, s_2) is the overall weight of the directed edge between s_1 and s_2 in our Social Network for a given forum f.

```
for each thread t in forum f do
    for each message m in t do
        a = author(m)
        for each previous sender s in thread t do
            SNW(a, s) = SNW(a, s) + w(a, s, m)
        end for
    end for
end for
```

For this evaluation we experimented with two distinct methods of handling when an author posted multiple times in the same thread (Repeated Author Resolution) and two distinct ways of increasing the weight between two members based on the Time difference between responses.

In order to illustrate the application of our methods, Figure 2 shows an example of a forum with two members and four posts.

1) **Repeated Author Resolution:** In Figure 2, our forum posts shows two posts for $sender1$ and two posts for $sender2$. Our experiment tested 2 ways of handling repeated authors within a thread.

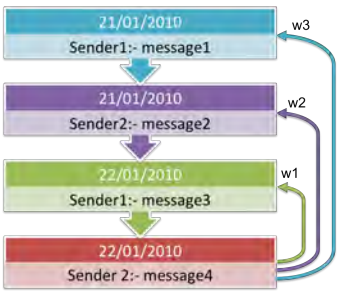


Figure 2. An example forum post with two members and four posts

Unique Senders: For this method, we only consider those senders that are unique within each thread. In this case, if we see an author reply to a thread they have previously commented on, we only add the weight for the most recent comment on the thread. In our example in Figure 2 when updating the scores after message4, only w_1 and w_2 would be added to SNA$(sender2, sender1)$ and SNA$(sender1, sender1)$ respectively and w_3 would not be calculated or added.

All Senders: For this method, we consider every sender within each thread and include duplicate senders. In this case, if we see an author reply to a thread they have previously commented on, we add the weight for each comment on the thread. In our example in figure 2 when updating the scores after $message4$, SNW$(sender2, sender1)$ is increased by $w_1 + w_3$ and SNW$(sender1, sender1)$ would be increased by w_2.

2) **Weighted Date Resolution:** When a new author a contributes a message m_a to a thread t where senders $s_1, s_2, \ldots s_k$ have previously commented, let d_{s_i} be the date that sender s_i sends a message m_{s_i} in the thread t. We then need to calculate w(a, s_i, m_a) for all i. In addition to evaluating the effectiveness of including all or unique senders, we also evaluated two distinct ways to calculate the weightings based on the time difference (d_t) between the two messages. The two methods we used to calculate the weightings were *'uniform weighting'* and *'inverse proportionality weighting'*.

Uniform weighting: Our first model assumed that the strength of the connection between two messages within a thread is independent on the amount of time taken to respond to a message. In this case, $\forall i$, w$(a, s_i, m_a) = 1$. Whilst this allows for a simple view of our forum, we set out to explore whether this model may not be representative as the same weight is given to a response to a message on the same day as a message with a response delay of 4 weeks.

Inverse Proportionality weighting: In order to overcome the issue outlined above, our second model was created on the assumption that the closer a response is in time, the stronger the weight of the connection between two users. In this case $\forall i$, w$(a, s_i, m_a) = \frac{1}{1+(d_a - d_{s_i})}$. This model ensures that $\forall i, 0 <$ w$(a, s_i, m_a) < 1$.

B. SNA Metrics

For the purpose of this paper, in order to measure the graph properties of each node that reflect importance, we used the

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-443-5

Appendix PHILLIPS02

TABLE I. A TABLE OUTLINING THE METRICS USED TO EVALUATE OUR SOCIAL NETWORK

Attribute Name	**Description**
Sent Messages (SM)	The number of emails sent by an employee.
Received Messages (RM)	The number of emails received by an employee.
Degree Centrality (DCS)	The number of distinct employees within the network that an employee has sent emails to.
Betweenness Centrality Score (BCS)	The betweenness centrality measure for an employee. [20]
Pagerank Score (PRS)	The PageRank score an employee. [21]
Markov Ranking (MR)	The markov ranking of an employee. [22]
HITS Authority Score (HAS)	The authority score for an employee (if several users with high hub weights send an email t the user then they will have a higher authority score). [23]
HITS Hub Score (HHS)	The hub score for an employee (if the user sends emails to users with high authority scores then they will have a higher hub score). [23]
Clique Score (CS)	The number of cliques (maximal subgraphs) an employee is in using the Bron and Kerbosch algorithm. [24]
Weighted Clique Score (WCS)	The weighted clique score for each user, weighted by the number of users within each clique.
Average Distance Score (ADS)	The average distance between the user and all other users in the graph.
Clustering Coefficient (CC)	The extent to which vertices in a graph tend to cluster together. [25]

SNA metrics outlined in [14] and are shown in Table I. These metrics can be split into five main categories that can be used to identify relevant properties of our network. These categories are highlighted in Figure 3. By selecting a cross-section of metrics that cover all five main categories of metrics, our machine learning model is able to capture as much information from our network as possible.

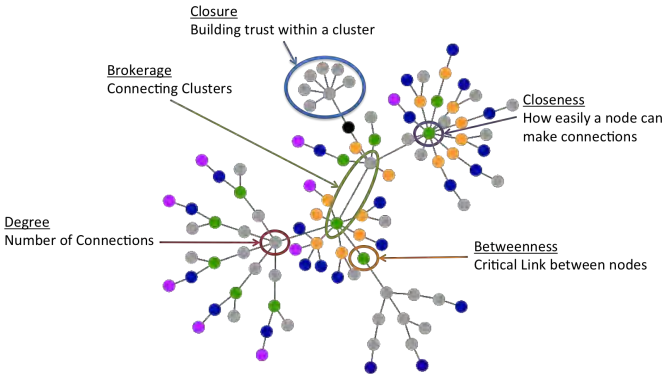


Figure 3. An outline of the five main categories of SNA metrics, namely brokerage, degree, closeness, closure and betweenness.

III. EXPERIMENTAL SETUP

A. Datasets

For this experiment we evaluated 10 of the forums collected as part of the Dark Web Forum Portal dataset[26]. This dataset was collected by crawling a variety of radical websites [27] and allows us to compare the results and its applicability over multiple forums. Table II shows a breakdown of each forum used as part of the experiment.

In order to ensure that the connections in our graph reflect meaningful connections within the network, members whose edge weights are less than 10 are pruned from our graph. This allows the graph to reflect the strong connections whilst removing those connections that do not play a central role within the network. As such, given the different weighting approaches for each scheme, after pruning edges with edge weight less than 10, each social network contained varying numbers of vertices with 1 or more edges.

For each of our seven terrorist network datasets, we created four social networks using the combination of metrics below. Once the network was created, we then calculated our SNA metrics and performed unsupervised machine learning.

- Uniform weighting and unique senders
- Uniform weighting and all senders
- Inverse-Proportionality weighting and unique senders
- Inverse-Proportionality and all senders

In order to evaluate the appropriateness of each of our methods, we performed Unsupervised Learning using Expectation Maximisation (EM) [28] in order to create clusters in order to classify forum members. EM was chosen as it is able to handle unobserved data and missing datapoints, which can reflect the fact that we may only have a partial view of the network by observing the online forums alone.

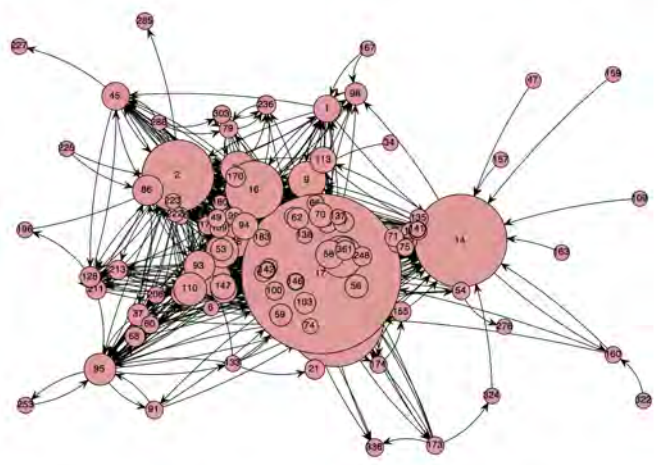


Figure 4. An example screenshot of our toolsuppport where each node is sized by the node's HITS hub score.

B. Tool Support

To allow us to visualise the results of our social network analysis and easily switch between the different weighting schemes and repeated author resolution, we built upon our existing tool [14] by adding additional features to allow users to provide their own weighting scheme to create their own Social Network Representation of the underlying forum data.

Figure 4 shows an example screenshot of our tool support displaying the social Network of the Ansar1 Network using repeated authors and inverse proportionality weighting.

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-443-5
Appendix PHILLIPS02

TABLE II. AN OVERVIEW OF THE ENGLISH LANGUAGE FORUMS USED AS PART OF OUR EXPERIMENTS

Forum	#Messages:	# Threads:	# Members:	Start Date:	End Date:	Forum URL:
Ansar AlJihad Network (Ansar1)	29492	11244	382	12/08/08	01/20/2010	http://wwww.ansar1.info/
Gawaher (Gawaher)	372499	53235	9269	10/24/2004	06/07/12	http://www.gawaher.com
Islamic Awakening (IslamicAwakening)	201287	32879	3964	04/28/2004	05/22/2012	http://forums.islamicawakening.com
Islamic Network (IslamicNetwork)	91874	13995	2082	06/09/04	11/10/10	http://talk.islamicnetwork.com
Islamic Web-Community (Myiwc)	25016	6310	756	11/05/00	02/19/2010	http://www.myiwc.com/forums/index.php
Turn To Islam (TurnToIslam)	335338	41654	10858	06/02/06	05/20/2012	http://www.turnintonislam.com/forum/
Ummah	1491957	91527	21013	04/01/02	05/18/2012	http://www.ummah.com/forum/

IV. RESULTS AND ANALYSIS

In this section we outline some of the main highlights of our research findings. For our Ansar1 Dataset, we identified five clusters. $Cluster0$ contained 26 members, $cluster1$ contained 37 members, $cluster2$ contained 1 member, $cluster3$ contained 5 members and $cluster4$ contained 42 members. Table III provides further detail on the distribution of each cluster identified by the EM model.

TABLE III. OUTLINE OF OUR CLUSTERING ALGORITHMS FOR OUR ANSAR1 DATASET USING A UNIFORMED WEIGHTING SCHEME AND INCLUDING ALL SENDERS.

Attribute	Cluster				
	0	1	2	3	4
degree					
mean	0.0953	0.0095	0.8559	0.3568	0.0293
std.dev	0.0419	0.0021	0.1104	0.0944	0.0124
betweenness					
mean	30.5428	0.0535	5549.8267	551.7925	0.0002
std.dev	42.8349	0.3254	541.1433	337.2082	0.013
PageRank					
mean	0.0105	0.0031	0.1488	0.0464	0.0053
std.dev	0.005	0.0007	0.017	0.0223	0.002
Markov					
mean	0.0113	0.002	0.1513	0.0538	0.0048
std.dev	0.0064	0.0009	0.0185	0.0258	0.0027
HITS_authority					
mean	0.0977	0.0168	0.4142	0.2655	0.0477
std.dev	0.0451	0.0118	0.0715	0.0601	0.0273
HITS_hub					
mean	0.1085	0.0149	0.4118	0.2502	0.0445
std.dev	0.0423	0.0126	0.0712	0.0506	0.0291
weighted_cliques					
mean	538.4462	2.107	5108	3875.3263	17.6769
std.dev	666.9511	0.458	997.1299	669.2559	28.3311
cliques					
mean	5.968	1.0532	110	42.0026	1
std.dev	4.5005	0.2258	13.7192	14.3096	13.7192
avgDistance					
mean	1.5016	1.4613	1.8868	1.5895	1.474
std.dev	0.027	0.0414	0.0586	0.0342	0.0306
clusteringCoeff					
mean	0.6634	0	0.0589	0.257	1
std.dev	0.1631	0.0029	0.4387	0.0935	0.4387

Our evaluator identified 5 main clusters ranging in size from 1 to 42 users. From our results, we can immediately identify that cluster 2 (with 1 user) has a significantly greater score on almost all metrics compared to the other clusters followed by cluster 3. This leads us to believe that user 17 (the sole user in cluster 2) has significantly higher influence within the network, compared followed by those members in cluster 3 and as such, these users are potential candidates for being "important" users within the Ansar1 network. The significantly lower standard deviation for clusters 1 and 4 indicate that these users are likely to be less important in the network with their low scores in the ranking metrics (e.g. PageRank, HITS Hub, etc.).

We also evaluated the difference in ranking of each created social network and Table IV shows the top ten ranked users

TABLE IV. THE TOP TEN USERS USING ALL FOUR SOCIAL NETWORK METRICS SCHEME USING OUR ANSAR1 DATASET RANKED USING THE PAGERANK ALGORITHM

Inverse-1Date	Inverse-AllDates	Uniform-AllDates	Uniform-1Date
17	17	17	17
14	14	0	0
0	0	14	14
16	2	2	2
2	16	16	16
11	39	39	39
9	11	11	11
39	9	12	9
33	33	33	33

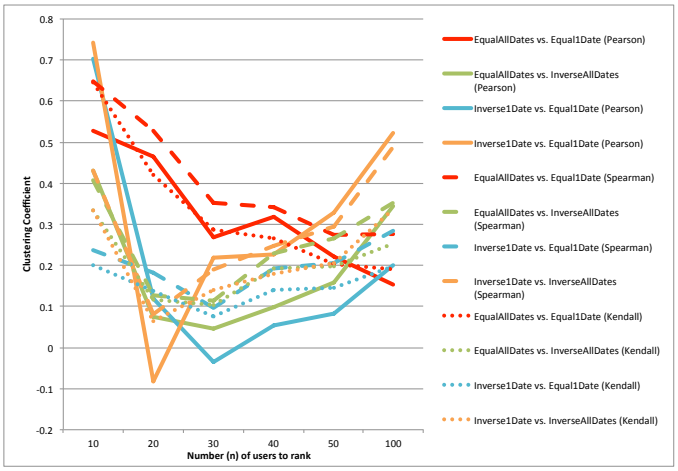


Figure 5. An outline of the correlation coefficients for each of our four networks using Spearman's, Kendall's and Pearson's correlation coefficients.

(using the PageRank algorithm) for each of the four networks using the Ansar1 dataset. The table shows that all four networks identified user 17 as the most influential user and agreed on the top three users (but had different orderings of users 0 and 14). Similarly, they all agree on the top 10 users (with the exception of the UniformAllDates network, which substitutes user 9 for user 12.

In order to evaluate the effect that weighting and repeated author resolution has on our model, we evaluated the consistency of our rankings with schemes by comparing the top n ranked users (for $n = 10, 20, 30, 40, 50$ and 100) using Spearman's Coefficient Ranking [29], Kendall's Tau Measure [30] and Pearson's Correlation Coefficient [31]. Table V and Figure 5 shows the ranking scores of each comparison with only one degree of freedom (either identical weighting and different repeated author resolution or vice versa). Our results showed that on average, varying the repeated author resolution caused less variation in the ranking of the metrics in the final social network when applied for $n < 50$. However, the difference between rankings based on their repeated author

TABLE V. COMPARISON OF CORRELATION BETWEEN RANKING ALGORITHMS USING OUR ANSAR1 DATASET

Attribute	Number of users (n) to compare ranking using PageRank																	
	10			20			30			40			50			100		
	P	S	K	P	S	K	P	S	K	P	S	K	P	S	K	P	S	K
UniformAllDates vs.Uniform1Date	0.529	0.648	0.644	0.466	0.528	0.421	0.270	0.354	0.287	0.318	0.343	0.267	0.221	0.274	0.203	0.152	0.276	0.191
UniformAllDates vs. InverseAllDates	0.431	0.406	0.333	0.074	0.128	0.116	0.045	0.113	0.103	0.099	0.231	0.192	0.157	0.265	0.198	0.346	0.352	0.253
Inverse1Date vs. Uniform1Date	0.704	0.236	0.200	0.119	0.182	0.137	-0.035	0.097	0.076	0.054	0.192	0.141	0.084	0.205	0.146	0.202	0.285	0.197
Inverse1Date vs. InverseAllDates	0.741	0.430	0.333	-0.083	0.081	0.063	0.218	0.191	0.140	0.227	0.247	0.179	0.328	0.294	0.207	0.521	0.489	0.338

resolution diminishes as n increases. This leads us to believe that performing two social networks, one using repeated authors and one using unique authors only may provide subtly different views of the graph, which in turn will allow us to gain more insight from the social network.

V. CONCLUSIONS AND FUTURE WORK

In our paper, we set out to investigate the effectiveness of using metadata alone to identify influential and important users on Dark Web Forums. We set out to investigate the effect of repeated author resolution and various weighting schemes on our rankings by creating four social networks per forum and evaluating the consistency of the top n users (for $n = 10, 20, 30, 40, 50, 100$). We also performed unsupervised machine learning for each of network in order to identify clusters by user's importance. Our results showed us that difference in rankings from different weighting schemes were more consistent on average than those using different repeated author resolution techniques.

In order to allow us to further evaluate the validity of our work, we hope to establish an authoritative ground truth in order to assess the relative performance of our work. Another direction we hope to take the research is to perform dynamic analysis on our network and assess how the network changes over time. This technique could then be applied to identify potential insider threats within the Dark Web Forums by observing abnormal dynamic behaviour. In order to allow us to further evaluate the validity of our work, we hope to establish an authoritative ground truth in order to assess the relative performance of our work. Another direction we hope to take the research is to perform dynamic analysis on our network and assess how the network changes over time. This technique could then be applied to identify potential insider threats within the Dark Web Forums by observing abnormal dynamic behaviour.

ACKNOWLEDGMENT

This work has been made possible through the EPSRC Scholarship as part of the Oxford University Centre for Doctoral Training in Cyber Security.

REFERENCES

[1] K. Crilley, "Information warfare: new battle fields terrorists, propaganda and the internet," in Aslib Proceedings, vol. 53, no. 7. MCB UP Ltd, 2001, pp. 250–264.

[2] Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, "Us domestic extremist groups on the web: link and content analysis," Intelligent Systems, IEEE, vol. 20, no. 5, 2005, pp. 44–51.

[3] P. B. Gerstenfeld, D. R. Grant, and C.-P. Chiang, "Hate online: A content analysis of extremist internet sites," Analyses of social issues and public policy, vol. 3, no. 1, 2003, pp. 29–44.

[4] S. Mehrotra, Intelligence and Security Informatics: IEEE International Conference on Intelligence and Security Informatics, ISI 2006, San Diego, CA, USA, May 23-24, 2006. Springer Science & Business Media, 2006, vol. 3975.

[5] H. Chen, S. Thoms, and T. Fu, "Cyber extremism in web 2.0: An exploratory study of international jihadist groups," in Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on. IEEE, 2008, pp. 98–103.

[6] V. E. Krebs, "Mapping networks of terrorist cells," Connections, vol. 24, no. 3, 2002, pp. 43–52.

[7] V. Krebs, "Uncloaking terrorist networks," First Monday, vol. 7, no. 4, 2002.

[8] A. Basu, "Social network analysis of terrorist organizations in india," in North American Association for Computational Social and Organizational Science (NAACSOS) Conference, 2005, pp. 26–28.

[9] C. C. Yang, N. Liu, and M. Sageman, "Analyzing the terrorist social networks with visualization tools," in Intelligence and security informatics. Springer, 2006, pp. 331–342.

[10] T. Stevens, "Regulating the ?dark web?: How a two-fold approach can tackle peer-to-peer radicalisation," The RUSI Journal, vol. 154, no. 2, 2009, pp. 28–33.

[11] J. Xu and H. Chen, "The topology of dark networks," Communications of the ACM, vol. 51, no. 10, 2008, pp. 58–65.

[12] G. L'Huillier, S. A. Ríos, H. Alvarez, and F. Aguilera, "Topic-based social network analysis for virtual communities of interests in the dark web," in ACM SIGKDD Workshop on Intelligence and Security Informatics, ser. ISI-KDD '10. New York, NY, USA: ACM, 2010, pp. 9:1–9:9. [Online]. Available: http://doi.acm.org/10.1145/1938606.1938615 [Accessed: 2015-10-19]

[13] T. Guardian, "Edward snowden | world news | the guardian," 06 2015. [Online]. Available: http://www.theguardian.com/world/edward-snowden [Accessed: 2015-10-10]

[14] E. Phillips, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Applying social network analysis to security," in International Conference on Cyber Security for Sustainable Society, 2015, pp. 11–27.

[15] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting structural re-identification in anonymized social networks," Proceedings of the VLDB Endowment, vol. 1, no. 1, 2008, pp. 102–114.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," ACM SIGCOMM Computer Communication Review, vol. 29, no. 4, 1999, pp. 251–262.

[17] A. Golightly and D. J. Wilkinson, "Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo," Interface Focus, 2011, p. rsfs20110047.

[18] L. Getoor and C. P. Diehl, "Link mining: A survey," SIGKDD Explor. Newsl., vol. 7, no. 2, Dec. 2005, pp. 3–12. [Online]. Available: http://doi.acm.org/10.1145/1117454.1117456 [Accessed: 2015-10-18]

[19] S. Wasserman, Social Network Analysis: Methods and Applications. Cambridge University Press, Nov. 1994.

[20] L. C. Freeman, "Centrality in social networks conceptual clarification," Social networks, vol. 1, no. 3, 1979, pp. 215–239.

[21] I. Rogers, The Google Pagerank algorithm and how it works, 2002. [Online]. Available: http://mira.sai.msu.ru/~megera/docs/IR/search/pagerank/pagerank_explained.pdf [Accessed: 2015-09-15]

[22] D. Koschtzki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, "Centrality Indices," in Network Analysis, ser. Lecture Notes in Computer Science, U. Brandes and T. Erlebach, Eds. Springer Berlin Heidelberg, 2005, no. 3418, pp. 16–61. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-31955-9_3 [Accessed: 2015-05-19]

[23] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," J. ACM, vol. 46, no. 5, Sep. 1999, pp. 604–632. [Online]. Available: http://doi.acm.org/10.1145/324133.324140 [Accessed: 2015-10-15]

[24] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an

Appendix PHILLIPS02

undirected graph," Communications of the ACM, vol. 16, no. 9, 1973, pp. 575 – 577. [Online]. Available: http://dl.acm.org/citation.cfm?id=362367 [Accessed: 2015-06-15]

[25] S. N. Soffer and A. Vazquez, "Network clustering coefficient without degree-correlation biases," Physical Review Series E-, vol. 71, no. 5, 2005, p. 057101.

[26] "Dark web forum portal." [Online]. Available: http://cri-portal.dyndns.org/portal/Home.action [Accessed: 2015-10-10]

[27] H. Chen, "Dark web forum portal," in Dark Web. Springer, 2012, pp. 257–270.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," Journal of the royal statistical society. Series B (methodological), 1977, pp. 1–38.

[29] S. B. Lyerly, "The average spearman rank correlation coefficient," Psychometrika, vol. 17, no. 4, 1952, pp. 421–428.

[30] G. S. Shieh, "A weighted kendall's tau statistic," Statistics & Probability Letters, vol. 39, no. 1, 1998, pp. 17–24.

[31] P. Sedgwick et al., "Pearsons correlation coefficient," BMJ, vol. 345, 2012.

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-443-5

Appendix PHILLIPS02

https://web.archive.org/web/20150315000000*/http://ora.ox.ac.uk/information/termsOfUse June 16, 2025

INTERNET ARCHIVE | WEB | TEXTS | VIDEO | AUDIO | SOFTWARE | IMAGES | SIGN UP | LOG IN | UPLOAD | Search

ABOUT | BLOG | PROJECTS | HELP | DONATE | CONTACT | JOBS | VOLUNTEER | PEOPLE

DONATE

INTERNET ARCHIVE WayBackMachine

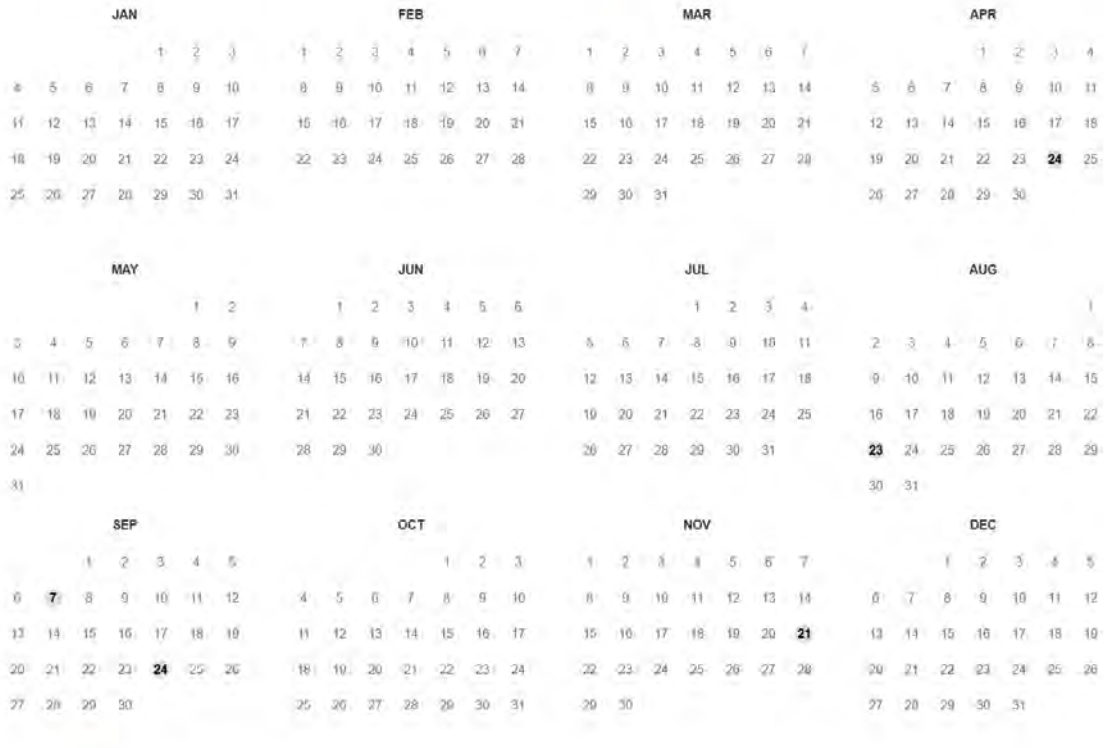
Explore more than 946 billion web pages saved over time

http://ora.ox.ac.uk/information/termsOfUse

Calendar · Collections · Changes · Summary · Site Map · URLs

Saved **23 times** between September 20, 2012 and January 25, 2025.





Note

This calendar view maps the number of times **http://ora.ox.ac.uk/information/termsOfUse** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the FAQ.

FAQ | Contact Us | Terms of Service (Dec 31, 2014)

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org

Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

https://web.archive.org/web/20151121124431/http://ora.ox.ac.uk:80/information/termsOfUse June 16, 2025

INTERNET ARCHIVE Wayback Machine http://ora.ox.ac.uk/information/termsOfUse Go SEP NOV APR
23 captures 21
20 Sep 2012 - 25 Jan 2025 2014 2015 2017 About this capture

Terms and Conditions of Use for Oxford University Research Archive

Terms of Use

The University of Oxford ('the University'), through the Bodleian Libraries, has developed Oxford University Research Archive ('ORA') so that users may access research outputs of the University.

Copyright and other rights in the items held in ORA are retained by the individual authors, the University, or other third parties. The extent of the rights reserved and the identity of the parties retaining those rights will vary on an item-by-item basis.

Users are required to comply with the permissions notice that applies to each item in ORA. Unless otherwise indicated in the relevant permissions notice on the item record, users may download and/or print one copy of any item in ORA to facilitate their private study or research for non-commercial purposes. Whilst items held in ORA may be subject to multiple intellectual property rights, depositors only licence the copyright through ORA, and no licence under any patents, design rights, trade marks or any other right that may exist in relation to such items is granted or implied. Some items in ORA are distributed under Creative Commons or other licences which may allow for further commercial or non-commercial use. Subject to such licences, users may not use material obtained from ORA for any profit-making activities or any commercial gain.

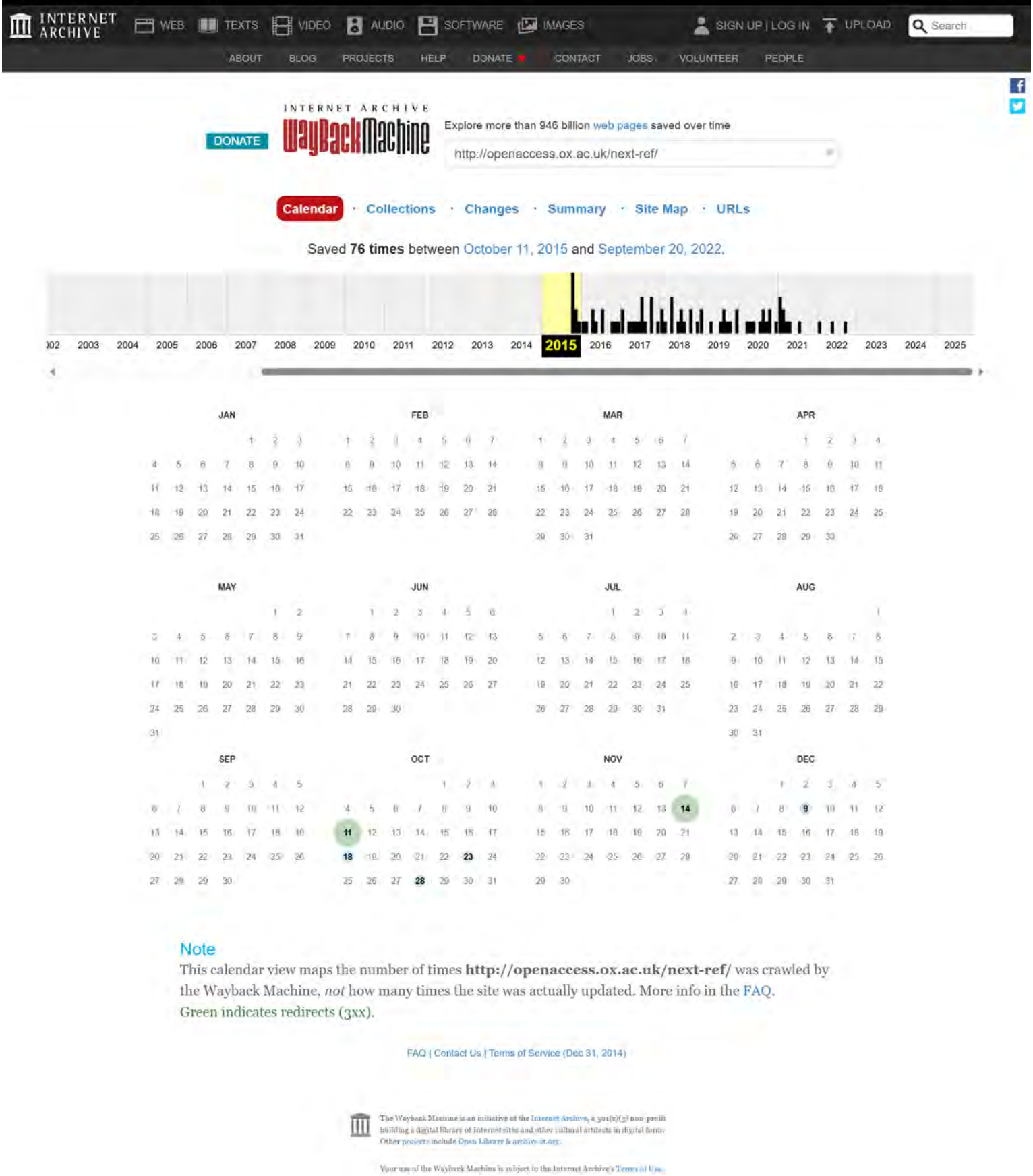
You may freely distribute the URL (http://ora.ox.ac.uk/) of the ORA website but if you wish to link to any item in ORA you must do so through the item record page.

These terms are governed by English law, and the Courts of England have exclusive jurisdiction in relation to them.

UNIVERSITY OF OXFORD

Site Copyright © 2008 Oxford University Research Archive. All rights reserved. Site not yet optimized for viewing in mobile devices

Disclaimer and Data Protection statement Cookies Accessibility statement Copyright Take-down procedure and complaints Contact


INTERNET ARCHIVE
WEB TEXTS VIDEO AUDIO SOFTWARE IMAGES
SIGN UP | LOG IN UPLOAD
Search
ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE
DONATE
INTERNET ARCHIVE
WayBackMachine
Explore more than 946 billion web pages saved over time
http://openaccess.ox.ac.uk/next-ref/
Calendar · Collections · Changes · Summary · Site Map · URLs
Saved 76 times between October 11, 2015 and September 20, 2022.
2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
JAN FEB MAR APR
MAY JUN JUL AUG
SEP OCT NOV DEC
Note
This calendar view maps the number of times http://openaccess.ox.ac.uk/next-ref/ was crawled by the Wayback Machine, not how many times the site was actually updated. More info in the FAQ.
Green indicates redirects (3xx).
FAQ | Contact Us | Terms of Service (Dec 31, 2014)
The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.
Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

https://web.archive.org/web/20151209231304/http://openaccess.ox.ac.uk/next-ref/ June 16, 2025

INTERNET ARCHIVE WayBack Machine http://openaccess.ox.ac.uk/next-ref/ Go OCT DEC 09 JAN 2014 2015 2016 About this capture
76 captures
11 Oct 2015 - 20 Sep 2022

Act on Acceptance: Open Access and the next REF

In order to be eligible for the next REF, the final peer-reviewed version of journal articles and conference papers (with an ISSN) must be deposited in an open access repository within three months of acceptance for publication. HEFCE's requirements are designed to 'increase substantially the amount of scholarly material that is made available in an open-access form'.

Key points of the policy

- applies to journal articles and conference papers (with an ISSN)
- requires researchers to act on acceptance to deposit their final peer-reviewed text (also known as the 'Author Accepted Manuscript') in a compliant open access repository within three months of the date of acceptance
- applies to article manuscripts accepted after 1 April 2016
- readers and search engines must be able to find out about the article straight away
- full-text must be accessible for anyone to read and download either immediately or after the end of an embargo period*

*Embargo periods should not exceed the following: 12 months for Science and Medicine (REF Main Panels A and B); 24 months for Social Sciences and Humanities (REF Main Panels C and D)

What this means for researchers

- failure to act will make research outputs ineligible for the next REF
- requires additional action; meeting the open access requirements of other funders is not enough
- you still need to meet the requirements of other funders such as RCUK and Wellcome Trust

Oxford's message: you deposit; we'll take care of the rest

- all academics and researchers employed by the collegiate University
- all journal articles and conference papers deposited in ORA
- deposit via Symplectic Elements
- start now

The Bodleian team will check copyright and licensing conditions, create an ORA record and make the full-text available if and when it complies with the requirements of your publisher.

Click here to Deposit your Work.

Need help?

- FAQs open access and next REF
- How to Deposit your Work
- Email openaccess@bodleian.ox.ac.uk

Related Links:

- Symplectic Elements at Oxford
- HEFCE's full policy and accompanying FAQ

About the Oxford Open Access Project

The Oxford Open Access project is a collaborative project involving Research Services, the Bodleian Libraries, IT Services, the Planning and Resource Allocation Section, OUP and the Academic Divisions, accountable to the Pro Vice-Chancellor (Research) and the Research Committee.

Tweets by @OAOxford

Copyright

Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 3.0 Unported License. (c) 2013, The University of Oxford.

Images are copyright (c) Zoe Power. All rights reserved.

Site Feedback

If you have any feedback on how the site could be more useful then please email Juliet Ralph at the address below.

Email : juliet.ralph@bodleian.ox.ac.uk

https://web.archive.org/web/20150701000000*/http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/ June 16, 2025

INTERNET ARCHIVE | WEB | TEXTS | VIDEO | AUDIO | SOFTWARE | IMAGES | SIGN UP | LOG IN | UPLOAD | Search

ABOUT | BLOG | PROJECTS | HELP | DONATE | CONTACT | JOBS | VOLUNTEER | PEOPLE

DONATE

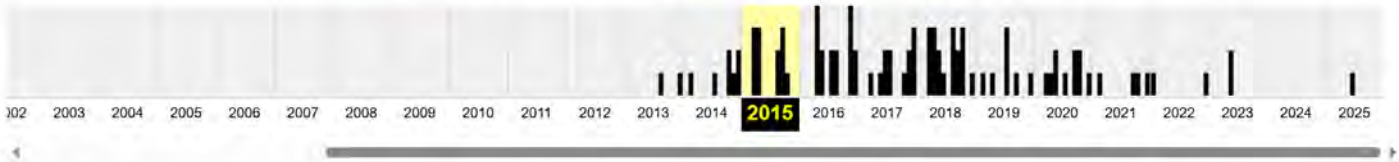
INTERNET ARCHIVE WayBackMachine

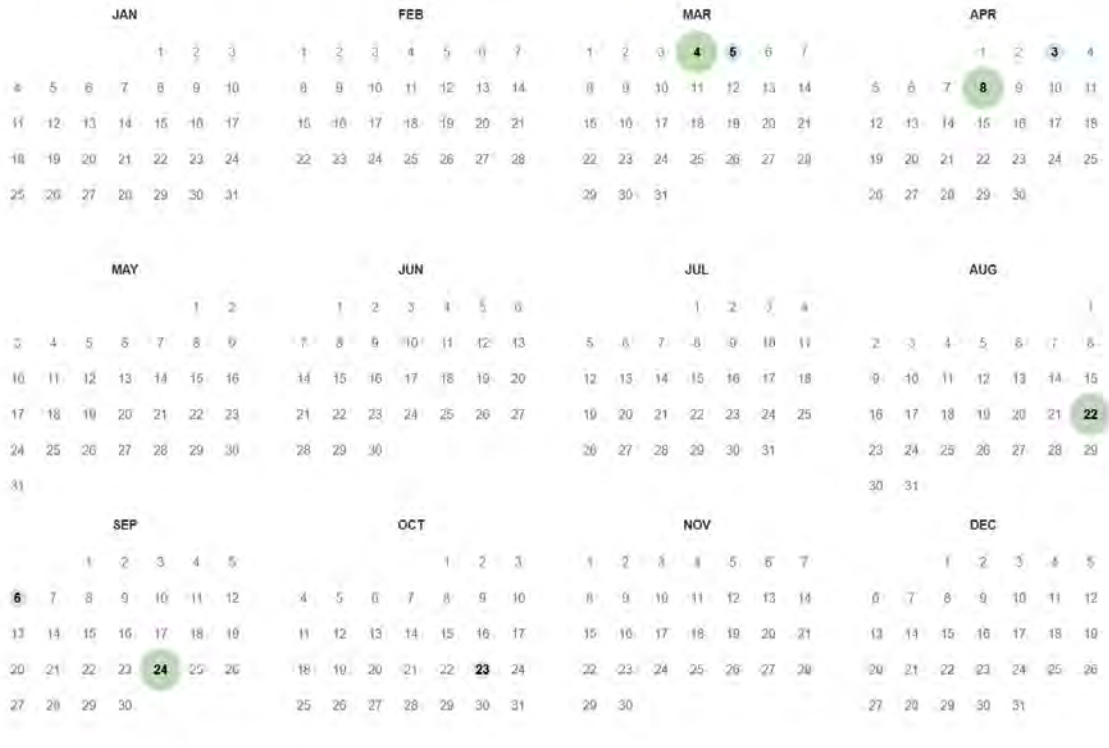
Explore more than 946 billion web pages saved over time

http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/

Calendar · Collections · Changes · Summary · Site Map · URLs

Saved **93 times** between August 8, 2013 and June 10, 2025.





Note

This calendar view maps the number of times **http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the FAQ.

Green indicates redirects (3xx).

FAQ | Contact Us | Terms of Service (Dec 31, 2014)

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org

Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

https://web.archive.org/web/20151023045941/http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/ June 16, 2025

INTERNET ARCHIVE WayBack Machine http://openaccess.ox.ac.uk/home-2/open-access-at-oxford/ Go SEP OCT APR 23 2014 2015 2016 About this capture
93 captures 8 Aug 2013 - 10 Jun 2025

What is Open Access?

Open-access (OA) literature is digital, online, free of charge including for those who do not have personal or institutional subscriptions to journals, and free of most copyright and licensing restrictions. What makes it possible is the internet and the consent of the author or copyright-holder. (See Peter Suber's Open Access Overview).

Read more about the development of Open Access at the Open Access Chronology blog post.

Open Access at Oxford

On 11 March 2013, Council approved the Statement on Open Access at the University of Oxford. Some key messages are:

- University of Oxford academics and researchers are strongly committed to, and already very active in, ensuring the widest possible access to research findings for the benefit of the international research community and society more widely. The value and utility of research outputs increases the more broadly they are available to be considered and used by others.
- Our academics, researchers, staff and students must be free to publish in the journal of their choice.
- The Oxford Research Archive was established in 2007 as a permanent and secure online archive of research materials produced by members of the University of Oxford. We are enhancing the coverage and features in ORA to provide a single point of public access to electronic copies of peer-reviewed journal articles, conference proceedings by Oxford authors and Oxford research theses. This also provides a means for institutional compliance with funders' Open Access requirements.
- The University favours, now and for the foreseeable future, open access by means of the **Green** Route. This enables authors to publish in any journal and to self-archive a version of the article for free public use. Thus the Green approach provides most of the benefits for a much smaller additional cost than Gold Open Access (which involves paying Article Processing Charges) (APCs).
- We will use special allocations from our funders in those cases where their policy requires that APCs be paid.

Useful Links

RCUK Block grant check eligibility and apply
Wellcome Trust / COAF Block Grant - check eligibility and request funds for APCs

Get Help - on the contact us page
Sponsor Requirements on the What do I need to do page
Glossary - Key Terms Explained
FAQ - Answers to your questions

Follow Us

Follow us on Twitter

What is Open Access?



Published on YouTube, Oct 25, 2012
What is open access? Nick Shockey and Jonathan Eisen take us through the world of open access publishing and explain just what it's all about. Make sure to watch it in Fullscreen using the icon at bottom right of the video.

About the Oxford Open Access Project

The Oxford Open Access project is a collaborative project involving Research Services, the Bodleian Libraries, IT Services, the Planning and Resource Allocation Section, OUP and the Academic Divisions, accountable to the Pro Vice-Chancellor (Research) and the Research Committee.

Tweets by @OAOxford

Site Feedback

If you have any feedback on how the site could be more useful then please email Juliet Ralph at the address below.

Email : juliet.ralph@bodleian.ox.ac.uk

Copyright

Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 3.0 Unported License. (c) 2013, The University of Oxford.

Images are copyright (c) Zoe Power. All rights reserved.

MARC View

LEADER 01547cam a2200361 i 4500
001 14769607
003 SIRSI
005 20170314130250.0
006 m d
007 cr n
008 141027t20122012njua sb 101 0 eng d
010 a| 2014207096
020 a| 9781467321051 (pbk.)
040 a| DLC b| eng c| DLC d| DLC d| WaSeSS
042 a| pcc
050 0 0 a| QA76.76.E95 b| I86 2012
082 0 0 a| 005.8 2| 23
111 2 a| IEEE International Conference on Intelligence and Security Informatics d| (2012 : c| Washington, D.C.)
245 1 0 a| 2012 IEEE International Conference on Intelligence and Security Informatics (ISI 2012) h| [electronic resource] : b| Washington, DC, USA, 11-14 June 2012.
246 3 0 a| ISI 2012
260 a| Piscataway, NJ : b| IEEE, c| [2012]
300 a| xv, 240 pages : b| illustrations ; c| 27 cm
500 a| "IEEE Catalog Number: CFP12ITI-POD."
504 a| Includes bibliographical references and index.
506 a| License restrictions may limit access.
650 0 a| Expert systems (Computer science) v| Congresses.
650 0 a| Data mining v| Congresses.
650 0 a| Computer security v| Congresses.
650 0 a| National security v| Congresses.
655 0 a| Electronic books.
856 4 0 u| http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000781181&title=2012%20IEEE%20International%20Conference%20on%20Intelligence%20and%20Security%20Informatics
037 b| 00066573
949 a| Electronic resource w| ASIS c| 1 i| 14769607-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 4/22/2015 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES
Libraries Home
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory
(814) 865-6368

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES
Facebook
Twitter
Instagram

2012 IEEE International Conference on Intelligence and Security Informatics (ISI 2012) [electronic resource] : Washington, DC, USA, 11-14 June 2012

Conference Author	IEEE International Conference on Intelligence and Security Informatics (2012 : Washington, D.C.)
Additional Titles	ISI 2012
Published	Piscataway, NJ : IEEE, [2012]
Physical Description	xv, 240 pages : illustrations ; 27 cm

ASK A LIBRARIAN

Access Online

serialssolutions.com

Restrictions on Access

License restrictions may limit access.

Subject(s)

Expert systems (Computer science)—Congresses
Data mining—Congresses
Computer security—Congresses
National security—Congresses

Genre(s)

Electronic books

ISBN

9781467321051 (pbk.)

Note

"IEEE Catalog Number: CFP12ITI-POD."

Bibliography Note

Includes bibliographical references and index.

View MARC record | catkey: 14769607

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

Securing Cyberspace: Identifying Key Actors in Hacker Communities

Victor Benjamin, Hsinchun Chen
Department of Management Information Systems
The University of Arizona
Tucson, AZ 85721, USA
vabenji@email.arizona.edu, hchen@eller.arizona.edu

***Abstract*— As the computer becomes more ubiquitous throughout society, the security of networks and information technologies is a growing concern. Recent research has found hackers making use of social media platforms to form communities where sharing of knowledge and tools that enable cybercriminal activity is common. However, past studies often report only generalized community behaviors and do not scrutinize individual members; in particular, current research has yet to explore the mechanisms in which some hackers become key actors within their communities. Here we explore two major hacker communities from the United States and China in order to identify potential cues for determining key actors. The relationships between various hacker posting behaviors and reputation are observed through the use of ordinary least squares regression. Results suggest that the hackers who contribute to the cognitive advance of their community are generally considered the most reputable and trustworthy among their peers. Conversely, the tenure of hackers and their discussion quality were not significantly correlated with reputation. Results are consistent across both forums, indicating the presence of a common hacker culture that spans multiple geopolitical regions.**

Keywords - Cyber security, Cybercrime, Hacker, Hacker community, Social media, Reputation, Key actor

I. INTRODUCTION

As the computer becomes more ubiquitous throughout society, the security of networks and information technologies is a growing concern. Critical infrastructures (e.g. power grids) are facing growing number of cyber-based threats that could result in service disruptions and physical damage. For example, the recent Stuxnet Worm was engineered to specifically target a country and destroy nuclear infrastructure. Further, keystroke logging software was found covertly spreading at the United States Drone Fleet Command. With growing security concerns, DARPA has emphasized that U.S. cyberspace is unsecure and more cyber security research is vital to national defense.

Advancing technology enables hackers to commit cybercrime much more easily now than in the past. Accessibility to technologies and methods to commit cybercrime has grown. Even legitimate tools, such as search engines, can be used to promote cybercriminal activity [1]. Malicious software, hacking tutorials, and other resources intended to help conduct cybercrime can be commonly found within hacker communities, often available for free or traded within black markets [2].

As a result, security researchers have taken interest in exploring hacker communities. Some research has found that community participants pool together skills and assets to form groups, often in order to accomplish operations much more advanced than any one individual could perform alone [3]. In particular, less skillful hackers often seek help from more experienced individuals, creating a meritocratic hierarchy within hacking culture [4]. Such skilled individuals could thus be considered as key actors within their communities, as they are heavily relied upon by peers. However, many participants of such communities aggressively protect their anonymity in order to minimize criminal evidence [5]. Given the importance of key actors in the formation of cybercriminal groups, as well as their pursuit of total anonymity, they make ideal subjects for further research aiming to secure cyberspace

II. LITERATURE REVIEW

To form the basis for this research, literature is reviewed from the following four areas:

- Past studies on hacker communities and cybercrime
- The significance of reputation within communities
- Control theory and criminal group leadership
- Previous literature on social media analytics

A. Hacker communities, and cybercrime

Hackers often congregate within underground communities, most commonly in the form of Internet Relay Chat (IRC) networks or online forums. Such communities are primarily used to share knowledge and resources [6]. Some communities evolve to include black markets where participants offer expertise, snippets of code, fully-developed applications, or stolen confidential data (e.g. credit card accounts) in exchange for other virtual goods or financial gain [2].

Due to the lucrative nature of cybercrime, some hackers use communities as a platform to organize into groups and launch sophisticated, financially-motivated cyber-attacks. Some past research has looked to develop methods for identifying collaboration between hackers, potentially revealing such groups [7]. Other research has studied group dynamics, revealing that peer approval and reputation are heavily integrated within cybercriminal group operations. This leads to a strong desire to increase reputation among hackers [8]. As the hierarchical chain of command within cybercriminal groups is heavily tied to member reputation, good reputation becomes an extremely valuable attribute hackers desire and pursue.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

B. The significance of reputation

Reputation has a strong influence on the dynamics of social groups. Individuals with good reputations are more able to cooperate or receive help from others, as they are more trusted [9]. Their reputation can be affected by various factors in a group setting; for example, seniority or tenure within a social group can often amplify the effects of good reputation [10]. Within the context of criminal groups, trust and reputation between both members and leadership is necessary for a group to remain functional [11].

Establishing a good reputation within a community or group is also a vehicle to obtain leadership. Emergence of leaders is often considered the outcome of a self-organizing process involving reputation, as reputation constitutes a precursor to leadership and affects how much stakeholders trust a leader [12]. Leaders are considered among their social group to be experts and hold great tacit knowledge [13]. They also contribute to cognitive advance of a community, and individuals characterized by high levels of activity are often more reputable and able to obtain leadership positions [14]. Reputation appears to underlie many hacker activities, including the formation of group leadership.

C. Control theory and criminal group leadership

Control theory states that a system is controllable if it can be driven from its initial state to another final state given specific inputs over some amount of time [15]. The general objective of control theory is to calculate how a system's inputs should be manipulated in order to achieve desired output, e.g. a car's velocity can be controlled by how far one pushes a gas pedal. Two different factors are said to contribute towards a system's controllability: the system's overall structure and the rules by which the system's components interact [16]. Thus, systems with different structures or component interactions are controlled in vastly different means.

When control theory is applied to social networks, identification and manipulation of driver nodes, or nodes that are said to control a system, can be used to guide a network's dynamics [17]. When considering cars, the steering wheel, gas pedal, brake pedal, and shifter are all considered driver nodes as they manipulate the remainder of their system. Unfortunately, driver nodes are generally unknown in real systems and methods to identify driver nodes need to be developed for each application context. However, past research has identified that criminal networks may have leaders that command and steer their entire group [18]. Criminal leaders exhibit characteristics that match the operational definition of driver nodes, and thus they may be considered as such. When attempting to identify the driver nodes or key actors of online social networks, such as hacker communities, it becomes necessary to develop a further understanding of methods used in social media analytics.

D. Previous studies on social media analytics

With the advent of Web 2.0, Internet users can publish content through a variety of means. Blogs, Wikis, and web forums are just a few mediums individuals can use to publish information. Each platform contains unique contents and offers opportunities for analyses. For example, in the context of eLearning forums, each student's discussion quality was found to be related to the number of posts they created, the average length of their posts, and the average number of replies they made within a discussion started by a peer [19] In programming communities, user reputation has been found closely tied to the breadth and depth of knowledge an individual possesses; users that can produce thoughtful comments across a variety of topics have been considered as domain experts [20].

User reputation has also traditionally had a role in social media analytics. Reputation is crucial as a means of establishing trust, status, and for fostering social interactions [9]. It is subject to many factors of user behaviors, such as activity levels and discussion quality. For example, past work found that forum users who mostly contributed short messages were unable to successfully maintain communication with peers [21]. Other research has considered the effects of tenure, or length of membership within a community, as an influencer of reputation. Conclusions suggest tenure may amplify reputation [22]. Finally, content quality has been observed as an important contributor towards reputation. For example, attaching or linking to additional resources in a message can positively influence its perceived quality [23]. This is especially important in hacker communities, as reputation is regarded as extremely valuable and often influences social interactions [8].

III. RESEARCH GAPS AND QUESTIONS

Researchers have been interested in exploring hacker communities and cybercriminal groups, but few inquiries have focused on identifying key actors, or hackers considered very reputable and have leadership potential or are already leaders. Law enforcement would benefit by focusing resources on apprehending key actors, as they may be the driver nodes of their respective networks and thus potentially responsible for directing cybercriminal activity. Additionally, there is a lack of work studying mechanisms which individuals to become key actors within hacker environments. Security researchers would benefit from knowledge of the processes hackers use to gain status, power, and leadership within their communities

In this study, we analyze the relationships between aspects of hacker behavior and reputation. Analysis is performed across two diverse hacking communities from the United States and China. We pose the following research questions:

1. How does the behavior of an individual hacker affect how other hackers perceive them?
2. In what ways can one build their reputation within hacking communities?
3. Are hacker communities different across cultures? In what ways are they similar?

IV. RESEARCH TESTBED AND DESIGN

The two communities used for this study are web forums based out of the United States and China. Communities were chosen for their size relative to other hacker communities, and

This work was supported in part by the National Science Foundation under Grant No. CBET-0730908, and by the Defense Threat Reduction Agency, Award No HDTRA1-09-1-0058.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

also for the growing relevance of cyber security in both countries. Other hacker communities suitable for this research could possess differing geopolitical origins, such as the Middle-East or Russia. Additionally, such communities may exist within entirely different mediums (e.g. IRC channels). However, web forums are particularly useful to study as they naturally provide access to data spanning multiple years into the past. Both communities we explored in this research also feature unique mechanisms allowing hackers to attach files or source code to their messages, promoting collaboration. We also perform analysis of hacker communities from differing geopolitical regions, as this allows for further observation of an expanding global phenomenon.

TABLE I. RESEARCH TESTBED

Forum Name	Language	# of Messages	# of Users	Forum Start Date
Hackhound.org	English	77,061	5,794	October 9, 2008
Unpack.cn	Chinese	646,494	22,743	October 12, 2004

written to extract relevant content. Text parsers were written to extract messages and user information from both of the communities.

Many previous studies have frequently used regression models to determine individuals' reputations. For example, ordinary least squares (OLS) regression was used to model leadership in business group settings [25]. The reputations and trustworthiness of online multiplayer gamers was modeled using a regression approach [26]. Additionally, to help find quality content on Yahoo! Answers, user reputations have been assessed through a series of regression models using message content features [27].

For this study, we borrow six features from past research and observe them within our hacker community dataset. As previous studies often aimed to observe user forum involvement and user discussion quality within virtual communities, we select related features. Two of our selected features aim to capture hacker discussion quality while the other four features are intended to model the extent of a hacker's involvement in their community.

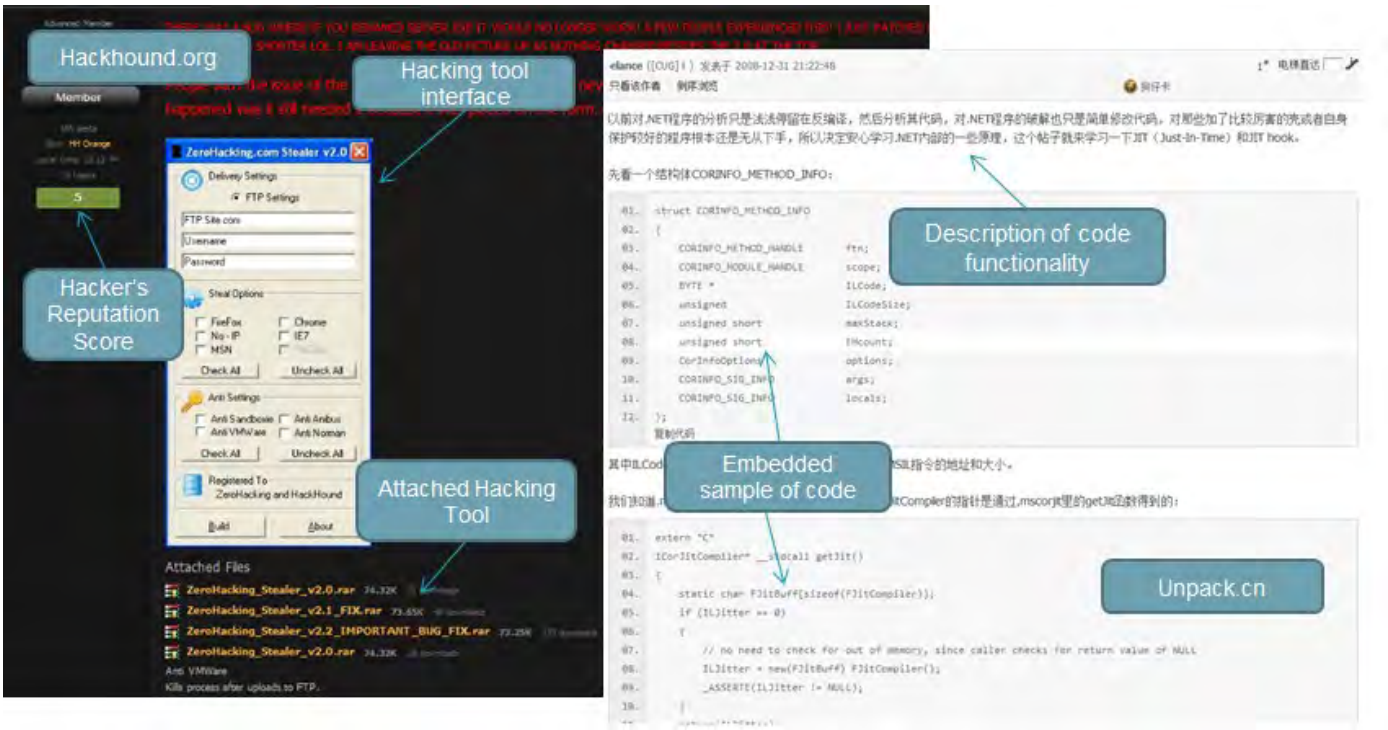


Figure 1. **Left**: A member of Hackhound.org publishes the latest version of his hacking tool meant to help others steal cached passwords on victims' computers **Right:** A hacker of the Chinese community Unpack.cn posts sample code for reverse engineerimg software written in the Microsoft .NET framework

Both communities allow for the unique feature for hackers to attach hacking tools and program source code to their messages for others to use (Figure 1). Additionally, both communities allow hackers to assign each other a reputation score in order to rate one another's usefulness and trustworthiness. Communities were identified through keyword searches similar to past research [24] Keywords such as "malware detection avoidance" were searched on Google.com and Bing.com to find relevant communities. Automated spiders were used to collect the communities and text parsers were

To conduct analysis of what features are related to hacker reputation, we employ an OLS regression similar to previous work. Each forum allows for hackers to assign each other reputation scores through forum mechanisms, resulting in reputation as a method for peers to evaluate each other. We observe the relationships between all selected features and reputation in attempt to reveal why certain hackers emerge as most reputable and trustworthy among their peers. Such hackers often have leadership potential or are already commanding less skilled individuals, and thus can be considered as key actors.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

A. Discussion Quality Features

Discussion quality features have been widely studied in past research to scrutinize the members of virtual communities. In particular, previous work has studied how user discussion quality can affect social behaviors and interactions. We observe two related features within this study:

1. *Average_Message_Length* – This feature is simply the average length of an individual hacker's messages. Past research has used average message length to determine the amount of content a user generally contributes within a single post, with the assumption that longer posts are generally more valuable to the cognitive advance of a community [21].
2. *Number_Of_Replies_Per_Thread* - The average number of replies a user posts within each discussion participated in. This feature has previously been used in research to measure the amount of dialogue a user contributes to a particular discussion. Users who contribute more dialogue to a discussion are thought to have a greater impact on the cognitive advance of a community [19].

B. Community Involvement Features

Along with observing the styles in which users communicate, previous research has also investigated the implications that differing levels of community involvement have on user reputation. Specifically, the breadth of involvement and the type of involvement have typically been considered important features used for analysis in social media analytics. We observe four features related to community involvement in this study:

1. *Number_Of_Threads_Involved* - User expertise is closely related to reputation in online communities; expert users are often involved in many discussions and are able to discuss various concepts and topics [20]. Users who post in many different threads take the initiative to often express their opinions and share their knowledge.
2. *Tenure* - Previous investigation has found that member tenure can amplify their reputation and the trust that others place into them [10].We measure tenure by observing the date that each hacker posted their first message within their community, as it marks their initial participation. To be able to quantify tenure and compare hackers, we use days as our unit of measurement.
3. *Sum_Of_Attachments* - The number of times a hacker attaches a file or code example to their post, a measure of cognitive advancement towards a hacker's community [14].
4. *Total_Messages* - Total number of messages a hacker has posted, relevant as previously highlighted literature states reputation may be a result of higher levels of activity [14].

C. Regression Model

$$\begin{aligned} Reputation = {} & \beta_1 Average\_Message\_Length \\ & + \beta_2 Number\_Of\_Replies \\ & + \beta_3 Number\_Of\_Threads\_Involved \\ & + \beta_4 Tenure \\ & + \beta_5 Sum\_Of\_Attachments \\ & + \beta_6 Total\_Messages \\ & + \varepsilon \end{aligned}$$

Regression analysis was performed on both forums. Only users with at least 1 reputation point were included in analysis in order to reduce statistical skewness that would result from members who join a community for only very briefly before becoming inactive. Additionally, in some instances, users were given negative reputations as a result of harmful or offensive postings. For example, one hacker was caught distributing a tool that contained a hidden level of malware used to infect other hackers. These users were excluded to more accurately depict the mechanisms in which reputation is gained. Lastly some community administrators disqualified from analysis as they were witnessed to assign themselves reputation points arbitrarily, which would skew results.

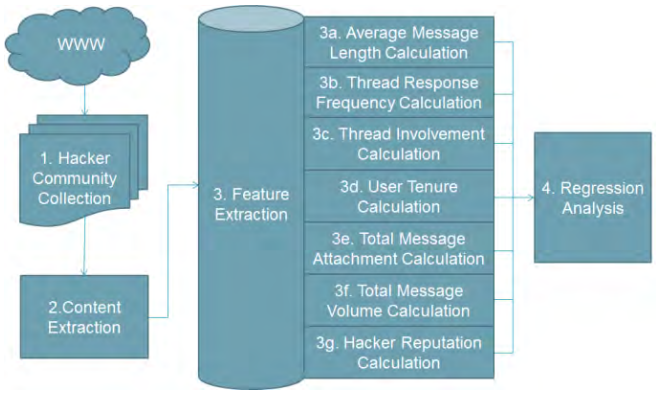


Figure 2. The research design, visualized. Two hacker communities are downloaded from the Internet and their contents extracted into a SQL database. Features identified through literature review are selected and calculated using forum contents. Finally, a regression analysis is run to identify relationships between features and hacker reputation.

Our first set of hypotheses delineates the relationship between a hacker's discussion quality and their reputation. Specifically, we observe how much content a hacker typically includes in a single message, as well as the amount of dialogue a hacker contributes to one discussion. Discussion quality has been previously been shown to affect social behaviors and interactions:

H1: Discussion quality will be a significant contributor towards hacker reputation as it relates to the cognitive advance of a community.

H1a: The average message length of a hacker will be a significant contributor of a hacker's reputation, as past research has highlighted that average message length is related to user discussion quality.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

H1b: The number of replies a hacker posts will affect their reputation. Number of replies has previously been used in research to measure the amount of dialogue a user contributes to a particular discussion.

Our second set of hypotheses is related to the relationship between a hacker's community involvement and their reputation among peers. Community involvement encompasses a hacker's activity level and their contributions to the cognitive advance of the community. Both have been previously found to influence user reputation within virtual communities:

H2: Community involvement will be a significant contributor influencer of hacker reputation as it encompasses user activity levels and contributions towards the cognitive advancement of the community.

H2a: The number of threads a hacker is involved in will impact their reputation; previous research has demonstrated that the ability to discuss a wide variety of topics and concepts is related to expertise.

H2b: The seniority of a hacker will be a significant contributor of a hacker's reputation, as previous literature states tenure generally amplifies an individual's expertise and trustworthiness.

H2c: The sum of attachments a hacker posts throughout their community activities will be an influencer of reputation as hackers are contributing to the cognitive advancement of their community.

H2d: Total message volume will be a significant contributor towards hacker reputation, as activity level is a previously observed contributor towards reputation.

Lastly, we expect to observe similar results in both hacking communities. Previous research exploring hacking communities and cybercriminal activity from the United States and China have unveiled similarities between hackers from both countries; for example, observations of hacker communities across multiple geopolitical regions has found black market trading to be universally popular [2,20].

H3: Results of the regression model will be consistent across both the English and Chinese forums, as both forums share aspects of hacker culture.

VI. RESULTS AND DISCUSSION

TABLE II. HACKHOUND.ORG RESULTS

Feature	Estimate	Std. Error	T Value
Average_Message_Length	-0.0083	0.0025	-0.968
Number_Of_Replies_Per_Thread	0.0188	0.0616	0.305
Number_Of_Threads_Involved	0.1689	0.0538	2.822***
Tenure	0.0041	0.0123	0.526
Sum_Of_Attachments	0.2786	0.1437	5.323***
Total_Messages	0.3396	0/0379	6.55 ***

Adjusted R-squared: 83.14%; ***p ≤ 0.001 ** p ≤ 0.01 *p ≤ 0.05)

TABLE III. UNPACK.CN RESULTS

Feature	Estimate	Std. Error	T Value
Average_Message_Length	0.0052	0.0027	0.125
Number_Of_Replies_Per_Thread	0.0372	0.0040	0.528
Number_Of_Threads_Involved	0.1403	0.0033	1.914*
Tenure	-0.0086	0.0135	-0.144
Sum_Of_Attachments	0.3805	0.1991	4.757***
Total_Messages	0.2838	0.0252	3.714**

Adjusted R-squared: 57.38 %; ***p ≤ 0.001 ** p ≤ 0.01 *p ≤ 0.05)

Involvement in various threads, the sum of attachments, and total message volume all appear to be significant contributors of reputation, supporting H2a, H2c, and H2d. This demonstrates support for theories tying higher user reputations to individuals who are active and contribute to cognitive advancement of their communities. Correlations are stronger in Hackhound.org, perhaps due to forum age; as Hackhound.org is younger and has a much lighter total message volume than U npack.cn, each message on Hackhound.org holds more potential weight for contributing towards a user's reputation. In hacker communities, knowledge appears to be power in the form of reputation and influence

Average post length, number of replies per thread, and tenure do not appear to be significant contributors towards reputation, thus not supporting H1a, H1b, and H2b. Discussion quality and seniority seem to be mostly irrelevant when considering reputation in the context of hacker communities. Results suggest reputation may be more heavily influenced by the content and diversity or novelty of information posted, rather than how information is delivered or who presents it.

Both communities share similar patterns in regards to how reputation is built by members, supporting H3. Despite obvious cultural differences between the English and Chinese forums, an overriding form of hacker culture appears to be experienced by both communities. A slight variation in R-squared is experienced when performing analyses on these two forums. The analysis of Hackhound.org has a higher R-squared than that of the Unpack.cn forum. Unpack.cn is a much older community and thus may have an accumulation of more variation among data over time. The differences in R-squared may also be related to cultural phenomena not yet identified.

VII. CONCLUSION AND CONTRIBUTIONS

As the computer becomes more ubiquitous within society, the security of networks and information technologies is a growing concern. Critical infrastructures and military interests are at an increasing risk. Tools to commit cybercrime are now more accessible and easy to use than any previous point in the Internet's history. In particular, key actors responsible for leading hacker communities and cybercriminal groups are of interest.

In this research, two major hacker communities from the United States and China are examined to identify the mechanisms in which key actors arise. These communities were selected due to unique characteristics of allowing the attachment of tools and code to messages, as well as a peer-

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

evaluated reputation system. Relationships between message volume, message lengths, community seniority, the inclusion of tools or code in messages, and reputation were studied. Hackers that contributed to cognitive advance of their community or were considerably active had the highest reputations.

Our experiment contributes to research in social media analytics, as results revealed correlations between user reputation and various user behaviors and social media features within web forums. Specifically, we confirm previous findings features reported to interact with increasing reputation. This study was unique from other social media research because it investigated the mechanisms built into hacker communities that allow users to share code and files as part of their message postings.

Further, this research contributes to advancing our understanding of hacker communities. Security researchers would benefit from knowledge of the processes hackers use to gain status, power, and leadership within their communities. Law enforcement would benefit by more efficiently focusing resources on cybercriminal pursuits. If key actors contributing towards the cognitive advance of their communities or groups are apprehended, novice hackers relying on such individuals could not as easily develop malicious capabilities. Lastly, the results of this study can also be used to identify key actors in other hacker forums where reputation-ranking systems are not provided.

This research can be expanded by additional investigation of hacker communities. Analytical tools developed from control theory can be used to provide additional scrutiny when attempting to identify driver nodes within communities of different types and sizes [17]. These tools are important when attempting to study hacker communities with unknown origins and context. Further, the perspectives of other social network theories can be used to extend current research on hacker communities, potentially revealing interesting phenomena. Deeper exploration is necessary to more fully understand hacker behaviors and methodologies.

REFERENCES

[1] T. Moore and R. Clayton, "Evil searching: compromise and recompromise of Internet hosts for phishing," in Financial Cryptography and Data Security, pp. 256-272, 2009.

[2] J. Radianti and J.J. Gonzalez, "A preliminary model of the vulnerability black market," Society, 2007.

[3] R. McCusker, "Transnational organised cyber crime: distinguishing threat from reality," in Crime, Law and Social Change, vol. 46, no. 4-5, pp. 257-273, December 2006.

[4] N. Kshetri, "The Simple Economics of Cybercrime," in IEEE Security & Privacy, vol. 4, no.1, pp. 33-39, January-February 2006.

[5] T. Holz, M. Engelberth, and F. Freiling, "Learning more about the underground economy: a case study of keyloggers and dropzones," in European Symposium on Research in Computer Security (ESORICS), 2009.

[6] D. Wall, "Cybercrime: the transformation of crime in the information age," Polity, September 2007.

[7] C. J. Mielke and H. Chen, "Botnet and the cybercriminal underground," in IEEE International Conference on Intelligence and Security Informatics 2008, pp. 206-211. June 2008.

[8] S. Gosh and E. Turrini, "Cybercrimes: a multidisciplinary analysis," Springer, October 2010.

[9] D. Semmans, H. Krambeck, and M. Milinksi, "Reputation is valuable within and outside one's own social group," in Behavioral Ecology and Sociobiology, vol.57, no. 6, April 2005.

[10] R. Lines, M. Selart, B. Espedal, S. Johansen, and T. Svein, "The production of trust during organizational change," in Journal of Change Management, vol. 5, no. 2, February 2005.

[11] K. V. Lampe and P. O. Johansen, "Organized crime and trust: on the conceptualization and empirical relevance of trust in the context of criminal networks," in Global Crime, vol. 6. no. 2, 2004.

[12] A. T. Hall, F. R. Blass, G. R. Ferris, and R. Massengale, "Leader reputation and accountability in organizations: implications for dysfunctional leader behavior," in The Leadership Quarterly, vol. 15, no. 4, August, 2004.

[13] S. J. Zaccaro, C. Kemp, and P. Bader, "Leader traits and attributes," in The Nature of Leadership, pp. 101-123, 2004.

[14] P. Muller, "Reputation, trust and the dynamics of leadership in communities of practice," in Journal of Management and Governance. Vol. 10, no. 4, November, 2006.

[15] R. E. Kalman, "Mathematical description of linear dynamical systems," in Society for Industrial and Applied Mathematics, 1963.

[16] D. G. Luenberger, "Introduction to dynamic systems: theory, models, & applications," Wiley, May 1979.

[17] Y.Y. Liu, J.J. Slotine, and A.L. Barabási, "Controllability of complex networks," in Nature, no. 473, May 2011.

[18] J. Xu, and H. Chen, "Criminal network analysis and visualization: a data mining perspective," in Communications of the ACM, vol. 48, no. 6, 2005.

[19] J. Kim, E. Shaw, D. Feng, C. Beal, and E. Hovy, "Modeling and assessing student activities in on-line discussions," in Proceedings of the Workshop on Educational Data Mining, 2006.

[20] J. Zhang, M. S. Ackerman, and L. A. Adamic, "Expertise networks in online communities: structure and algorithms," in World Wide Web Conference., May 2007.

[21] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and Yahoo Answers: everyone knows something," in World Wide Conference, April 2008.

[22] O. Nov, M. Naaman, and C. Ye, "Motivational, structural and tenure factors that impact online community photo sharing," in Proceedings of AAAI International Conference on Weblogs and Social Media, May 2009.

[23] K. Chai, C. Wu, V. Potdar, and P. Hayati, "Automatically measuring the quality of user generated content in forums," in Proceedings of Australasian Conference on Artificial Intelligence, pp. 51-60, 2011.

[24] T. J. Holt, E. Lampke, "Exploring stolen data markets online: products and market forces," in Criminal Justice Studies, vol. 23, no. 1, March 2010.

[25] A. Mehra, A. L. Dixon, D. J. Brass, and B. Robertson, "The social network ties of group leaders; implications for group performance and leader reputation," in Organization Science, vol. 17, no. 1, February 2006.

[26] G.Y. Huang, S.Y. Hu, and J.R. Jiang, "Scalable reputation management with trustworthy user selection for P2P MMOGs," in International Journal of Advanced Media and Communication, vol. 2, no. 4, 2008.

[27] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha, "Learning to recognize reliable users and content in social media with coupled mutual reinforcement," in World Wide Web Conference, May 2009.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:07:16 UTC from IEEE Xplore. Restrictions apply.

MARC View

LEADER 02582nam a2200337 i 4500				
001 12138452				
003 SIRSI				
005 20140415021552.0				
006 m d				
007 cr n				
008 140228s2010 nyu||||s||||1||||eng d				
020				a| 9781450300551 (print)
040				a| WaSeSS b| eng c| WaSeSS d| WaSeSS
050	1	4		a| QA76.9.D343
111	2			a| International Conference on Knowledge Discovery & Data Mining n| (16th : d| 2010 : c| Washington, D.C.)
245	1	0		a| KDD '10 h| [electronic resource] : b| proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining : June 25-28, 2010, Washington, DC, USA / c| sponsored by ACM SIGKDD & SIGMOD.
246	1			a| Knowledge Discovery and Data Mining 2010 : b| proceedings of the 16th Association for Computing Machinery Special Interest Group on Management of Data International Conference on Knowledge Discovery and Data Mining : June 25-28, 2010, Washington, District of Columbia, United States of America
246	1			a| Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
246	1			a| KDD '10 the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Washington DC, DC, USA - July 25 - 28, 2010
260				a| New York : b| ACM, c| 2010.
300				a| 1 online resource (1214 pages)
506				a| License restrictions may limit access.
588				a| Description based on: online resource; title from pdf title page (ACM Digital Library, viewed Feb. 12, 2014).
650		0		a| Data mining v| Congresses.
650		0		a| Database management v| Congresses.
655		0		a| Electronic books.
710	2			a| Association for Computing Machinery. b| Special Interest Group on Knowledge Discovery & Data Mining.
710	2			a| Association for Computing Machinery. b| Special Interest Group on Management of Data.
776	0	8		i| Print version: a| International Conference on Knowledge Discovery & Data Mining (16th : 2010 : Washington, D.C.) t| KDD '10 : proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining : June 25-28, 2010, Washington, DC, USA. d| New York : ACM, 2010 z| 9781450300551
856	4	0		u| http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001139208&title=KDD%20%2710%20%3A%20proceedings%20of%20the%2016th%20ACM%20SIGKDD%20International%20Conference%20on%20Knowledge%20Discovery%20and%20Data%20Mining%20%3A%20June%2025-28%2C%202010%2C%20Washington%2C%20DC%2C%20USA
949				a| Electronic resource w| ASIS c| 1 i| 12138452-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 4/16/2014 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState
Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.
Legal Statements | PSU Hotlines
PENN STATE UNIVERSITY LIBRARIES
Libraries Home (814) 865-6364
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory
CONNECT WITH PENN STATE UNIVERSITY LIBRARIES
Facebook
Twitter
Instagram

KDD '10 [electronic resource] : proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining : June 25-28, 2010, Washington, DC, USA / sponsored by ACM SIGKDD & SIGMOD

ASK A LIBRARIAN

Conference Author	International Conference on Knowledge Discovery & Data Mining (16th : 2010 : Washington, D.C.)
Additional Titles	Knowledge Discovery and Data Mining 2010 : proceedings of the 16th Association for Computing Machinery Special Interest Group on Management of Data International Conference on Knowledge Discovery and Data Mining : June 25-28, 2010, Washington, District of Columbia, United States of America, Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, and KDD '10 the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Washington DC, DC, USA - July 25 - 28, 2010
Published	New York : ACM, 2010.
Physical Description	1 online resource (1214 pages)
Additional Creators	Association for Computing Machinery. Special Interest Group on Knowledge Discovery & Data Mining and Association for Computing Machinery. Special Interest Group on Management of Data

Access Online

serialssolutions.com

Restrictions on Access

License restrictions may limit access.

Subject(s)

Data mining—Congresses

Database management—Congresses

Genre(s)

Electronic books

ISBN

9781450300551 (print)

View MARC record | catkey: 12138452

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368

Libraries Intranet (Staff Only)

Accessibility Help

Website Feedback

Policies and Guidelines

Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook

Twitter

Instagram

Appendix BOZORGI01



Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits

Mehran Bozorgi, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker
Department of Computer Science and Engineering
University of California, San Diego
mehranbozorgi@google.com, {saul,savage,voelker}@cs.ucsd.edu

ABSTRACT

The security demands on modern system administration are enormous and getting worse. Chief among these demands, administrators must monitor the continual ongoing disclosure of software vulnerabilities that have the potential to compromise their systems in some way. Such vulnerabilities include buffer overflow errors, improperly validated inputs, and other unanticipated attack modalities. In 2008, over 7,400 new vulnerabilities were disclosed—well over 100 per week. While no enterprise is affected by all of these disclosures, administrators commonly face many outstanding vulnerabilities across the software systems they manage. Vulnerabilities can be addressed by patches, reconfigurations, and other workarounds; however, these actions may incur down-time or unforeseen side-effects. Thus, a key question for systems administrators is which vulnerabilities to prioritize. From publicly available databases that document past vulnerabilities, we show how to train classifiers that predict whether and how soon a vulnerability is likely to be exploited. As input, our classifiers operate on high dimensional feature vectors that we extract from the text fields, time stamps, cross-references, and other entries in existing vulnerability disclosure reports. Compared to current industry-standard heuristics based on expert knowledge and static formulas, our classifiers predict much more accurately whether and how soon individual vulnerabilities are likely to be exploited.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; I.5.1 [**Pattern Recognition**]: Models—*Statistical*; I.5.2 [**Pattern Recognition**]: Design Methodology—*Feature evaluation and selection*

General Terms

Algorithms, Security

Keywords

supervised learning, SVM, vulnerabilities, exploits

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
KDD'10, July 25–28, 2010, Washington, DC, USA.
Copyright 2010 ACM 978-1-4503-0055-1/10/07 ...$10.00.

1. INTRODUCTION

Among the many requests made of researchers in computer security, few are as frequent or as urgent as the call for meaningful security metrics. The requests are driven by a widespread need to quantify security risks ("how likely is it that an attacker will thwart my security measures?") in a way that informs operational policy choices. Unfortunately, the adversarial nature of security has resisted traditional methods of quantifying risk and has even led some to argue that such metrics are inherently unattainable [4]. Nevertheless, even absent a comprehensive solution to this conundrum, there remains a need to evaluate the value of distinct operational security choices. Thus, a range of ad hoc approaches have emerged in individual domains where the needs are particularly acute. In this paper we focus on one such domain — evaluating vulnerability disclosures — and we show that it is possible to make meaningful predictions using tools from data mining and machine learning.

Public vulnerability disclosure has long been a staple of the software security industry, with many thousands of new software vulnerabilities identified and publicized each year [11]. In turn, these vulnerabilities are communicated, via a variety of channels, to system administrators who must then determine if they have susceptible systems and decide what action to take if so. Unfortunately, patching and other mitigations can incur significant manpower overheads (even more so for mission critical services that require quality assurance testing before deploying new software). Since few organizations have the resources to address every vulnerability disclosure that might impact their enterprise, administrators must prioritize their efforts, triaging the most critical vulnerabilities to address first.

To inform these decisions, a variety of "vulnerability scoring" frameworks have been designed to assess risk qualitatively (e.g., Microsoft's critical, important, moderate, or low severity rating) or quantitatively (e.g., US-CERT's severity metric). Indeed, one such framework — FIRST's Common Vulnerability Scoring System (CVSS) [9] — appears to be emerging as the *de facto* standard in the community. The use of CVSS is mandated in the Payment Card Industry's Data Security Standard (PCI-DSS), it is the ranking used in NIST's National Vulnerability Database (NVD), and its use is recommended by a wide range of computer, networking and software vendors (e.g., Cisco's Risk Triage whitepaper [5]).

However, while these systems are carefully designed using expert knowledge, they are inherently ad hoc in nature. For example, the CVSS (v2) overall "Base Score" is expressed in terms of Impact (I) and Exploitability (E) components by:

$$\text{BaseScore} = (1.176) * \left(\frac{3I}{5} + \frac{2E}{5} - \frac{3}{2}\right). \qquad (1)$$

By convention, this score is rounded to one decimal place, and it

Appendix BOZORGI02

is set to zero (regardless of the above formula) if $I = 0$. We note further that the Impact and Exploitability components in eq. (1) are themselves combinations of categorical magic numbers. (For example, the Exploitability component depends on an Access Vector score which takes the value 0.395 if the vulnerability requires local access, 0.646 if the vulnerability requires adjacent network access, and 1.0 if the vulnerability global network access). While we have little doubt that these scoring metrics were carefully considered and of great value when first developed, we suspect that any single fixed equation, such as eq. (1), is unlikely to provide a robust and lasting model of vulnerability severity.

To this end, our paper seeks to place this problem on a more systematic footing. Using tools from machine learning, we show how to train classifiers that predict whether vulnerabilities are likely to be exploited, and if so, how soon. Our results suggest that our trained classifiers are likely to outperform current measures of exploitability. In particular, our classifier outputs correlate much better with vulnerability outcomes than the "Exploitability" component of the CVSS Base Score in eq. (1).

2. BACKGROUND

Software vulnerabilities are exploitable flaws in software systems that pose significant security risks. Production software inevitably ships with many such flaws, a subset of which are subsequently discovered and become known over time. When flaws are discovered, vendors distribute patches and mitigations to their customers, who ideally implement such measures before an *exploit* is developed and targeted against them. This vulnerability lifecycle (described in more detail by Arbaugh *et al.* [1]) has in turn driven the creation of a complex ecosystem of players: vulnerability researchers, software and security vendors, security information providers and a range of networks, information sources and markets connecting them together (for a comprehensive analysis of the vulnerability ecosystem see Frei *et al.* [10]).

2.1 Public Vulnerability Disclosures

At the end of this process, vulnerabilities are documented and disclosed to the public. These reports not only list various discrete attributes of each vulnerability (e.g., software affected, date disclosed), they also describe (in plain text) how each vulnerability works, why it presents a threat, and how it can be mitigated. This information is disclosed to the public via multiple sources, including moderated forums (e.g., Bugtraq [21]), individual vendors (e.g., Microsoft [14], commercial aggregators (e.g., Secunia [20]), and open source databases (e.g., OSVDB [17]). Vulnerabilities are also quickly assigned a unique identifier — both local to individual information providers/repositories and global across multiple vulnerability databases using MITRE's Common Vulnerabilities and Exposures (CVE) service [6].

The precise timing of vulnerability disclosures depends considerably on how they were found and the policy of the organizations involved. Some vulnerabilities are disclosed immediately upon discovery while others may be kept private for significant periods of time to allow vendors to develop and test appropriate patches and mitigations. Still other vulnerabilities are exploited before or at the same time as public disclosure (so-called 0-day exploits). Vulnerability discovery and disclosure policy has generated a great deal of both controversy and research [2, 3, 18, 16]. A number of studies have also examined the probability that vulnerabilities are able to be patched [15, 19]. By contrast, our work focuses on predicting if a vulnerability is likely to be exploited shortly (thus meriting immediate attention from system administrators).

2.2 Rating Vulnerabilities

To aid system administrators, vulnerability disclosures typically include a qualitative or quantitative assessment of each vulnerability's severity. The overall severity score depends on both *impact* (how significant are the consequences of exploitation) and *exploitability* (how difficult is the vulnerability to exploit). Severity scores are derived primarily from expert knowledge and/or communal input. For example, US-CERT generates a quantitative severity score ranging from 0 to 180, calculated directly from answers to a range of qualitative questions (e.g., "Is information about the vulnerability widely available or known?" and "What are the preconditions required to exploit the vulnerability?") [7]. Microsoft's Security Bulletin documents vulnerability severity using a qualitative scheme (critical, important, moderate, or low) as do Secunia's reports (extremely critical, highly critical, moderately critical, less critical, or not critical). More recently, a group of vendors and researchers came together, under the sponsorship of the Forum of Incident Response and Security Teams (FIRST), to create a new severity metric, the Common Vulnerability Scoring System (CVSS). Now in its second iteration, CVSS defines several independent metrics, but it is the "base metric" which is typically used in third-party vulnerability databases. This score combines impact and exploitability components according to a carefully designed formula [9].

Unfortunately, each of these systems measures different things and weights them in different ways. We are unaware of any empirical study evaluating the effectiveness of any of these metrics or comparing them to one another. Thus, it is hard to make concrete statements about which approach is best, or why. Indeed, this problem is inherently difficult since some aspects of "severity" are either context dependent (e.g., a mission critical server being shut down may be more "severe" than a print server) or may be inherently difficult to quantify. However, the issue of exploitation is far more clear cut — a vulnerability is either exploited or not and the date upon which a working exploit becomes known is frequently documented. Thus, in this paper we focus on the exploitability aspect of severity.

Given this scope, we argue that existing scoring systems are probably too limited to offer strong predictive power. They include only a few factors in each vulnerability's assessment—which may not be the key distinguishing features and frequently depend on the judgment of evaluators—and they combine these features in the same way to produce a score for widely different sorts of vulnerabilities. For example, the current CVSS Exploitability score is calculated as follows [13]:

$$\text{Exploitability} = 20 * AV * AC * \text{Authentication} \quad (2)$$

where AV stands for `Access Vector` and AC stands for `Access Complexity`, and each of these variables are assigned particular fixed values based on other qualitative or subjective evaluations. For example, AC is set to 0.35 if access complexity is deemed to be "high", 0.61 if "medium" and 0.71 if "low". It is not entirely clear how this formula or its constants were designed. Moreover, it seems unlikely that this simple formula can model the probability of exploitation across many different sorts of vulnerabilities.

3. VULNERABILITY DATA

In this study, we use two well-known, online sources of vulnerability data, the Open Source Vulnerability Database (OSVDB) [17] and the MITRE Common Vulnerabilities and Exposures (CVE) database [6].

Appendix BOZORGI02

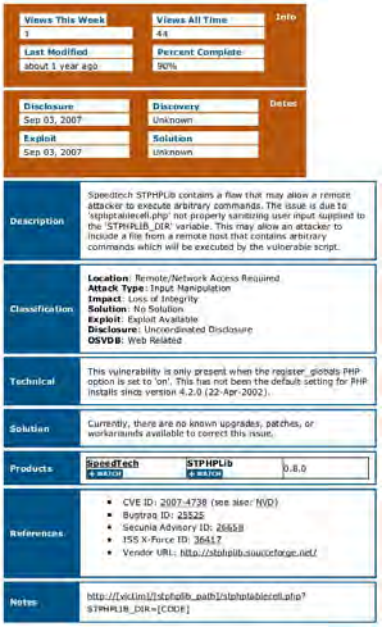


Figure 1: An example OSVDB vulnerability report.

Exploit Category	# Vulnerabilities	Label
Exploit Available	8,537	Positive
Exploit Rumored / Private	1,483	Positive
Exploit Unavailable	536	Negative
Exploit Unknown	3,209	Negative
No Category	999	Not Used
Total	14,764	

Table 1: Categories of exploited vulnerabilities.

OSVDB is a large database containing reports on over 57,000 vulnerabilities. As an example, figure 1 shows the OSVDB report for a vulnerability in a Web services library. These reports contain a wealth of information about each vulnerability, indexed using a unique OSVDB ID, including a detailed description, technical details, the software products affected, solutions (such as patches and mitigations to prevent exploitation), and references to other sources of information about the vulnerability. The reports also include classification information, such as the type of attack required to exploit the vulnerability, the origins from which an attack can be launched, and the manner in which the vulnerability was disclosed. Section 4.1 describes how we extract information from these reports as features for classification and prediction.

The reports also provide temporal information for each vulnerability, such as the dates of discovery, disclosure, and first known exploits. We augment this data with additional temporal information provided by Frei *et al.* as described in their WEIS 2009 paper [10]. As described below, we use this additional information to create labeled training and test sets.

From the OSVDB database we extracted a large set of vulnerabilities for classification and prediction. We used only vulnerabilities that were disclosed during the years 1991–2007, inclusive. Vulnerabilities before 1991 represent a different era of software; we excluded later vulnerabilities because, at the time we started the project, they were recent and still in flux (e.g., many of them had undetermined outcomes). We also excluded vulnerabilities that did not have a description.

Table 1 shows the number of vulnerabilities we use in our experiments and how we label them. It categorizes vulnerabilities using their OSVDB "Exploit Classification" status. If a vulnerability has an available, rumored, or private exploit, we label it as a "positive" vulnerability, indicating that it has been exploited. Similarly, if a vulnerability has no known exploits or exploits are unavailable, we label it a "negative" vulnerability indicating that it is not exploited. If a vulnerability report does not classify its exploit status, we exclude it from consideration since we cannot determine the accuracy of our predictions. Section 4.2 describes how we train a classifier from these labeled examples of vulnerabilities.

We use the CVE database to augment the vulnerability reports from the OSVDB database. Similar to OSVDB, CVE entries include summaries, references to related products and reports, information about the type of vulnerability, time stamps, and severity scores. In addition to providing more information that can be extracted as features for classification, for some vulnerabilities the CVE entries also provides information missing in the OSVDB reports. We integrate these records by cross-referencing their CVE and OSVDB identifiers. Most OSVDB reports reference the corresponding CVE reports for the same vulnerability and conversely, some CVE entries have corresponding OSVDB IDs as well.

Finally, we note that the quality of our results are inherently tied to the quality of this disclosure data and in particular the quality of the temporal labels (when a vulnerability was disclosed and exploited). This creates two potential classes of problems. In principle, there are adversarial training risks since bad vulnerability data could influence what the classifier learns during training. However, we believe this is a particularly unlikely scenario since vulnerability databases are generated by large numbers of independent actors. It seems unlikely that an adversary would discover and disclose enough new vulnerabilities (in turn validated and accepted by third parties) to influence the overall feature set used in training. Similarly, while an adversary might try to "game" our predictions (e.g., by only exploiting vulnerabilities which we had classified as unlikely to be exploited), the risk seems low, and certainly such a counterstrategy is no easier than it is under current vulnerability scoring systems. A somewhat more likely limitation is systematic bias. In particular, we note that large numbers of vulnerabilities in the complete database have unknown exploitation status or dates, which limits our ability to train on these records. In this work, we assume that the remainder of disclosures (with known status and dates) are representative and accurate. A selection bias would emerge if the omitted records were systematically different than complete records; however, we do not believe such a bias exists.

4. MACHINE LEARNING FOR VULNERABILITY CLASSIFICATION

We aim to improve on existing approaches by casting vulnerability classification as a problem in machine learning. In a nutshell, our goal is to replace small-scale heuristics by large-scale statistics. This section describes our statistical model for vulnerability classification. The model is estimated from a large database of vulnerabilities that have been labeled as "exploited" or "not exploited". Section 4.1 describes how we extract information from this database and distill it into feature vectors for classification, and section 4.2 describes how we classify these feature vectors using support vector machines [22]. The training and test sets of feature vectors in our experiments are available at `http://www.sysnet.ucsd.edu/projects/exploit-learn/`.

Appendix BOZORGI02

Feature Family	Count	Database Source
Summary (B)	14883	CVE
Full Product Name (B)	13040	OSV - Obj. Correls.
Description (B)	11573	OSV - Vulnerabilities
Title (B)	9812	OSV - Vulnerabilities
Short Description (B)	9761	OSV - Vulnerabilities
Manual Notes (B)	6576	OSV - Vulnerabilities
Product Versions (B)	5388	OSV - Obj. Versions
Related Products (B)	5057	CVE
Product Names (B)	3661	OSV - Obj. Products
Tech. Description (B)	3479	OSV - Vulnerabilities
Solution (B)	3474	OSV - Vulnerabilities
Product Vendors (B)	2500	OSV - Obj. Vendors
Authors (B)	2368	OSV - Credits
Keywords (B)	1556	OSV - Online
References (B)	267	CVE
Classifications (B)	69	CVE
External Refs (B)	31	OSV - Ext. Refs.
OSVDB Dates (I)	15	OSV - Vulnerabilities
Attack Type (B)	11	OSV - Classifications
Category (B)	9	OSV - Classifications
Location (B)	8	OSV - Classifications
Solution Category (B)	8	OSV - Classifications
Disclosure Type (B)	8	OSV - Classifications
CVE Dates (I)	6	CVE
Impact (B)	4	OSV - Classifications
Scores (I)	3	CVE
Effect on Products (B)	3	OSV - Aff. Types
Other (I)	8	OSV & CVE
Total	93578	

Table 2: Extracted features from the vulnerability data. (B) denotes binary and (I) denotes integer features.

4.1 Feature extraction

Our database of vulnerabilities contains a wealth of information, both factual and textual, about their histories and distinguishing characteristics. For each vulnerability, we extract a high-dimensional ($d = 93578$) feature vector of binary and integer-valued features. Though many of these features will ultimately turn out to be irrelevant or redundant for classification, the goal of our feature extraction is to distill as much information as possible for subsequent statistical analysis.

Much of our information about vulnerabilities is contained in text fields. We derive binary features using a bag-of-words representation for each text field [12]. Essentially, these features record whether or not particular tokens (e.g., "buffer", "heap", "DNS") appear in specific text fields (e.g., "title", "solution", "product name") associated with each vulnerability.

Table 2 shows the breakdown of features that we extract for each vulnerability in our database. Each row in the table indicates the number of features derived from a particular type of information. Most of the features are generated from bag-of-words representations of text fields. However, integer-valued features also encode useful information, such as the date when a vulnerability was first disclosed, the length of text describing its symptoms, or the ranking of its severity according to other popular heuristics.

4.2 Large margin classification

We build classifiers by training linear support vector machines (SVMs) [22] on the feature vectors described in the previous section. (As preprocessing, however, the non-binary features are normalized to lie between zero and one so that they do not overshadow the binary features.) Linear SVMs are trained by computing the maximum margin hyperplane that separates the positive and negative examples in feature space. The decision rule mapping feature vectors $\mathbf{x} \in \Re^d$ to labels $y \in \{-1, +1\}$ is given by:

$$y = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b), \tag{3}$$

where $\mathbf{w} \in \Re^d$ is the normal (weight) vector to the separating hyperplane and b is the distance of the separating hyperplane from the origin.

Linear SVMs are particularly appropriate for our application to vulnerability classification because we have many more input features (d) than training examples (n). In particular, for the experiments in section 5, the ratio of features to examples is never less than 10-to-1. In this regime of small sample size ($n \ll d$), there are many hyperplane decision boundaries that can perfectly separate all n examples $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ in our training sets. Linear SVMs compute the hyperplane that (roughly speaking) maximizes the distance of the most borderline training examples to the linear decision boundary. This hyperplane is not only uniquely specified, but a large body of work in statistical learning theory also shows that it generalizes better to new data, yielding lower expected error rates when used to classify previously unseen examples [22]. Many software packages are available for fitting models of this form; for the results in this paper, we used the LIBLINEAR implementation of SVMs [8].

5. EVALUATION

In this section we present our experimental results using SVMs for vulnerability classification. We consider several different scenarios. We first evaluate the prediction accuracy in an offline experiment, representing a best-case scenario where we consider the data set of vulnerabilities as a single, static snapshot; we also examine the features that have the most prominent role in making predictions. We then evaluate the prediction accuracy of our model in an online experiment emulating a real-world deployment: we dynamically update the classifier and make predictions over time as new vulnerabilities appear. We also use SVMs to predict if vulnerabilities will be exploited within a particular time frame. Finally, we compare the results from SVMs to current heuristic approaches to vulnerability classification.

5.1 Methodology

As discussed in Section 3, we use the OSVDB and CVE databases as our data set of vulnerability examples. In addition to providing the features we use for learning and classification, they also provide the ground truth for evaluating the accuracy of our classifiers. In general, we label those vulnerabilities that have exploits as positive examples and vulnerabilities that do not have exploits as negative examples. Table 1 shows a breakdown of the vulnerabilities based on these labels.

Note that there are more positive examples (10,020) than negative examples (3,745), i.e., more vulnerabilities with exploits than those without. When conducting balanced experiments, which remove any such bias in the data used for classification, we randomly choose the same number of examples from both sets multiple times and average the results. When conducting unbalanced experiments, an unavoidable aspect of practical deployments, we explicitly quantify and report the bias in the input data. Finally, a subset of the vulnerabilities do not have exploit information; we do not use these examples in our experiments because, without true labels, we cannot evaluate the accuracy of classification.

Appendix BOZORGI02

	Training	Testing
Positive Examples ($\|P\|$)	1600	2000
Negative Examples ($\|N\|$)	1500	1874
Total Examples	3100	3874
True Negatives	100%	92.2%
True Positives	100%	87.5%
False Negatives	0%	7.79%
False Positives	0%	12.5%
Total Accuracy	100%	89.8%

Table 3: Prediction accuracy in the offline experiment.

5.2 Offline Exploit Prediction

In our first experiment, we evaluate how well SVMs classify vulnerabilities in an offline setting. For this experiment, we use a balanced data set of roughly 4000 positive and negative examples—that is, divided almost evenly between vulnerabilities with and without exploits. We train SVMs on 40% of these examples (as training data) and evaluate them on 50% of these examples (as test data). We use the remaining 10% of examples as a development set to choose tuning parameters for the SVMs. We report averaged results from ten-fold cross validation: that is, we learn ten different classifiers, randomly choosing which examples fall into the training, test, and development sets, then average the results across all runs.

Table 3 shows the results. Here, *true positives* are positive examples correctly classified as vulnerabilities that will be exploited; *true negatives* are negative examples correctly classified as vulnerabilities with no known exploits; *false positives* are positive examples incorrectly classified as vulnerabilities that will be exploited, but were not; and *false negatives* are negative examples that were incorrectly classified as vulnerabilities that have no known exploits, but in fact do. The overall accuracy is nearly 90%, demonstrating the viability of statistical methods for vulnerability classification.

5.3 Feature Inspection

The offline classification results show that the vulnerability reports contain useful features for predicting whether a vulnerability will be exploited. We now examine which features play a prominent role in these predictions. In the linear SVMs that we use, the decision rule in eq. (3) multiplies each feature by a positive or negative weight. Recall that all features are normalized to lie between zero and one before the weights are learned. Thus the magnitudes of these weights reflect the relative contribution of each feature to the decision rule.

Tables 4 and 5 show the top 10 features with the highest positive and negative normalized weights, respectively, from the experiment in Section 5.2. Positively weighted features suggest to the classifier that the vulnerability has an exploit; negatively weighted features suggest to the classifier that the vulnerability does not. The first column lists the feature family (Table 2) and the second column the specific feature in the family. For example, the feature "References: BUGTRAQ ID" in the first row of Table 4 corresponds to the token "BUGTRAQ ID" appearing in the "References" feature family of a vulnerability report. The third column lists the number of vulnerabilities N_j in which feature j appears in the training data. For example, the feature "References: BUGTRAQ ID" occurs in 2,045 of the 3,100 vulnerabilities in the training set. The fourth column lists the *normalized* weight of the feature. The normalized weight $\tilde{w}_j = w_j(N_j/N)$ is the raw weight w_j learned by the classifier multiplied by the ratio of the number of vulnerabilities N_j whose jth feature is non-zero divided by the total number of vulnerabilities in the training set ($N = 3100$). By sorting the normalized weights, we reveal the features with the largest overall effect across all vulnerabilities (as opposed to the largest effect on a possibly miniscule number of vulnerabilities).

Feature Family	Feature	N_j	Weight
References	BUGTRAQ ID	2045	0.3674
Modified Date − Create Date	(Time Difference)	3096	0.1860
Authors	(Number of Tokens)	1858	0.1461
Class. Types	LOCATION	2509	0.1373
References	(Number of Tokens)	3054	0.1292
Title	(Number of Tokens)	3085	0.1229
CVE Summary	ALLOWS	1983	0.1146
Notes	(Number of Tokens)	925	0.0971
Modified Date − Disclosure Date	(Time Difference)	3098	0.0902
References	SECUNIA ADV. ID	2107	0.0840

Table 4: Top 10 features with the highest positive normalized weights, and the number of vulnerabilities N_j in which they appear in the training set. Features prefixed with "CVE" are derived from CVE entries, otherwise they come from OSVDB reports.

The features in Table 4 are those that suggest most strongly to the classifier that a vulnerability will be exploited. Many of them correspond to the number of tokens in particular feature families, such as the "Authors", "Title", and "Notes" sections of the vulnerability reports. These weights suggest that vulnerabilities with exploits generally have longer reports in the vulnerability databases than those without exploits; when looking at the vulnerability reports manually, we find that this situation is indeed the case. Other top features are references to other security databases, suggesting that vulnerabilities with exploits are often tracked by multiple sources. Finally, we note that there are many features with positive weights beyond those in Table 4. The top 100 features span nearly all of the feature families listed in Table 2; in other words, there are useful features in all parts of the vulnerability reports.

The features in Table 5 are those that suggest most strongly to the classifier that a vulnerability will not be exploited. Among these features, we observe two trends. First, we see that multiple features measuring the passage of time have strongly negative weights. Thus it appears that vulnerabilities with "dusty" reports are less likely to be exploited. Second, we see that vulnerabilities whose product-related fields are undefined ("Full Product Name", "Products", "Vendors") also appear less likely to be exploited. Presumably, such "incomplete" reports indicate vulnerabilities that have not received much attention from the community (nor from attackers).

5.4 Online Exploit Prediction

The offline experiment in Section 5.2 showed the potential for learning to classify vulnerabilities that were randomly divided into training, test, and development sets. In a real-world deployment, however, system administrators would train the classifier on known vulnerabilities to make predictions about new ones. Moreover, as time goes, the knowledge that vulnerabilities have or have not been exploited can be used to create new training examples. We can then extend the training set with these new vulnerabilities and learn a new classifier based on the most up-to-date information. This process can continue indefinitely as time progresses.

Feature Family	Feature	N_j	Weight
Today − Last Modified Date	(Time Difference)	3097	-0.9432
CVE Today − Generate Date	(Time Difference)	3045	-0.1478
Class. Types	ATTACK TYPE	3052	-0.1439
CVE Mod. Date − Generate Date	(Time Difference)	3044	-0.1412
Classifications	ATTACK TYPE INPUT MANIP	2158	-0.08260
References	RELATED OSVDB ID	1486	-0.07923
Create Date − Disclosure Date	(Time Difference)	3098	-0.06584
Description	CODE	872	-0.05499
CVE References	VUPEN	994	-0.04956
Full Product Name	(Not Defined)	2322	-0.04770
Products	(Not Defined)	2322	-0.04770
Vendors	(Not Defined)	2322	-0.04770

Table 5: Top 10 features (including ties) with the lowest negative normalized weights, and the number of vulnerabilities N_j in which they appear in the training set. Features prefixed with "CVE" are derived from CVE entries, otherwise they come from OSVDB reports.

In our next experiment we emulate this online scenario using the time-stamp information in our vulnerability databases:

1. Initialize a baseline classifier — Consider all vulnerabilities $\{V\}_0^t$ reported between time 0 and time t that have known outcomes (exploited or not) represented with labels $\{L\}_0^t$. We build a baseline classifier C_t based upon these vulnerability examples and labels $\{V, L\}_0^t$.

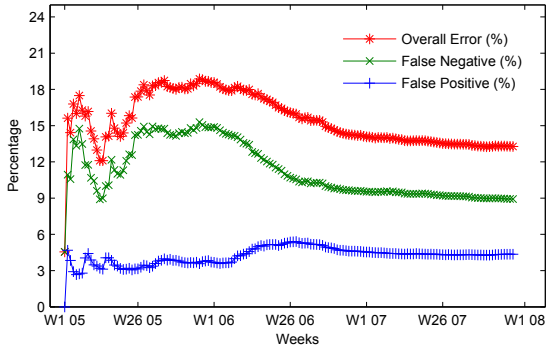
2. Predict exploited vulnerabilities that appear in the next time interval — Suppose new vulnerabilities $\{V\}_{t+1}^{t+T}$ arrive after a duration of T time elapses. We can use C_t to predict the labels $\{L'\}_{t+1}^{t+T}$ for these examples.

3. Update with known vulnerability outcomes — Once we know whether or not vulnerabilities are exploited, we know the true labels $\{L\}_{t+1}^{t+T}$ of the vulnerabilities. With their true labels, we can now include these vulnerabilities $\{V\}_{t+1}^{t+T}$ in the training set $\{V, L\}_t^{t+T}$ and rebuild a new classifier C_{t+T}.

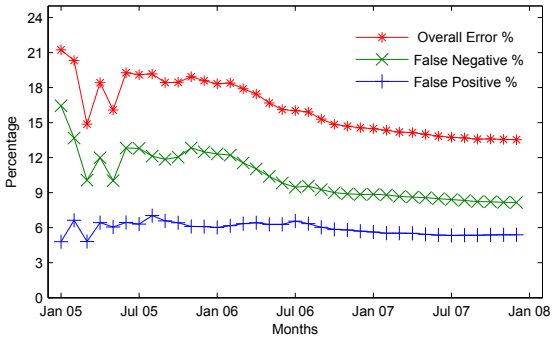
4. Calculate error — We also calculate the cumulative error of the classifier C_t. For each time interval of duration T, we count the number of predicted labels $\{L'\}$ that differ from their true labels $\{L\}$ in that interval, and sum all of the counts across all intervals. We then divide the sum by the total number of vulnerabilities seen up until that time. Calculating the cumulative error shows the stability of the classifier over time.

Figure 2 shows the results of this experiment. We train a classifier starting at January 2005 and initialize it with all prior vulnerabilities appearing before 2005. We then emulate the appearance of vulnerabilities and online reclassification and prediction through December 2007 (the end of our data set). We evaluate two update intervals T, once a week (Figure 2a) and once a month (Figure 2b). We show three curves for the total classification error as well as the false positive and negative rates over time.

These results show that, after initial fluctuations, the classifier stabilizes and improves its accuracy with more examples over time. At the end the classifier has an overall error rate of 14%, a false negative rate of 9% and a false positive rate of 5%. Further, classification accuracy is relatively insensitive to the update period: when



(a) Weekly Training



(b) Monthly Training

Figure 2: Cumulative error, false negative, and false positive percentages for predicting whether vulnerabilities will be exploited in an online, deployed setting. We evaluate two time intervals for updating the classifier, every (a) week and (b) month.

the classifier stabilizes, the weekly and monthly results differ very little. These results demonstrate the viability of deploying a classifier in an online setting to predict whether vulnerabilities will be exploited.

5.5 Predicting Time to Exploit

Next we use SVMs to predict other metrics that help assess the severity of vulnerabilities. In practice, in addition to knowing whether a vulnerability will be exploited, it is also useful to know how soon it will be exploited. (Even if all vulnerabilities will eventually be exploited, it is valuable to know when.) With this knowledge, software vendors can prioritize the patches they release; system administrators can similarly prioritize the installation of these patches.

In general, there are three kinds of time-dependent exploits: positive-day exploits where an exploit is reported after the vulnerability is disclosed; 0-day exploits where an exploit is reported at the same time that the vulnerability is disclosed; and negative-day exploits where the exploit precedes the vulnerability disclosure date (e.g., an attacker exploits a vulnerability before the software vendor realizes the existence and nature of the vulnerability). Ideally, for each vulnerability, we would like know the probability distribution over days when it will be exploited. Such a distribution cannot

t (days)	$\|P\|$	$\|N\|$	Bias	SVM
2	1,404	2,632	65.21%	78.01%
7	1,960	2,076	51.44%	75.78%
14	2,330	1,706	57.73%	77.06%
30	2,733	1,303	67.71%	79.82%

Table 6: Predicting whether vulnerabilities will be exploited within t days.

be modeled by SVMs, which are designed for binary classification. However, we can use SVMs to make similarly relevant predictions.

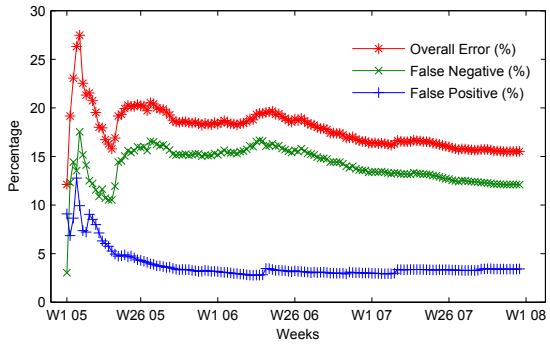
Next we use SVMs to predict whether or not a vulnerability will be exploited within some time t, where t is the difference between the exploit and disclosure dates of the vulnerability. To train such SVMs, we use the same examples as before, merely altering the labels to reflect whether a vulnerability has been exploited within some time frame (as opposed to whether it has been exploited at all). The set of "positive" examples P contains all vulnerabilities with positive-day exploits that are exploited within time t. The set of "negative" examples N contains vulnerabilities with positive-day exploits that are not exploited within time t. (Those that have 0-day or negative-day exploits need no prediction since their reports arrive with the vulnerability already exploited.) We then evaluate a range of time frames for t, from two days to one month.

For this experiment, we used an additional source of information with more accurate dates of vulnerability events. (Unfortunately, the OSVDB database has mixed-quality date information.) From his recent work developing a detailed empirical model of the vulnerability discovery, disclosure, and patch process [10], Stefan Frei generously shared the date information on vulnerabilities with CVE identifiers from his carefully collected data sets. We incorporated his data on discovery, exploit, and disclosure dates for the vulnerabilities contained in our data sets (Table 1).

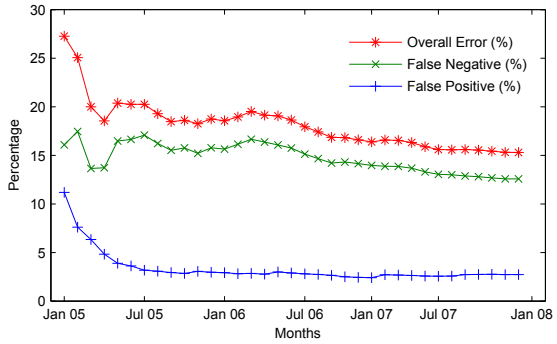
To evaluate the accuracy of predicting time-to-exploit for vulnerabilities, we perform both offline and online experiments similar to those in Sections 5.2 and 5.4. In the offline experiment we train and test classifiers on the entire data set, and in the online experiment we retrain the classifiers and make predictions on vulnerabilities over time.

Table 6 shows the results of the offline experiment. As in Section 5.2, we partition the examples into training and testing sets and report averaged results from cross-validation with ten different random partitions. For each experiment in the table, we show the predicted time frame t, the number of positive $|P|$ and negative $|N|$ examples, the accuracy $\max(|P|,|N|)/(|P|+|N|)$ of the default classifier that always predict the dominant label, and the accuracy from SVMs. The predictions from SVMs are 75–80% accurate across the different time frames; note that these results are significantly better than the raw bias induced from the imbalance of positive and negative training examples. Considering that we have not tuned the classifier, features, or thresholds to optimize the accuracy for this scenario, we believe that these results demonstrate the viability of predicting time-to-exploit from statistical analyses of vulnerability disclosure reports.

Figure 3 shows the results for the online version of the experiment. In the online version, we emulate a real-world deployment where we dynamically update the classifier and make predictions over time as new vulnerabilities appear. We show the results for predicting whether a vulnerability will be exploited within $t = 2$ days, the most severe positive-day case. (Other time frames, not shown, yielded similar results.) The classifier fluctuates initially, then stabilizes after training on a sufficient number of examples. The long-term trend shows a decrease in the false negative and



(a) Weekly Training



(b) Monthly Training

Figure 3: Cumulative error, false negative, and false positive percentages for predicting time to exploit in an online, deployed setting. We evaluate two time intervals for updating the classifier, every (a) week and (b) month.

cumulative error rates while the false positive error rate remains flat. For a simple linear classifier, the overall results are extremely promising: at the end of training, the classifier has an overall cumulative error rate of 15%. Finally, in terms of errors, there are many more false negatives (13%) than false positives (2%).

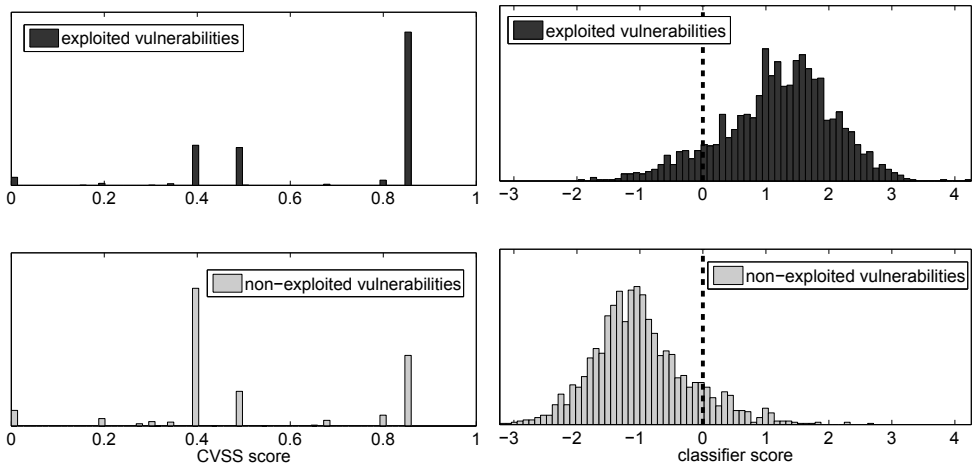
5.6 Exploitability Metrics

Finally, we consider the issue of scoring metrics for vulnerabilities. Specifically, we compare two metrics for assessing how likely a reported vulnerability is likely to be exploited: one based on prior (expert) knowledge and handcrafted formulas, the other based on statistical methods and data mining.

As discussed in Section 2.2, the Common Vulnerabilities Scoring System (CVSS) defines a metric for scoring the "Exploitability" of a vulnerability; see eq. 2. We use this CVSS score as a representative formula-based metric. To be fair, the CVSS specification does not state how to interpret the "Exploitability" score; its intended purpose may not have been to represent the likelihood that a vulnerability is exploited. However, given its name and the factors that determine the score — e.g., difficulty and complexity of programmatically accessing the vulnerability in an exploit attempt — it seems reasonable to expect that the score correlates with exploit likelihood.

Our data-driven approach to vulnerability classification suggests an alternative scoring method. Recall that the decision rule in

Appendix BOZORGI02



(a) CVSS (b) Classifier

Figure 4: Histograms of exploitability scores computed on the vulnerabilities in our data set: (a) computed using the CVSS "Exploitability" formula (Eq. 2) with values normalized to 1; (b) computed using the classifier score ($\mathbf{w}\cdot\mathbf{x}+b$).

eq. (3) computes the signed distance to the maximum margin hyperplane separating positive and negative examples. The signed distance ($\mathbf{w}\cdot\mathbf{x}+b$) serves as a natural score for the exploitability of a vulnerability: the sign predicts whether it will be exploited, and for positively labeled examples, the magnitude indicates its severity.

We compare the effectiveness of these scoring methods by illustrating the distributions of their scores computed on the vulnerabilities in our data set. Visually, these distributions tell a compelling story.

Figure 4(a) shows histograms of CVSS "Exploitability" scores for exploited vulnerabilities (top) and vulnerabilities without exploits (bottom); we have normalized the scores to a maximum of 1. Note that CVSS automatically assigns a normalized score of 1 to all newly discovered vulnerabilities as a precautionary step. Vulnerabilities with that default score dominate the distribution, though, so we have removed them from the histogram to more clearly show the distribution of values computed by the CVSS formula. Figure 4(a) suggests that the CVSS exploitability scores on known vulnerabilities do not consistently reflect what happens in practice. Many vulnerabilities without exploits have high CVSS scores, and many vulnerabilities with exploits have low CVSS scores. As a result, no threshold CVSS score can differentiate well between the exploited and non-exploited vulnerabilities.

Figure 4(b) shows histograms of the classifier scores (i.e., the signed distances $\mathbf{w}\cdot\mathbf{x}+b$) for the same vulnerabilities. The vertical dashed line indicates the default threshold of zero used to predict whether a vulnerability should be labeled as "exploited" or not: values above the threshold are predicted as "exploited", and values below the threshold as "not exploited". As suggested by the results in Section 5.2, the histograms show that the classifier separates these distributions well: few exploited vulnerabilities have scores below the threshold (the false negatives), and few non-exploited vulnerabilities have scores above the threshold (the false positives). We note that preceding experiments included the CVSS score as a feature since it is available when a vulnerability is reported. However, we found that excluding the CVSS score as a feature did not noticeably change any of the results.

Overall, our results suggest that the security community should consider statistical models in addition, or as an alternative to current scoring practices. Such models have many compelling features. First, with little tuning, standard models such as SVMs can provide metrics that correlate well with exploit behavior. Second, the models can dynamically adapt over time to incorporate new features and data sets. Third, such models can be flexibly adapted to yield a variety of predictions—for example, whether a vulnerability will be exploited, or in what time frame it will be exploited. Fourth, the models provide real-valued scores that practitioners can use to prioritize vulnerabilities. Finally, these models can integrate the results from other scoring systems simply by incorporating the metrics defined by other systems as additional features used for classification.

6. CONCLUSION

Ranking vulnerabilities is a critical task for software companies. With thousands of vulnerabilities in hand and limited resources to fix them, it is important to prioritize any operational actions. Current methods, while easy to calculate, rely on static combinations of a small number of human-mediated qualitative variables that seem unlikely to capture the full complexity that drives vulnerability exploitation. In this paper we have described a complementary approach for vulnerability assessment using tools from data mining and machine learning. By considering a far broader range of features and relying on contemporary empirical data rather than "gut instinct" to determine their importance, we demonstrate that this approach can classify vulnerabilities significantly better than at least one currently (and widely) used system for severity scoring.

In general, we believe that machine learning is well-suited to many such security assessment tasks and offers considerable flexibility for consolidating disparate data sources so long as desirable security outcomes can be identified. For example, while this paper has focused specifically on exploitability, it would be straightforward for software vendors to use our approach in triaging *discovered* vulnerabilities to determine how to prioritize the development and deployment of patches. Finally, one limitation with existing vulnerability scoring approaches is they are generally "one size fits all"; they do not provide an easy mechanism for incorporating environment or context-specific information (aside from manually ad-

Appendix BOZORGI02

justing the ad hoc magic numbers in the formulas). In contrast, our data-driven approach provides a consistent way to integrate many local data sources, such as vulnerability scanners, IDS logs and incident ticketing systems, to specialize vulnerability assessment to a particular organization.

For many years, security assessment activities have been more art than science. While we concede that the "holy grail" security metric remains elusive, we see no reason to ignore the power of well-founded statistical methods that can improve the state of the practice.

Acknowledgments

We gratefully acknowledge the assistance of Stefan Frei, who generously shared his carefully collected data on vulnerability event dates [10].

7. REFERENCES

[1] W. A. Arbaugh, W. L. Fithen, and J. McHugh. Windows of vulnerability: A case study analysis. *Computer*, 33(12):52–59, 2000.

[2] A. Arora, A. Nandkumar, and R. Telang. Does information security attack frequency increase with vulnerability disclosure? an empirical analysis. *Information Systems Frontiers*, 8(5), 2006.

[3] A. Arora, R. Telang, and H. Xu. Optimal policy for software vulnerability disclosure. In *Workshop on Economics and Information Security (WEIS'04)*, 2004.

[4] S. M. Bellovin. On the Brittleness of Software and the Infeasibility of Security Metrics. *IEEE Security and Privacy*, 4(4), July 2006.

[5] Cisco. Risk Assessment: Risk Triage for Security Vulnerability Announcements. Cisco Whitepaper, Accessed September, 2009. http://www.cisco.com/web/about/security/intelligence/vulnerability-risk%-triage.html.

[6] CVE Editorial Board. Common Vulnerabilities and Exposures: The Standard for Information Security Vulnerability Names. http://cve.mitre.org/.

[7] C. Dougherty. Vulnerability metric, Updated on July 24, 2008. https://www.securecoding.cert.org/confluence/display/seccode/Vulnerabil%ity+Metric.

[8] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR – A Library for Large Linear Classification. http://www.csie.ntu.edu.tw/~cjlin/liblinear/.

[9] Forum of Incident Response and Security Teams (FIRST). Common Vulnerabilities Scoring System (CVSS). http://www.first.org/cvss/.

[10] S. Frei, D. Schatzmann, B. Plattner, and B. Trammel. Modeling the Security Ecosystem — The Dynamics of (In)Security. In *Proc. of the Workshop on the Economics of Information Security (WEIS)*, June 2009.

[11] IBM. IBM Internet Security Systems X-Force 2008 Trend and Risk Report. White paper, Jan. 2009. http://www-935.ibm.com/services/us/iss/xforce/trendreports/xforce-2008-%annual-report.pdf.

[12] D. Lewis. Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval. In *Proceedings of ECML-98, the 10th European Conference on Machine Learning*, pages 4–15, 1998.

[13] P. Mell, K. Scarfone, and S. Romanosky. A complete guide to the common vulnerability scoring system version 2.0, June, 2007. http://www.first.org/cvss/cvss-guide.html.

[14] Microsoft TechNet Security Team. Microsoft Security Bulletin. http://www.microsoft.com/technet/security/current.aspx.

[15] D. Moore, C. Shannon, and k. claffy. Code-red: a case study on the spread and victims of an internet worm. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 273–284, 2002.

[16] D. Nizovtsev and M. Thursby. Economic analysis of incentives to disclose software vulnerabilities. In *Proc. of the Workshop on the Economics of Information Security*, 2005.

[17] OSVDB. The Open Source Vulnerability Database. http://osvdb.org/.

[18] A. Ozment. The likelihood of vulnerability rediscovery and the social utility of vulnerability hunting. In *Proc. of the Workshop on the Economics of Information Security*, 2005.

[19] E. Rescorla. Security holes... who cares? In *Proc. of the 12th conference on USENIX Security Symposium*, 2003.

[20] Secunia Corporation. Secunia Advisories. http://secunia.com.

[21] Symantec Corporation. Security Focus. http://www.securityfocus.com.

[22] V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, NY, 1998.

MARC View

LEADER 01663nam a22002652 4500			
001 22614607			
003 SIRSI			
005 20180314130447.0			
006 m d			
007 cr n			
008 161108e20160928nju s|||||||| 2|eng|d			
020			a| 9781509038664
020			a| 1509038663 (Spiral) c| USD 231.00 Retail Price (Publisher) 9| Active Record
037			b| 00066573
040			a| BIP US d| WaSeSS
110	2		a| IEEE Staff e| Author
245	1	0	a| 2016 IEEE Conference on Intelligence and Security Informatics (ISI) h| [electronic resource]
260			a| Piscataway : b| IEEE c| Sept. 2016
506			a| License restrictions may limit access.
520	8		a| Annotation b| Intelligence and Security Informatics (ISI)is an interdisciplinary research field involving academic researchers in information technologies, computer science, public policy, bioinformatics, medical informatics, and social and behavior studies local, state, and federal law enforcement and intelligence experts and information technology industry consultants and practitioners ISI supports counterterrorism and homeland security s missions of anticipation, prevention, preparedness and response to security events, in physical, cyber, enterprise, and societal spaces
521			a| Scholarly & Professional b| IEEE
655		0	a| Electronic books.
856	4	0	u| http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001947262&title=2016%20IEEE%20Conference%20on%20Intelligence%20and%20Security%20Informatics%20%28ISI%29
955			a| Books In Print, (c) 2017 R.R. Bowker LLC
949			a| Electronic resource w| ASIS c| 1 i| 22614607-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 3/20/2018 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

2016 IEEE Conference on Intelligence and Security Informatics (ISI) [electronic resource]

Corporate Author IEEE Staff

Published Piscataway : IEEE Sept. 2016

Access Online

serialssolutions.com

ASK A LIBRARIAN

Restrictions on Access

License restrictions may limit access.

Summary

Annotation Intelligence and Security Informatics (ISI)is an interdisciplinary research field involving academic researchers in information technologies, computer science, public policy, bioinformatics, medical informatics, and social and behavior studies local, state, and federal law enforcement and intelligence experts and information technology industry consultants and practitioners ISI supports counterterrorism and homeland security s missions of anticipation, prevention, preparedness and response to security events, in physical, cyber, enterprise, and societal spaces

Genre(s)

Electronic books

ISBN

9781509038664

1509038663 (Spiral)

Audience Notes

Scholarly & Professional IEEE

View MARC record | catkey: 22614607

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368

Libraries Intranet (Staff Only)

Accessibility Help

Website Feedback

Policies and Guidelines

Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook

Twitter

Instagram

Product Offerings in Malicious Hacker Markets

Ericsson Marin, Ahmad Diab and Paulo Shakarian
Arizona State University
Tempe, Arizona
{ericsson.marin, ahmad.diab, shak}@asu.edu

***Abstract*—Marketplaces specializing in malicious hacking products - including malware and exploits - have recently become more prominent on the darkweb and deepweb. We scrape 17 such sites and collect information about such products in a unified database schema. Using a combination of manual labeling and unsupervised clustering, we examine a corpus of products in order to understand their various categories and how they become specialized with respect to vendor and marketplace. This initial study presents how we effectively employed unsupervised techniques to this data as well as the types of insights we gained on various categories of malicious hacking products.**

I. Introduction

Websites on the deepweb and darkweb specializing in the sale of malicious hacking products - such as malware platforms, software exploits and botnet rental - have become the venue of choice for online purchase of these items by cyber criminals. In this paper, we leverage unsupervised learning to categorize and study the product offerings of 17 of these online markets. Specifically, we describe how we used manual labeling combined with clustering techniques to identify product categories (Section II), and then we analyze the results both quantitatively and qualitatively (Section III). We identify categories of products that are highly specialized with respect to particular vendors and markets. We also highlight other interesting facets of this ecosystem - for instance, vendors who habitually cross-list products on multiple sites and nearly identical products for sale by multiple vendors.

Background and Related Work. The darkweb refers to the anonymous communication provided by crypto-network tools such as "The Onion Router" (Tor), which is free software dedicated to protect the privacy of its users by obscuring traffic analysis as a form of network surveillance [1]. On the other hand, the deepweb refers to sites not indexed by common search engines due to a variety of reasons (e.g. password protections), that not necessarily rely on additional protocols.

The sites on the darkweb and deepweb explored in this study comprise marketplaces [2]. In these websites, vendors advertise and sell their goods and services relating to malicious hacking, drugs, pornography, weapons and software services. Products are most often verified before any funds are released to the seller. The main engine of these environments is trust. If a seller is misleading or fails to deliver the appropriate item, he is banned from the site. Similarly, buyers can be banned for not complying with the transaction rules. Basically, all marketplaces in darkweb require a registration and a valid account to get access. Sometimes, this registration process is not trivial, including effort to answer questions, solve puzzles, mathematical equation or CAPTCHA.

Most related work on darkweb markets such as [3] focus on a single market and do not restrict their study to malicious hacking products. Our previous work on markets [4] focused on a game theoretic analysis of a small subset of the data in this paper - and did not attempt to categorize the products for sale. Additionally, there is a complementary lines of work on malicious hacking forums (i.e. [5], [6], [7], [8], [9], [10], [11], [12]) - which is a related but different topic from this paper.

II. Malicious Hacking Product Categorization

In this section, we describe our malicious hacker product dataset and our use of clustering to identify malicious hacker product categories. We examined 17 malicious hacker marketplaces crawled over a 6 month period. The crawled information was then parsed and stored in a relational database. Relevant tables record the marketplaces themselves, the existing vendors and the items/products for sale. Each item is associated with a vendor and a marketplace, allowing for join queries. Some of the more relevant fields for marketplace items include the price, title, description, rating, posting date. In this work, we primarily use the product title/name to generate features.

We note that many items are cross-posted and are nearly identical. Fig. 1(a) shows the distribution of vendors who use the same screen-name across multiple marketplaces. These are the ones responsible for the duplicated (cross-posted) products, reported on the size of our dataset in Table I. Note that only 56.4% of products are distinct by simple comparison methods.

In addition, as we collect data from different sites, there is inconsistency as to how products are categorized on each site - if such non-trivial categorization even exists for a given site. Furthermore, there is a clear absence of a standardized method for vendors to register their products. As a consequence, the great majority of the distinct products are unique when compared with simple matching or regular expression technique. It is valid even in the case where a pair of vendors with different screen names post what a human would determine to be the same product. Fig. 1(b) shows the distribution of products and number of vendors shared among them. The distribution follows a power-law. From the 9,093 distinct products, about 77% are unique (exclusive), not shared between vendors.

Clustering approach. Using product names, we engineer features that represent each product as a vector. A set of pilot experiments suggests that word and character n-grams

978-1-5090-3865-7/16/$31.00 ©2016 IEEE

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:43 UTC from IEEE Xplore. Restrictions apply.

Appendix MARIN02

TABLE I
SCRAPED DATA FROM MARKETPLACES IN DARKWEB.

Marketplaces	17
Products (Total)	16,122
Products (Distinct)	9,093
Vendors	1,332

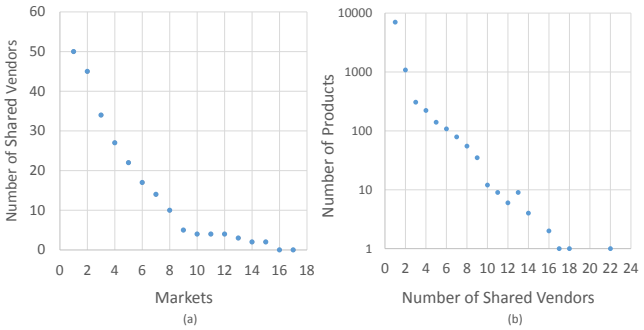


Fig. 1. Distribution of (a) Shared Vendors over Markets. (b) Products over Shared Vendors.

would provide more pure clusters compared with other feature engineering methods, such as meta-data or domain-specific keywords. These features were valued using standard term frequency - inverse document frequency (TF-IDF), after the elimination of stopping words and the execution of steaming. We evaluated word n-gram feature vectors of length up to 1 and up to 2 words and many character n-gram features in the ranges from 3 to 7 and from 4 to 7. This gave us 10 different feature vectors in all. To verify which of these strategies could reach the best performance in our dataset, we evaluated the effect of the different types of feature vectors on the accuracy and purity of clusters produced by the K-means algorithm.

To determine the best feature vector, we manually labeled 500 samples using 34 labeled groups (listed in Table III). We used 400 of the samples to determine centroids for each of the 34 groups, and then we evaluated the resulting clustering on the remaining 100 samples. We examined the accuracy of the different approaches when compared to ground truth using the Rand-index method [13]. This method is defined as the number of pairs correctly considered in the same class or correctly considered in different classes divided by $\binom{n}{2}$, where n is the number of samples. In addition, we used standard entropy measurements to examine the purity of the clusters. Entropy measures the amount of disorder in a cluster. A zero-value for this metric means the clusters are formed by just one class. The formal definition is as follows:

$$entropy(D_i) = -\sum_{i=1}^{k} Pr_i(c_j) \log_2 Pr_i(c_j), \qquad (1)$$

where $Pr_i(c_j)$ is the proportion of class c_j data points in cluster i or D_i. The total entropy (considering all clusters) is:

$$entropy_{total}(D) = \sum_{i=1}^{k} \frac{|D_i|}{|D|} \; x \; entropy(D_i) \qquad (2)$$

Table II shows the performance of each TF-IDF vectorization using Rand-index and entropy, when K-means starts with the 34 fixed centroids. For Rand-index, character n-grams in the range from 3 to 4, 3 to 5, and 3 to 6, when K-means used cosine similarity reached a high best performance (0.986). In addition, we also found the best entropy (0.067) when K-means uses the same specification. This way, K-means configuration with character n-grams in the range from 3 to 6 for vectorization, cosine similarity for distance function and the 34 points for the starting centroids was our natural choice to produce the clusters in the entire dataset.

We also examined the performance of our approach using random centroids. As expected, it performs worse than using the centroids derived from products. Additionally, we examined products (from the full dataset) with a cosine similarity of less than 0.1 from the calculated centroids. There were 410 such distinct products (4.51% of the dataset). These were then manually examined and most were found to be irrelevant to our target domain - and we did not consider them further.

III. ANALYSTS' INTERPRETATION OF PRODUCT CLUSTERS

In this section, we examine the results of clustering based on character n-grams in the range 3 to 6 using initial centroids determined from the labeled data. In order to analyze the information of these clusters, we calculated their entropy with respect to two different criteria: marketplaces and vendors. We also checked in the database the number of distinct marketplaces and vendors inside each cluster. The idea was to understand the diversity of the clusters regarding these two facets. A low marketplace entropy for a given cluster would mean its products were mainly found in a particular marketplace. Similarly, low vendor entropy would mean the cluster's products were mainly sold by a particular vendor. Table III presents the results.

As shown in Table III, *Links* holds the lowest entropy when we analyzed the marketplaces, suggesting the great majority of products come from the same market. In this cluster, 80% of products came from only 2 markets. However, when we check the vendor entropy for this same cluster, we can observe a higher value, suggesting that many vendors are actually offering products related to *Links*. It is possible that many markets discourage the re-selling of lists of links, as much of this information can be found on darkweb Wiki's for free.

Similarly, *Hacking Tools* holds the lowest entropy for the vendor criteria. This suggest that only a few vendors are present in that cluster. Specifically, only 2 vendors author $416(50\%)$ of this type of products. At first glance, this may be surprising as this appears to be a very general group. However, upon inspection of the contents, we find that many authors of these products are actually organizations. These organizations use similar language in their product description in an effort to brand their wares. This could indicate the presence of hacking-collectives that author products as well as the limitations of our text-based approach - which can potentially cluster products branded in a similar fashion. We also note one of the most

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:43 UTC from IEEE Xplore. Restrictions apply.

Appendix MARIN02

TABLE II
K-MEANS EVALUATION (FIXED CENTROIDS).

	Rand-index										
	word(1,1)	word(1,2)	char(3,4)	char(3,5)	char(3,6)	char(3,7)	char(4,4)	char(4,5)	char(4,6)	char(4,7)	Random
Cosine	0.986	0.985	0.986	0.986	0.986	0.985	0.985	0.985	0.984	0.982	0.933
Euclidean	0.986	0.977	0.976	0.973	0.973	0.974	0.975	0.975	0.977	0.971	0.933
	Entropy										
Cosine	0.075	0.079	0.067	0.067	0.067	0.075	0.072	0.079	0.088	0.088	0.423
Euclidean	0.224	0.110	0.153	0.156	0.156	0.141	0.134	0.134	0.137	0.175	0.423

TABLE III
CLUSTERS ENTROPY (14486).

Rank	Cluster Name	Nº of Products	Nº of Markets	Market Entropy	Nº of Vendors	Vendor Entropy
1	Carding	1263	16	0.320	315	0.720
2	PayPal-related	1103	16	0.340	335	0.754
3	Cashing Credit Cards	867	16	0.351	256	0.738
4	PGP	865	15	0.347	203	0.696
5	Netflix-related	846	14	0.270	351	0.805
6	Hacking Tools - General	825	15	0.331	132	0.516
7	Dumps - General	749	12	0.289	280	0.777
8	Linux-related	561	16	0.372	117	0.758
9	Email Hacking Tools	547	13	0.335	196	0.738
10	Network Security Tools	539	15	0.366	117	0.621
11	Ebay-related	472	15	0.385	163	0.772
12	Amazon-related	456	16	0.391	197	0.825
13	Bitcoin	443	15	0.360	201	0.823
14	Links (Lists)	422	12	0.211	221	0.838
15	Banking	384	13	0.349	186	0.840
16	Point of Sale	375	15	0.384	181	0.841
17	VPN	272	12	0.413	130	0.827
18	Botnet	257	12	0.291	110	0.796
19	Hacking Groups Invitation	251	14	0.387	143	0.865
20	RATs	249	15	0.453	99	0.797
21	Browser-related	249	12	0.380	134	0.857
22	Physical Layer Hacking	237	13	0.408	122	0.856
23	Password Cracking	230	13	0.434	100	0.781
24	Smartphone - General	223	14	0.408	110	0.816
25	Wireless Hacking	222	13	0.389	56	0.601
26	Phishing	218	13	0.403	111	0.849
27	Exploit Kits	218	14	0.413	91	0.795
28	Viruses/Counter AntiVirus	210	14	0.413	60	0.684
29	Network Layer Hacking	205	14	0.459	60	0.716
30	RDP Servers	191	12	0.405	124	0.895
31	Android-related	156	11	0.429	60	0.770
32	Keyloggers	143	13	0.496	77	0.862
33	Windows-related	119	12	0.464	50	0.717
34	Facebook-related	119	15	0.501	67	0.876

prominent vendor in this cluster was itself a marketplace - which is also reflected in the low marketplace entropy.

In our analysis of the *Facebook* and *Keylogger* clusters, we can see that they point to the other direction. They have high values for both entropy, a clear sign about the diversity with respect to both vendors and markets. For example, in cluster *Facebook*, there were 119 products and 67 vendors, and the most prolific vendor for this cluster authored only 8 products. In the same cluster, products were also spread across 15 markets - and the most well-represented market was associated with 30 products. This analysis also indicates widespread prevalence of keyloggers - which is not surprising as it is a well established hacking technique. However, observing the similar trend for the *Facebook* cluster could be indicative of an increase in demand for Facebook-directed social media hacking products and information.

Conclusion and Future Work. In this paper, we conducted an initial examination of malware products from 17 malicious hacker markets through unsupervised learning. Using manually-labeled data, we studied the effect of feature vector on cluster purity using text-based features. We then analyzed the impurity of clusters in our corpus of over $8,000$ malicious hacking products with respect to vendor and marketplace, and finally, we identified several interesting characteristics of how the products were grouped. Currently, we are examining other methods for grouping these products using matrix factorization and supervised techniques. Additionally, we are studying the underlying social network of vendors through relationships based on similar product offerings.

Acknowledgments. Some of the authors were supported by the Office of Naval Research (ONR) Neptune program, the Arizona State University Global Security Initiative (ASU GSI), and CNPq-Brazil.

REFERENCES

[1] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-generation Onion Router," in *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, ser. SSYM'04. Berkeley, CA, USA: USENIX Association, 2004, pp. 21–21.

[2] V. Ciancaglini, M. Balduzzi, R. McArdle, and M. Rösler, "Below the Surface: Exploring the Deep Web," 2015.

[3] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13. New York, NY, USA: ACM, 2013, pp. 213–224.

[4] J. Robertson, V. Paliath, J. Shakarian, A. Thart, and P. Shakarian, "Data Driven Game Theoretic Cyber Threat Mitigation," in *Proc. 28th Innovative Applications of Artificial Intelligence (IAAI-16)*, 2016.

[5] C. C. Yang, X. Tang, and X. Gong, "Identifying dark web clusters with temporal coherence analysis," in *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*, July 2011, pp. 167–172.

[6] V. Benjamin, W. Li, T. Holt, and H. Chen, "Exploring threats and vulnerabilities in hacker web: Forums, irc and carding shops," in *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*, May 2015, pp. 85–90.

[7] M. Macdonald, R. Frank, J. Mei, and B. Monk, "Identifying digital threats in a hacker web forum," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ser. ASONAM '15. New York, NY, USA: ACM, 2015, pp. 926–933.

[8] Z. Zhao, G.-J. Ahn, H. Hu, and D. Mahi, "SocialImpact: Systematic Analysis of Underground Social Dynamics." in *ESORICS*, ser. Lecture Notes in Computer Science, S. Foresti, M. Yung, and F. Martinelli, Eds., vol. 7459. Springer, 2012, pp. 877–894.

[9] H. Chen, "Dark web: Exploring and mining the dark side of the web," in *Intelligence and Security Informatics Conference (EISIC), 2011 European*, Sept 2011, pp. 1–2.

[10] J. Shakarian, P. Shakarian, and A. Ruef, "Cyber attacks and public embarrassment: A survey of some notable hacks," *Elsevier SciTechConnect*, 2015.

[11] P. Shakarian and J. Shakarian, "Socio-cultural modeling for cyber threat actors," in *AAAI Workshop on Artificial Intelligence and Cyber Security (AICS)*, 2016.

[12] J. Shakarian, A. Gunn, and P. Shakarian, "Exploring malicious hacker forums," in *Cyber Deception: Building the Scientific Foundation*, S. Jajodia, V. Subrahmanian, V. Swarup, and C. Wang, Eds. Springer, 2016.

[13] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, Dec. 1971.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:43 UTC from IEEE Xplore. Restrictions apply.

Appendix MARIN02

MARC View

LEADER 01663nam a22002652 4500				
001 22614607				
003 SIRSI				
005 20180314130447.0				
006 m d				
007 cr n				
008 161108e20160928nju s|||||||| 2|eng|d				
020			a| 9781509038664	
020			a| 1509038663 (Spiral) c| USD 231.00 Retail Price (Publisher) 9| Active Record	
037			b| 00066573	
040			a| BIP US d| WaSeSS	
110	2		a| IEEE Staff e| Author	
245	1	0	a| 2016 IEEE Conference on Intelligence and Security Informatics (ISI) h| [electronic resource]	
260			a| Piscataway : b| IEEE c| Sept. 2016	
506			a| License restrictions may limit access.	
520	8		a| Annotation b| Intelligence and Security Informatics (ISI)is an interdisciplinary research field involving academic researchers in information technologies, computer science, public policy, bioinformatics, medical informatics, and social and behavior studies local, state, and federal law enforcement and intelligence experts and information technology industry consultants and practitioners ISI supports counterterrorism and homeland security s missions of anticipation, prevention, preparedness and response to security events, in physical, cyber, enterprise, and societal spaces	
521			a| Scholarly & Professional b| IEEE	
655		0	a| Electronic books.	
856	4	0	u| http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001947262&title=2016%20IEEE%20Conference%20on%20Intelligence%20and%20Security%20Informatics%20%28ISI%29	
955			a| Books In Print, (c) 2017 R.R. Bowker LLC	
949			a| Electronic resource w| ASIS c| 1 i| 22614607-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 3/20/2018 o| .STAFF. 360 MARC	

ASK A LIBRARIAN

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

PennState University Libraries | Catalog

Bookmarks 0 Course Reserves My Account

Keyword Search... Search Advanced search Start Over

← Previous | 1 of 325 | Next →

Share Bookmark Report an Issue

2016 IEEE Conference on Intelligence and Security Informatics (ISI) [electronic resource]

Corporate Author IEEE Staff

Published Piscataway : IEEE Sept. 2016

Access Online

serialssolutions.com

ASK A LIBRARIAN

Restrictions on Access

License restrictions may limit access.

Summary

Annotation Intelligence and Security Informatics (ISI)is an interdisciplinary research field involving academic researchers in information technologies, computer science, public policy, bioinformatics, medical informatics, and social and behavior studies local, state, and federal law enforcement and intelligence experts and information technology industry consultants and practitioners ISI supports counterterrorism and homeland security s missions of anticipation, prevention, preparedness and response to security events, in physical, cyber, enterprise, and societal spaces

Genre(s)

Electronic books

ISBN

9781509038664
1509038663 (Spiral)

Audience Notes

Scholarly & Professional IEEE

View MARC record | catkey: 22614607

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence

Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin, Vineet Mishra,
Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, Paulo Shakarian
Arizona State University
Tempe, AZ 85281, USA
Email: {enunes1, ahmad.diab, andrewgunn, ericsson.marin, vvmishra,
vivin.paliath, jj.robertson, jshak, amanda.thart, shak} @asu.edu

***Abstract*—In this paper, we present an operational system for cyber threat intelligence gathering from various social platforms on the Internet particularly sites on the darknet and deepnet. We focus our attention to collecting information from hacker forum discussions and marketplaces offering products and services focusing on malicious hacking. We have developed an operational system for obtaining information from these sites for the purposes of identifying emerging cyber threats. Currently, this system collects on average 305 high-quality cyber threat warnings each week. These threat warnings include information on newly developed malware and exploits that have not yet been deployed in a cyber-attack. This provides a significant service to cyber-defenders. The system is significantly augmented through the use of various data mining and machine learning techniques. With the use of machine learning models, we are able to recall 92% of products in marketplaces and 80% of discussions on forums relating to malicious hacking with high precision. We perform preliminary analysis on the data collected, demonstrating its application to aid a security expert for better threat analysis.**

I. Introduction

Pre-reconnaissance cyber threat intelligence refers to information gathered before a malicious party interacts with the defended computer system. An example demonstrating the importance of cyber threat intelligence is shown in Table 1. A Microsoft Windows vulnerability was identified in Feb. 2015. The release of the vulnerability was essentially Microsoft warning its customers of a security flaw. Note that at this time, there was no publicly known method to leverage this flaw in a cyber-attack (i.e. an available exploit). However, about a month later an exploit was found to be on sale in darknet market. It was not until July when FireEye, a major cybersecurity firm, identified that the Dyre Banking Trojan designed to steal credit cards exploited this vulnerability - the first time an exploit was reported. This vignette demonstrates how threat warnings gathered from the darknet can provide valuable information for security professionals. The average global exposure of the Dyre Banking Trojan was 57.3% along with another banking malware Dridex[1]. It means that nearly 6 out of 10 organizations in the world were affected, and this is a significantly high number on a global level.

In this paper, we examine how such intelligence can be gathered and analyzed from various social platforms on the Internet particularly sites on the darknet and deepnet. In doing so, we encounter several problems that we addressed with various data mining techniques. Our current system is operational and actively collecting approximately 305 cyber threats each week. Table 2 shows the current database statistics. It shows the total data collected and the data related to malicious hacking. The vendor and user statistics cited only consider those individuals associated in the discussion or sale of malicious hacking-related material, as identified by the system. The data is collected from two sources on the darknet/deepnet: markets and forums.

TABLE 1: Exploit example.

Timeline	Event
Feb. 2015	Microsoft identifies Windows vulnerability MS15-010/CVE 2015-0057 for remote code execution. There was no publicly known exploit at the time the vulnerability was released.
April 2015	An exploit for MS15-010/CVE 2015-0057 was found on a darknet market on sale for 48 BTC (around $10,000-15,000).
July 2015	FireEye identified that the Dyre Banking Trojan, designed to steal credit card number, actually exploited this vulnerability[1].

TABLE 2: Current Database Status

Markets	Total Number	17
	Total products	11991
	Hacking related	1573
	Vendors	434
Forums	Total Number	21
	Topics/Posts	23780/162872
	Hacking related	4423/31168
	Users	5491

We are providing this information to cyber-security professionals to support their strategic cyder-defense planning to address questions such as, *1) What vendors and users have a presence in multiple darknet/deepnet markets/ forums? 2)What zero-day exploits are being developed by malicious hackers? 3) What vulnerabilities do the latest exploits target?*

[1]https://www.fireeye.com/blog/threat-research/2015/06/evolution_of_dridex.html

978-1-5090-3865-7/16/$31.00 ©2016 IEEE

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

Specific contributions of this paper include, 1) Description of a system for cyber threat intelligence gathering from various social platforms from the Internet such as deepnet and darknet websites. 2) The implementation and evaluation of learning models to separate relevant information from noise in the data collected from these online platforms. 3) A series of case studies showcasing various findings relating to malicious hacker behavior resulting from the data collected by our operational system.

Background: Many of the individuals behind cyber-operations – originating outside of government run labs or military commands – rely on a significant community of hackers. They interact through a variety of online forums (as means to both stay anonymous and to reach geographically dispersed collaborators).

Darknet and Deepnet Sites: Widely used for underground communication, "The Onion Router" (Tor) is free software dedicated to protect the privacy of its users by obscuring traffic analysis as a form of network surveillance [9]. The network traffic in Tor is guided through a number of volunteer-operated servers (also called "nodes"). Each node of the network encrypts the information it blindly passes on neither registering where the traffic came from nor where it is headed [9], disallowing any tracking. Effectively, this allows not only for anonymized browsing (the IP-address revealed will only be that of the last node), but also for circumvention of censorship[2]. Here, we will use "darknet" to denote the anonymous communication provided by crypto-networks like "Tor", which stands in contrast to "deepnet" which commonly refers to websites hosted on the open portion of the Internet (the "Clearnet"), but not indexed by search engines [15].

Markets: Users advertise and sell their wares on marketplaces. Darknet marketplaces provide a new avenue to gather information about the cyber threat landscape. The marketplaces sell goods and services relating to malicious hacking, drugs, pornography, weapons and software services. Only a small fraction of products (13% in our collected data to date) are related to malicious hacking. Vendors often advertise their products on forums to attract attention towards their goods and services.

Forums. Forums are user-oriented platforms that have the sole purpose of enabling communication. It provides the opportunity for the emergence of a community of like-minded individuals - regardless of their geophysical location. Administrators set up Darknet forums with communication safety for their members in mind. While structure and organization of Darknet-hosted forums might be very similar to more familiar web-forums, the topics and concerns of the users vary distinctly. Forums addressing malicious hackers feature discussions on programming, hacking, and cyber-security. Threads are dedicated to security concerns like privacy and online-safety - topics which plug back into and determine the structures and usage of the platforms.

II. SYSTEM OVERVIEW

Fig. 1 gives the overview of the system. Through search engines and spider services on the Tor network, human analysts were able to find forums and marketplaces populated by malicious hackers. Other platforms were discovered through links posted on forums either on the Tor-network or on the Clearnet. The system consists of three main modules built independently before integration. The system is currently fully integrated and actively collecting cyber threat intelligence.

Crawler: The crawler is a program designed to traverse the website and retrieve HTML documents. Topic based crawlers have been used for focused crawling where only webpages of interest are retrieved [17], [6]. More recently, focused crawling was employed to collect forum discussions from darknet [10]. We have designed separate crawlers for different platforms (markets/forums) identified by experts due to the structural difference and access control measures for each platform. In our crawler, we address design challenges like accessibility, unresponsive server, repeating links creating a loop etc. to gather information regarding products from markets and discussions on forums.

Parser: We designed a parser to extract specific information from marketplaces (regarding sale of malware/exploits) and hacker forums (discussion regarding services and threats). This well-structured information is stored in a relational database. We maintain two databases, one for marketplaces and the other for forums. Like the crawler, each platform has its own parser. The parser also communicates with the crawler from time to time for collection of temporal data. The parser communicates a list of relevant webpages to the crawler, which are re-crawled to get time-varying data. For markets we collect the following important products fields: {*item_title, item_description, vendor_name, shipping_details, item_reviews, items_sold, CVE, items_left, transaction_details, ratings*}. For forums we collect the following fields: {*topic_content, post_content, topic_author, post_author, author_status, reputation, topic_interest*}.

Classifier: We employ a machine learning technique using an expert-labeled dataset to detect relevant products and topics from marketplaces and forums respectively discussed in Section III. These classifiers are integrated into the parser to filter out products and topics relating to drugs, weapons, etc. not relevant to malicious hacking.

III. EVALUATION

We consider the classification of identifying relevant products in darknet/deepnet marketplaces and relevant topics on forum post containing communication relevant to malicious hacking in this paper. It is a binary classification problem with the data sample (in this case products/forum topics) being relevant or not. We look at both supervised and semi-supervised approaches to address the classification.

A. Machine Learning Approaches

In this work, we leverage a combination of supervised and semi-supervised methods. Supervised methods include the well-known classification techniques of Naive Bayes (NB), random forest (RF), support vector machine (SVM) and logistic regression (LOG-REG). However, supervised techniques required labeled data, and this is expensive and often requires expert knowledge. Semi-supervised approaches work with limited labeled data by leveraging information from unlabeled data. We discuss popular semi-supervised

[2]See the Tor Project's official website (https://www.torproject.org/)

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

Appendix NUNES02

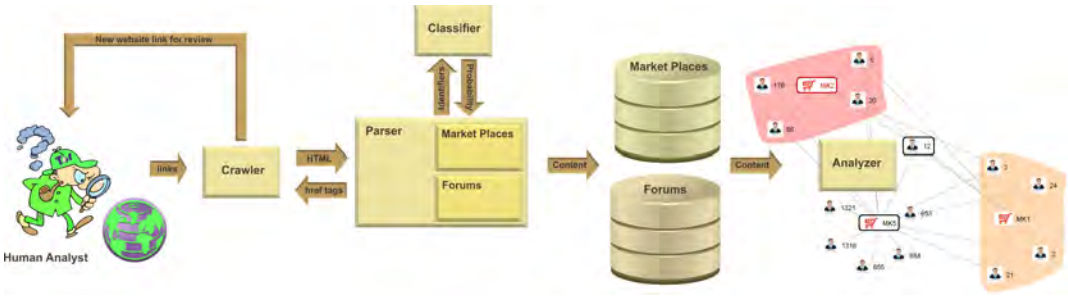


Fig. 1: System overview

approaches used in this work. We perform a grid search to find optimal parameters for the learning techniques.

Label propagation (LP). The label propagation approach [22] has been widely used for semi-supervised classification task [3], [16], [21], [8]. It estimates the label values based on graph Laplacian [1] where the model is represented by a weighted graph $G = (V, E)$, where V indicates the vertices representing the samples, while the edges E are the weights indicating the similarity between points. A subset of these vertices are labeled and these vertices are then used to estimate the labels of the remaining under the assumption that the edges are able to capture the similarity between samples. Hence, the performance of these methods depends on the similarity measure used. The most commonly used similarity measures include k-NN and Gaussian kernel.

Co-training (CT). The Co-training approach was proposed by Blum and Mitchell [4]. In this approach, the feature set is divided into two sets (assumed to be independent), and two classifiers are trained using the limited labeled set denoted by L . These trained classifiers are then used to estimate the labels for the unlabeled points. High confidence label estimates from classifier-1 are added to the labeled set L of classifier-2 and vice versa. For the current setting we set the confidence to 70%. Every time the labeled set L is updated, the classifiers are retrained. This procedure repeats until all of the unlabeled points are labeled. It can be viewed as two classifiers teaching each other.

B. Experiments: Marketplaces

Marketplaces sell goods and services that do not relate to malicious hacking, including drugs, pornography, weapons and software services. Only a small fraction of products (13%) are related to malicious hacking. We thus require a model that can separate relevant products from the non-relevant ones. The data collected from marketplaces is noisy and hence not suitable to use directly as input to a learning model. Hence, the raw information undergoes several steps of automated data cleaning. We now discuss the challenges associated with the dataset obtained and the data processing steps taken to address them. We note that similar challenges occur for forum data.

Text Cleaning. Product title and descriptions on marketplaces often have much text that serves as noise to the classifier (e.g. *****SALE*****). To deal with these instances, we first removed all non-alphanumeric characters from the title and description. This, in tandem with standard stop-word removal, greatly improved classification performance.

Misspellings and Word Variations. Misspellings frequently occur on forums and marketplaces, which is an obstacle for the standard bag-of-words classification approach. Additionally, with the standard bag-of-words approach, variations of words are considered separately (e.g. hacker, hack, hackers, etc.). Word stemming mitigates these issue of word variations, but fails to fix the issue of misspellings. To address this we use character n-gram features. As an example of character n-gram features, consider the word “hacker”. If we were using tri-gram character features, the word “hacker” would yield the features “hac”, “ack”, “cke”, “ker”. The benefit of this being that the variations or misspellings of the word in the forms “hack”, “hackz”, “”hackker”, will all have some common features. We found that using character n-grams in the range (3, 7) outperformed word stemming in our experiments.

Large Feature Space. In standard bag-of-words approach, as opposed to the character n-gram approach, the feature matrix gets very large as the number of words increase. As the number of unique words grow, this bloated feature matrix begins to greatly degrade performance. Using n-gram features further increases the already over-sized feature matrix. To address this issue, we leveraged the sparse matrix data structure in the scipy[3] library, which leverages the fact that most of the entries will be zero. If a word or n-gram feature is not present in a given sample, there is simply no entry for that feature in the sparse matrix.

Preserving Title Feature Context. As the title and description of the product are disjoint, we found that simply concatenating the description to the product title before extracting features led to sub-optimal classification performance. We believe that by doing a simple concatenation, we were losing important contextual information. There are features that should be interpreted differently should they appear in the title versus the description. Initially, we used two separate classifiers: one for the title and one for the description. With this construction, when an unknown product was being classified, we would pass the title to the title classifier and the description to the description classifier. If either classifier returned a positive classification, we would assign the product a positive classification. However, we believe that this again led to the loss of important contextual information. To fix this, we independently extract character n-gram features from the title and description.

[3]http://www.scipy.org/

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

TABLE 3: Markets and Number of products collected.

Markets	Products
Market-1	439
Market-2	1329
Market-3	455
Market-4	4018
Market-5	876

Markets	Products
Market-6	497
Market-7	491
Market-8	764
Market-9	2014
Market-10	600

This step yields a title feature vector and a description feature vector. We then horizontally concatenate these vectors, forming a single feature vector which includes separate feature sets for the title and description.

Results: We consider 10 marketplaces to train and test our learning model. A summary of these marketplaces is shown in Table 3. Table 4 gives an instance of products defined as being relevant or not. With the help of security experts we label 25% of the products from each marketplace. The experimental setup is as follows. We perform a leave-one-marketplace-out cross-validation. In other words, given n marketplaces we train on $n-1$ and test on the remaining one. We repeat this experiment for all the marketplaces. For the supervised experiment, we only use the 25% labeled data from each marketplace. We evaluate the performance based primarily on three metrics: precision, recall and unbiased F1. Precision indicates the fraction of products that were relevant from the predicted ones. Recall is the fraction of relevant products retrieved. F1 is the harmonic mean of precision and recall. The results are averaged and weighted by the number of samples in each market. In this application, a high recall is desirable as we do not want to omit relevant products. In the supervised approaches, SVM with linear kernel performed the best, recalling 87% of the relevant products while maintaining a precision of 85% (Fig. 2). SVM performed the best likely due to the fact it maximizes generality as opposed to minimizing error.

TABLE 4: Example of Products.

Product Title	Relevant
20+ Hacking Tools (Botnets Keyloggers Worms and More!)	YES
5 gm Colombian Cocaine	NO

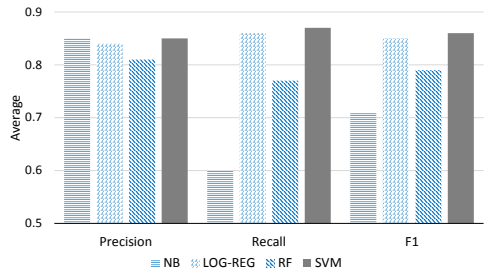


Fig. 2: Average Precision, Recall and F1 comparisons for NB, LOG-REG, RF and SVM for product classification.

As stated, only 25% of the data is labeled, as labeling often requires expert knowledge. However, this significant cost and time investment can be reduced by applying a semi-supervised approach which leverages the unlabeled data to aid in classification. It takes approximately one minute for a human to label 5 marketplace products or 2 topics on forums as relevant or not, highlighting the costliness of manual labeling. The experimental setup is similar to the supervised approach, but this time we also utilize the large unlabeled data from each marketplace (75%) for training.

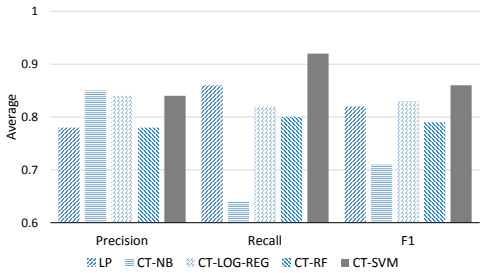


Fig. 3: Average Precision, Recall and F1 comparisons for LP, CT-NB, CT-LOG-REG, CT-RF and CT-SVM for product classification.

Fig. 3 shows the performance comparison for the semi-supervised approaches. For the co-training approach, we divide the feature space into two sets. The two feature sets used are both based on character n-grams. However, the set of words from which the character n-grams are derived are disjoint between the two sets. In this way, the two corresponding feature vectors can be treated as being independent from one another. Hence we get two views of the same sample. Co-training with Linear SVM is able to recall 92% of the relevant products as compared to label propagation and other variants of co-training while maintaining a precision of 82%, which is desirable. In this case, the unlabeled data aided the classification in improving the recall to 92% without significantly reducing the precision.

C. Experiment: Forums

In addition to the darknet/deepnet marketplaces that we have already discussed, there are also numerous darknet forums on which users discuss malicious hacking related topics. Again, there is the issue that only a fraction of these topics with posts on these forums contain information that is relevant to malicious hacking or the trading of exploits. Hence, we need a classifier to identify relevant topics. This classification problem is very similar to the product classification problem previously discussed, with similar set of challenges.

We performed evaluation on two such English forums. The dataset consisted of 781 topics with 5373 posts. Table 5 gives instance of topics defined as being relevant or not. We label 25% of the topics and perform a 10-fold cross validation using supervised methods. We show the results from the top two performing supervised and semi-supervised methods. In the supervised setting, LOG-REG performed the best with 80% precision and 68% recall (Fig. 4). On the other hand, leveraging unlabeled data in a semi-supervised technique improved the recall while maintaining the precision. We note that in this case the 10-fold cross validation was performed only on the labeled points. In the semi-supervised domain

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

Appendix NUNES02

co-training with LOG-REG improved the recall to 80% with precision of 78%.

TABLE 5: Example of Topics.

Topic	Relevant
Bitcoin Mixing services	YES
Looking for MDE/MDEA shipped to Aus	NO

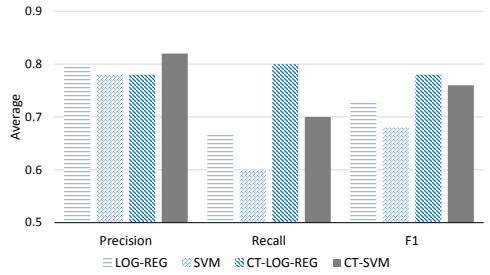


Fig. 4: Average Precision, Recall and F1 comparisons for LOG-REG, SVM, CT-LOG-REG, and CT-SVM for English forum topic classification.

IV. Case Studies

We analyze the data with the purpose of answering the questions raised in the Section I. We will be using the following key security terms. *Vulnerability* is a security flaw that allows an attacker to compromise a software or an operating system. *Exploit* is a piece of software that takes advantage of a vulnerability in a piece of software or operating system to compromise it. *Patch* is a piece of software used to improve existing software by fixing vulnerabilities to improve security. We discuss the following case-studies.

A. Discovery of Zero-Day Exploits.

Over a 4 week period, we detected 16 zero-day exploits from the marketplace data. Zero-day exploits leverage vulnerabilities that are unknown to the vendor. Table 6 shows a sample of zero-day exploits with their selling price in Bitcoin. The Android WebView zero-day affects a vulnerability in the rendering of web pages in Android devices. It affects devices running on Android 4.3 Jelly Bean or earlier versions of the operating system. This comprised of more than 60% of the Android devices in 2015. After the original posting of this zero-day, a patch was released in Android KitKat 4.4 and Lollipop 5.0 which required devices to upgrade their operating system. As not all users have/will update to the new operating system, the exploit continues to be sold for a high price. Detection of these zero-day exploits at an earlier stage can help organizations avoid an attack on their system or minimize the damage. For instance, in this case, an organization may decide to prioritize patching, updating, or replacing certain systems using the Android operating system.

B. Users having presence in markets/ forums.

Previous studies on darknet crawling [10], [2] explore a single domain, namely forums. We create a social network that includes both types of information studied in this paper: marketplaces and forums. We can thus study and find these

TABLE 6: Example of Zero-day exploits.

Zero-day exploit	Price (BTC)
Internet Explorer 11 Remote Code Execution 0day	20.4676
Android WebView 0day RCE	40.8956

cross-site connections that were previously unstudied. We are able to produce this connected graph using the "usernames" used by vendors and users in each domain. A subgraph of this network containing some of the individuals who are simultaneously selling products related to malicious hacking and publishing in hacking related forums is shown in Fig. 5. In most cases, the vendors are trying to advertise/discuss their products on the forums, demonstrating their expertise. Using these integrated graphic representations, one can visualize the individuals' participation in both domains, making the right associations that lead to a better comprehension of the malicious hacker networks. It is helpful in determining social groups within the forums of user interaction. The presence of users on multiple markets and forums follows a power law. From Fig. 6, majority of users only belong to a single market or forum. We note that there are 751 users that are present in more than two platforms. Fig. 7 considers one such user/vendor. The vendor is active in 7 marketplaces and 1 forum . The vendor offers 82 malicious hacking related products and discusses these products on the forum. The vendor has an average rating of 4.7/5.0, rated by customers on the marketplace with more than 7000 successful transactions, indicating the reliability of the products and the popularity of the vendor.

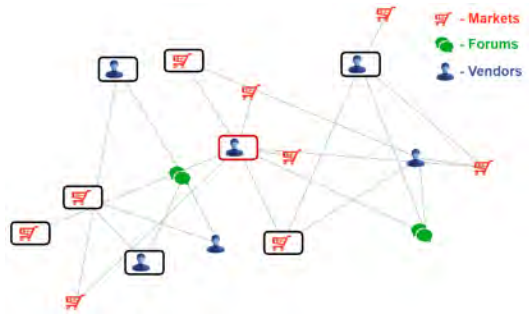


Fig. 5: Vendor/User network in marketplace and forum.

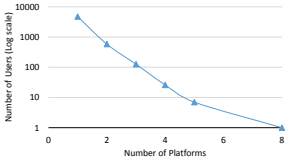


Fig. 6: Users in multiple markets and forums.

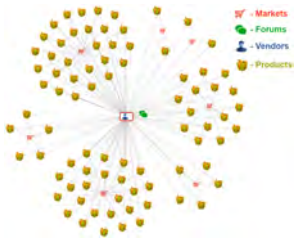


Fig. 7: A centric network of a Vendor.

V. Related Work

Web crawling is a popular way of collecting large amounts of data from the Internet. In many applications, researchers

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

Appendix NUNES02

are interested in specific topics for their application. Hence, the need for a topic-based crawler popularly referred to as a focused crawler [6], [5]. Most of the focused crawlers are designed to collect information from the *surface web* with little concentration on the darknet websites. More recently, a focused crawler concentrating on dark web forums was designed [10]. This research primarily concentrated on forums, collecting data over a period of time and then performing static analysis to study online communities. The authors also describe different data mining techniques for these forums in [7]. We, on the other hand, not only look at darknet forums but also collect information from marketplaces hosting a range of products relating to malicious hacking. Another application of leveraging darknet information to counter human trafficking is developed by DARPA through the Memex program[4] - a program with different goals than the work described in this paper.

Previous work leverages the exploit information from marketplaces in a game theoretic framework to formulate system configurations that minimize the potential damage of a malicious cyber attack [19]. Work analyzing hacker forums to detect threats that pose great risk to individuals, businesses, and government have been discussed in [2]. It further states that knowledge is distributed in forums. That minimally skilled people could learn enough by simply frequenting such platforms. Studying these hacker communities gives insights in the social relationships. Also, the distribution of information amongst users in these communities based on their skill level and reputation [13], [14], [11]. These forums also serve as markets where malware and stolen personal information are shared / sold [12]. Samtani et al. analyze hacker assets in underground forums [20]. They discuss the dynamics and nature of sharing of tutorials, source code, and "attachments" (e.g. e-books, system security tools, hardware/software). Tutorials appear to be the most common way of sharing resources for malicious attacks. Source code found on these particular forums was not related to specific attacks. Additionally underground (not malicious hacking related) forums have also been analyzed to captures the dynamic trust relationships forged between mutually distrustful parties [18].

VI. Conclusion

In this paper, we implement a system for intelligence gathering related to malicious hacking. Our system is currently operational. We are in the process of transitioning this system to a commercial partner. We consider social platforms on darknet and deepnet for data collection. We address various design challenges to develop a focused crawler using data mining and machine learning techniques. The constructed database is made available to security professionals in order to identify emerging cyber-threats and capabilities.

Acknowledgments: Some of this work is supported by ONR NEPTUNE, ASU GSI, ASU ISSR and CNPq-Brazil.

References

[1] M. Belkin and P. Niyogi. Using manifold structure for partially labelled classification. In *Advances in NIPS*, 2002.

[2] V. Benjamin, W. Li, T. Holt, and H. Chen. Exploring threats and vulnerabilities in hacker web: Forums, irc and carding shops. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*, pages 85–90. IEEE, 2015.

[3] C. M. Bishop and I. Ulusoy. Object recognition via local patch labelling. In *Deterministic and Statistical Methods in Machine Learning*, pages 1–21, 2004.

[4] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pages 92–100, New York, NY, USA, 1998. ACM.

[5] S. Chakrabarti, K. Punera, and M. Subramanyam. Accelerated focused crawling through online relevance feedback. In *Proceedings of the 11th international conference on World Wide Web*, pages 148–159. ACM, 2002.

[6] S. Chakrabarti, M. Van den Berg, and B. Dom. Focused crawling: a new approach to topic-specific web resource discovery. *Computer Networks*, 31(11):1623–1640, 1999.

[7] H. Chen. *Dark web: Exploring and data mining the dark side of the web*, volume 30. Springer Science & Business Media, 2011.

[8] H. Cheng, Z. Liu, and J. Y. 0001. Sparsity induced similarity measure for label propagation. In *ICCV*, pages 317–324. IEEE, 2009.

[9] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, pages 21–21, 2004.

[10] T. Fu, A. Abbasi, and H. Chen. A focused crawler for dark web forums. *Journal of the American Society for Information Science and Technology*, 61(6):1213–1231, 2010.

[11] T. J. Holt. Subcultural evolution? examining the influence of on-and off-line experiences on deviant subcultures. *Deviant Behavior*, 28(2):171–198, 2007.

[12] T. J. Holt and E. Lampke. Exploring stolen data markets online: products and market forces. *Criminal Justice Studies*, 23(1):33–50, 2010.

[13] T. J. Holt, D. Strumsky, O. Smirnova, and M. Kilger. Examining the social networks of malware writers and hackers. *International Journal of Cyber Criminology*, 6(1):891–903, 2012.

[14] T. Jordan and P. Taylor. A sociology of hackers. *The Sociological Review*, 46(4):757–780, 1998.

[15] D. Lacey and P. M. Salmon. It's dark in there: Using systems analysis to investigate trust and engagement in dark web forums. In D. Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, volume 9174 of *Lecture Notes in Computer Science*, pages 117–128. Springer International Publishing, 2015.

[16] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, CVPR '06, pages 61–68, Washington, DC, USA, 2006. IEEE Computer Society.

[17] F. Menczer, G. Pant, and P. Srinivasan. Topical web crawlers: Evaluating adaptive algorithms. *ACM Transactions on Internet Technology (TOIT)*, 4(4):378–419, 2004.

[18] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker. An analysis of underground forums. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 71–80. ACM, 2011.

[19] J. Robertson, V. Paliath, J. Shakarian, A. Thart, and P. Shakarian. Data driven game theoretic cyber threat mitigation. In *IAAI*, 2016.

[20] S. Samtani, R. Chinn, and H. Chen. Exploring hacker assets in underground forums. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*, pages 31–36. IEEE, 2015.

[21] C. Wang, S. Yan, L. Z. 0001, and H.-J. Zhang. Multi-label sparse coding for automatic image annotation. In *CVPR*, pages 1643–1650. IEEE, 2009.

[22] X. Zhu, J. Lafferty, and Z. Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In *ICML 2003 workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, pages 58–65, 2003.

[4] http://opencatalog.darpa.mil/MEMEX.html

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:06:01 UTC from IEEE Xplore. Restrictions apply.

Appendix NUNES02

MARC View

LEADER 01238nas a2200301 a 4500				
001 47479703				
003 SIRSI				
005 20250514090049.0				
006 m				
007 cr n				
008 100304c20079999nyu|| p|s|||||||||0eng|d				
019			a| DO NOT EDIT this 360 MARC RECORD	
035			a| (WaSeSS)ssj0000329167	
040			a| WaSeSS c| WaSeSS d| WaSeSS	
050		4	a| HV6431	
245	0	0	a| CTC sentinel h| [electronic resource].	
246	1	3	a| Combating Terrorism Center	
260			a| West Point, N.Y. : b| Combating Terrorism Center	
310			a| Monthly	
362	1		a| Began with Vol. 1, issue 1(Dec. 2007).	
500			a| Description based on: Vol. 1, issue 1(Dec. 2007); title from volume contents page (publisher's website, viewed Mar. 5, 2010).	
500			a| Latest issue consulted: Vol. 3, issue 1 (Jan. 2010) (viewed Mar. 5, 2010).	
525			a| Includes special issues.	
538			a| Mode of access: World Wide Web.	
650		0	a| Terrorism x| Prevention v| Periodicals.	
856	4	0	u| http://SK8ES4MC2L.search.serialssolutions.com/?V=1.0&L=SK8ES4MC2L&S=JCs&C=TC0000329167&T=marc	
910			a| Serials Solutions original	
949			a| Electronic resource w| ASIS c| 1 i| 47479703-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 5/17/2025 o| .STAFF. 360 MARC	

ASK A LIBRARIAN

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

CTC sentinel [electronic resource].

Additional Titles: Combating Terrorism Center

Published: West Point, N.Y. : Combating Terrorism Center

Access Online

serialssolutions.com

ASK A LIBRARIAN

Dates of Publication and/or Sequential Designation
Began with Vol. 1, issue 1(Dec. 2007).

Subject(s)
Terrorism—Prevention—Periodicals

Publication Frequency
Monthly

Note
Description based on: Vol. 1, issue 1(Dec. 2007); title from volume contents page (publisher's website, viewed Mar. 5, 2010).
Latest issue consulted: Vol. 3, issue 1 (Jan. 2010) (viewed Mar. 5, 2010).

Supplement Note
Includes special issues.

Technical Details
Mode of access: World Wide Web.

View MARC record | catkey: 47479703

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES
Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES
Facebook
Twitter
Instagram

All Journals Only Books Only Videos Only

Title begins with Search

Show more search options ▼

Showing results 1 through 1 of 1

for the search: "TC0000329167"

Refine Results: All | Journals Only | Books Only | Videos Only

Limit by: Peer Reviewed | Open Access

CTC sentinel

2007 to 2020 in Columbia International Affairs Online Journals

Back to Top

Copyright ©2025 The Pennsylvania State University.
All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Last Updated on 06-17-2025 © Ex Libris, a ProQuest Company, 2000-2025

COMBATING TERRORISM CENTER AT WEST POINT

CTC SENTINEL

OBJECTIVE · RELEVANT · RIGOROUS

Contents

FEATURE ARTICLE

1 **The Risks Posed by Jihadist Hackers**
By Christopher Heffelfinger

REPORTS

5 **Fire as a Weapon in Terrorist Attacks**
By Joseph W. Pfeifer

8 **The Syrian Spillover and Salafist Radicalization in Lebanon**
By Bilal Y. Saab

12 **Anti-Muslim Attacks in Myanmar Threaten Uptick in Regional Violence and Islamist Activism**
By James Brandon

16 **Zaynab's Guardians: The Emergence of Shi` a Militias in Syria**
By Christopher Anzalone

21 **Al-Qa` ida in the Islamic Maghreb's Tunisia Strategy**
By Aaron Y. Zelin, Daveed Gartenstein-Ross and Andrew Lebovich

26 **Recent Highlights in Terrorist Activity**

28 **CTC Sentinel Staff & Contacts**

About the CTC Sentinel

The Combating Terrorism Center is an independent educational and research institution based in the Department of Social Sciences at the United States Military Academy, West Point. The CTC Sentinel harnesses the Center's global network of scholars and practitioners to understand and confront contemporary threats posed by terrorism and other forms of political violence.

The views expressed in this report are those of the authors and not of the U.S. Military Academy, the Department of the Army, or any other agency of the U.S. Government.

The Risks Posed by Jihadist Hackers

By Christopher Heffelfinger



THE CURRENT POOL of jihadist hackers (or jihadist hacktivists[1]) is youthful, ambitious in its goals, and largely lagging in terms of its technical capabilities. This is best illustrated by the fact that these hackers have carried out few effective large-scale attacks to date. Jihadist hacktivists remain a loosely to disorganized set of individual hackers who form and disband hacking groups they create, and frequently enter into counterproductive rivalries with fellow hackers. Perhaps as a result, despite more than seven years of efforts to construct and recruit for jihadist hacking attacks via online forums,[2] they have yet to form a jihadist hacking group that can demonstrably perform effective cyber attacks.

There are a range of skillsets, leadership abilities and ideologies among jihadist-inspired hacktivists, and some individual hackers have carried out small- to medium-scale cyber attacks against U.S. government and private sector targets, with moderate impact in terms of data loss and exposure. Those attacks also provided jihadist hacktivists with clout and a media platform (often predominately social

1 A hacktivist is a hacker who performs cyber attacks for a movement or cause. The U.S. Computer Emergency Response Team (US-CERT) includes in their definition of hacktivist: "Hacktivists form a small, foreign population of politically active hackers that includes individuals and groups with anti-U.S. motives," although the term does not explicitly connote an anti-American agenda. For the US-CERT definition, see "Cyber Threat Source Descriptions," U.S. Department of Homeland Security, May 2005.

2 Calls for jihadist hacking units and electronic mujahidin armies have been circulating in al-Qa` ida magazines since late 2005.

Appendix HEFFELFINGER02

media) from which to promote their message. The range of ideological beliefs among jihadist-inspired hackers is varied; some hacking groups embrace hard line militant Islamist imagery and messaging (such as that of al-Qa`ida), while concurrently incorporating the imagery, ethos and slogans of secular hacking collectives such as Anonymous into the informational aspects of their cyber attack campaigns.

This article evaluates existing jihadist cyber attack capabilities, offers a case study on a leading pro-jihadist hacktivist, and examines the rise in interest in cyber attacks among proponents of jihadist activism generally. It finds that although cyber attacks are becoming a more common and desirable means of furthering the global jihadist agenda, the overall impact and sophistication of jihadist hacktivists' attacks have been relatively low and will likely remain as such in the near term.

Evaluating Existing Jihadist Cyber Attack Capabilities

In comparison to hackers and hacking groups sponsored or controlled by state actors, jihadist hacktivists are clearly behind in terms of the impact of their attacks, their diminished technical skillset, and their overall weak organizational and recruitment abilities. Their hacking activities frequently include website defacements (usually against poorly secured websites), wherein the attackers leave antagonistic imagery and comments on the victimized websites.[3] Yet the activities of some jihadist hacktivists indicate there is a gradual sophistication of attack modes and intended attack impacts, occurring alongside a growing contingent of young jihadist enthusiasts who see cyber attacks as an increasingly effective and relatively easy way to contribute to the liberation or support of "oppressed" Muslims around the globe, which can frequently fall under the designation of "cyber terrorism."[4]

While jihadist-themed cyber attacks have been modest and often rudimentary over the past decade, the advancement and ambitions of certain jihadist hacking groups, individual hacktivists and proponents of cyber jihad over the past one to two years give some cause for concern in this area, particularly as those adversaries are growing more adept at identifying vulnerabilities in U.S. and other government targets, as well as those in the private sector. Clearly, the damage caused by jihadist cyber attacks pales in comparison to those under state sponsorship (prominent cases of the latter include Stuxnet in Iran, the highly destructive Saudi Aramco malware attack, operations conducted by the Syrian Electronic Army, and various data breaches performed by the Chinese People's Liberation Army Unit 61398, among others). Jihadist cyber attacks also trail those of better known hacktivist groups such as Anonymous or LulzSec, both in terms of the volume of attacks, sophistication, and impact. Yet the prospect of jihadists conducting a high impact cyber attack—such as one against an industrial control system (ICS) target or a series of high profile financial attacks—should not be dismissed.

Junaid Hussain (aka TriCk): Pro-Jihadist Hacktivist, Cyber Criminal

One prominent jihadist-inspired hacktivist was Junaid Hussain. Born in 1994, Hussain founded the hacking group TeaMpOisoN. Between the ages of 13-17, Hussain was a highly active hacker using the online moniker "TriCk." Between 2010 and 2012, he targeted NATO, officials and agencies of the UK government and a United States emergency response call line, among others, carrying out cyber attacks that were typically loosely jihadist-themed and promoting the liberation of Muslims in Palestine, Kashmir and other Muslim-majority conflict zones. Hussain's attacks also included the publication of personally identifiable information on the leadership of the English Defense League (EDL), an "anti-Islamist extremism" group, in April 2011, and the theft of hundreds of Israeli credit card holders' data as part of "Operation Free Palestine" in November 2011.[5] He stated that he became political when he was 15-years-old, after "watching videos of children getting killed in countries like Kashmir and Palestine."[6] He described his actions on behalf of TeaMpOisoN as "internet guerrilla warfare."[7]

Hussain was sentenced on July 27, 2012, to six months imprisonment after he pleaded guilty in a London court to conspiring to commit a public nuisance between January 1, 2010, and April 14, 2012, and "causing a computer to perform a function to secure unauthorized access to a program or data" under the UK Computer Misuse Act.[8] A resident of Birmingham, England, Hussain had turned 18 shortly before the trial date in late June 2012.[9] His initial arrest—after years of maintaining his anonymity as a teenage hacker—was prompted by a phone-based hack and concurrent telephonic denial-of-service (DoS) attack targeting the Anti-Terrorist Hotline of the Metropolitan Police Service (MPS) on April 10-11, 2012.[10]

"Although cyber attacks are becoming a more common and desirable means of furthering the global jihadist agenda, the overall impact and sophistication of jihadist hacktivists' attacks have been relatively low and will likely remain as such in the near term."

3 Such imagery typically includes the attacking group's logo and/or a satire of the victim's logo, or statements criticizing the victim's security.

4 There is often a thin line between cyberterrorism and online activism for humanitarian or social causes. As with kinetic activities, however, terrorism is distinguished by causing harm to another person or their property in the act of conveying a message, or political agenda.

5 Hannah Furness, "Team Poison: Profile of the Hackers," *Telegraph*, April 12, 2012.

6 Eduard Kovacs, "Hackers Around the World: It's No TriCk, He's Among the Best in the UK," Softpedia, February 18, 2012.

7 Ibid.

8 Caroline Grant, "Hacker's Facing Jail Over Blair Email Raid," *The Sun*, June 30, 2012.

9 Tammy Hughes, "'Team Poison' Hacker, 18, who Published Tony Blair's Address Book Online Faces Jail," *Daily Mail*, June 30, 2012.

10 On the evening of April 12, 2012, TriCk and another teenager aged 16 (who was later released on bail after police said they did not believe him to be a member of TeaMpOisoN) were arrested by police at a residence in Birmingham, in West Midlands. See "Two Arrested After Hackers Attacked Anti-Terror Hotline," *Telegraph*,

Following Hussain's arrest, his previous hacking activities carried out under the name TriCk and his group, TeaMpOisoN, were handled by the Police Central eCrime Unit, which investigates major cyber crimes.[11] This was due in part to the fact that Hussain had, in June 2011, breached the personal e-mail account of Katy Kay, a former special adviser to Prime Minister Tony Blair, and stole home addresses, phone numbers and e-mail addresses of Tony Blair, his wife and sister-in-law, as well as the personal information of other relatives, friends, and contacts in the House of Lords and Parliament.[12]

Shortly after the telephonic DoS attack on April 11, 2012, ended, Hussain posted a four-minute audio recording on YouTube entitled "Mi6: Counter Terrorism Command Phones Hacked - Leaked Call Discussing TeaMpOisoN," that contained an intercepted conversation among counterterrorism staff discussing the automated call "hoax."[13] During that recording, one employee of the counterterrorism office is heard telling a colleague that the anti-terrorism hotline had been inundated with "about 700 calls" from TeaMpOisoN over the previous two nights.[14] The employees also acknowledged that legitimate callers had been effectively denied access to the anti-terrorism hotline that TriCk and TeaMpOisoN had targeted.[15]

The phone-based attacks, which resulted in the breach and subsequent publication of sensitive conversations among British counterterrorism employees over the victimized phone lines, used a somewhat novel hacking technique that had been popular among the earliest generation of hackers in the 1980s known as "phreaking."[16] While Hussain did not disclose any further specifics about the method he and other TeaMpOisoN members claimed to use to record the phone call, he said "the conversation was tapped into via a private phreaking method, their phone system is old and we found a way to get in via basic but private phreaking technique."[17] As claimed, this hacking method likely enabled the hackers to eavesdrop on and record the conversations of officers in London's MPS.

The targeting of the hotline, Hussain said in an interview on April 11, 2012, occurred in retaliation to the fact that "the UK court system has extradited Babar Ahmad, Adel Abdel Bary and a few others" to face unfair treatment in the United States.[18] Babar Ahmad was allegedly involved in promoting militant jihadist materials online through a prominent website called "Azzam Publications."[19] Another of the five men to be extradited was a well-known radical cleric named Abu Hamza al-Masri who had established links to known militant groups including al-Qa`ida.[20]

Implications of TeaMp0isoN's Phone Hacking

TeaMpOisoN attempted a similar telephonic DoS attack on the 10th anniversary of the 9/11 attacks, but it failed, possibly due to the fact that it relied on a participatory model that required individual volunteers to work, and it ultimately lacked adequate numbers of supporters for the DoS attack to have any substantial impact.[21] In the successful telephonic DoS attack on the MPS Anti-Terrorist Hotline in April 2012, however, the calls were made by an automated caller program (which continuously repeated the phrase "Team Poison" in a computer-generated voice), using a compromised server based in Malaysia[22] running Asterisk software.[23] That attack also demonstrated an ability to learn from past mistakes and deliver an improved attack mechanism within a short-time frame on the part of TriCk and supporting TeaMpOisoN members.

In addition to affecting the MPS' anti-terrorism hotline over a two-day period, another significant aspect of the operation from a capabilities perspective is that Hussain was able to eavesdrop on sensitive, confidential phone conversations among counterterrorism and law enforcement officials. That capability was illustrated by the leaked recordings on YouTube and further evinced when MPS officials acknowledged it in the media—with the implication being that other hacktivists, including those directly supporting militant groups such as al-Qa`ida, could use it for counterintelligence purposes.[24] While such a capability may pose a risk to the security of law enforcement and other government agencies' communications if used effectively by adversaries, gaining highly-sensitive data from this tactic is challenging and uncommon, making it unlikely to be employed in a widespread fashion.[25]

Similar Groups and Offshoots

Around TeaMpOisoN emerged a number of like-minded hacking groups (in addition to several others that formed organically, unrelated to TriCk or TeaMpOisoN) that have carried out similar jihadist-oriented cyber attack campaigns. These groups are often composed of a majority of young Sunni Muslims with membership and support from non-Muslims in various countries.[26] ZCompany Hacking Crew (ZHC), which began in June 2010 as a spin-off hacking collective from TeaMpOisoN, aims to "end injustice, extremism, Zionism, illegal occupation" and other "evils," with a primary focus on Kashmir and Palestine.[27] In a December 2011 manifesto, it called for

April 12, 2012.

11 Hughes.

12 "'Team Poison' Hacker Who Posted Tony Blair's Details is Jailed," *Telegraph*, July 27, 2012.

13 Martin Evans, Christopher Williams and Hannah Furness, "Two Arrested after Hackers Attacked Anti-Terror Hotline," *Telegraph*, April 12, 2012.

14 Ibid.

15 Ibid.

16 Phreaking is the act of hacking into or manipulating a telephone network. Popularized in the 1980s, it is often considered the precursor to computer hacking.

17 "Hackers 'Eavesdrop on Terror Line,'" *Guardian*, April 11, 2012.

18 Evans et al.

19 "Babar Ahmad and Abu Hamza Among Terror Suspects to be Sent to US," BBC, October 5, 2012.

20 Ibid.

21 Message posted by TeaMpOisoN members announcing "Op911" on September 7, 2011, on pastebin.com, a site commonly used by hacktivists to post campaign messages.

22 Paul Roberts, "UK Teen, TeaMpOisoN Member, Arrested For 'Phone Bomb' Attack," ThreatPost.com, April 17, 2012; "Hackers 'Eavesdrop on Terror Line.'"

23 Asterisk software is an open source software product that enables users to convert a Linux-run server into a VoIP (Voice over Internet Protocol) telephone exchange.

24 Evans et al.

25 Jeremy Kirk, "How Anonymous Hacked the FBI-SOCA Conference Call," IDG News Service, March 6, 2012.

26 As observed from group membership and participation in social media during 2011-2012.

27 These details are from a manifesto posted by ZHC members on pastebin.com and various social media outlets in December 2011.

Appendix HEFFELFINGER02

members and supporters to "hack USA websites against Quran burning/draw Muhammad (PBUH) or for protesting against the killings of innocents in IRAQ, AFG and Pakistan" and to "hack France websites for protesting ban on *hijab*."[28] The following month, the group launched a widespread campaign (which was again primarily composed of defacement attacks) against Western targets called "Operation 1M_vs_NATO." On January 9, 2012, it claimed to compromise credit card holders' data from servers located in the United States, United Kingdom and Australia. The attackers released a list of targeted servers on Pastebin,[29] but did not readily disclose the method used in the alleged credit card theft. While it may have been exaggerated or ineffective, their claim nonetheless marked ZHC's first declared foray into causing financial loss as a facet of their attacks.[30]

"These attacks have resulted in low- to medium-level data and privacy loss, but a significant distance remains between jihadist hacktivists' demonstrated abilities and the capability to conduct an effective cyber attack on critical infrastructure."

One curious offshoot from TeaMpOisoN was "PoisAnon," a collaboration between TeaMpOisoN and purported Anonymous members that emerged in late 2011 to carry out shared cyber operations such as "OpCensorThis" and "OpRobinHood." During the week of December 1, 2012, as part of "OpRobinHood," PoisAnon identified and published a purported SQL[31] vulnerability in the First National Bank of Long Island website, then demonstrated the same SQLi (SQL injection) vulnerability on a webpage belonging to the BCD Credit Union in the United Kingdom.[32] A TeaMpOisoN member also claimed to identify a similar vulnerability on the website of the National Bank of California.[33] Yet in these cases, the attackers published statements saying that they would not release innocent peoples' credit card information, instead urging them to withdraw their money from those and other financial institutions.[34] Despite demonstrating a capability, the group members shied away from publishing breached credit card data, perhaps out of caution. More significant than the attacks, however, is the amalgamation of the sometimes jihadist-themed TeaMpOisoN and Anonymous, through a shared ethos of aiding the oppressed and confronting corrupt governments. Given how such a movement could appeal strongly to younger recruits, it could be an early indication of what the next, younger generation of jihadist enthusiasts-cum-hacktivists looks like.

TeaMpOisoN and ZHC have also been affiliated with smaller offshoot hacking groups such as the Mujahideen Hacking Unit (MHU) and Muslim Liberation Army (MLA) that had some overlapping membership with TeaMpOisoN and ZHC. These groups have largely been composed of young Pakistani Muslims espousing a hard line Salafi-jihadi message and functioned more as propaganda units than hacking cells, carrying out low level website defacements and other generally low impact activities.[35] In addition to these groups, individual hacktivists have, at times, had an impact on the security environment.

Jihadist DDoS Attacks Gain Global Attention

On January 16, 2012, a Saudi hacker with the moniker "OxOmar" conducted a high profile distributed denial-of-service (DDoS)[36] attack on the websites of Israeli national airline El Al and the Tel Aviv Stock Exchange.[37] The DDoS attack also targeted the websites of three Israeli banks that same day.[38] The DDoS attacks did not impact trading on the stock exchange, nor did it affect the operations of the airline; only the front-facing websites victimized in the attack were temporarily inaccessible.

OxOmar claimed a group called "Nightmare" assisted him in the attacks, but little information is available to confirm that such a hacking group existed that was in contact with him at the time of the attacks. The DDoS attack tool used by OxOmar on January 16 was not disclosed. The affected websites were largely restored to normal operations within one business day, but the media impact of the attacks was felt throughout the Middle East and bolstered other pro-jihadist and Islamist-inspired hacking groups.

One prominent Muslim cleric who commented on the media coverage was Kuwaiti imam Dr. Tariq al-Suwaidan, who also hosts a popular television show. One day after OxOmar's DDoS attacks on Israeli targets, al-Suwaidan posted on his Twitter account, which had some 240,000 followers, a call "to unify the efforts of [Muslim] hackers in the endeavor of electronic jihad against the Zionist enemy."[39] Al-Suwaidan is also a leading member of the Kuwaiti Muslim Brotherhood, and in May 2007 he was listed as an unindicted co-conspirator in the U.S. Department of Justice's case against the Muslim-American charity, the Holy Land Foundation.[40]

28 Ibid.

29 Pastebin.com is a website that allows users to store text for a certain period of time. The website is mainly used by programmers to store pieces of source code, but is open to any user to paste any type of text.

30 The theft of databases of credit card holder information is not uncommon among hacktivists. Many employ programs to scan for vulnerabilities in SQL or XSS to exfiltrate sensitive customer data.

31 Structured Query Language, or SQL, is a programming language designed for managing data in relational databases.

32 "First National Bank of Long Island, Operation Robin Hood Victim," Softpedia, November 29, 2011. The original statement is online at www.pastebin.com/g0Ckrq3u.

33 The claim was made on a TeaMpOisoN member's Twitter page.

34 Ibid.

35 As observed from group membership and participation in social media during 2011-2012.

36 A DDoS attack employs a botnet of compromised or voluntary machines as "bots," which simultaneously send requests to a specified server and, if successful, render it unresponsive.

37 Gianluca Mezzofiore, "'Nightmare' OxOmar Hackers Attack Israel's Stock Exchange and El Al," *International Business Times*, January 16, 2012.

38 "'I Want to Harm Israel,' Saudi Hacker Tells 'Post,'" *Jerusalem Post*, January 16, 2012.

39 In addition to his actual Twitter post, also see "Kuwaiti Imam: Cyber Jihad Effective," ynetnews.com, January 18, 2012.

40 *U.S.A. v. Holy Land Foundation for Relief and Development*, Northern District of Texas, 2007.

Appendix HEFFELFINGER02

Perhaps unsurprisingly, al-Suwaidan's call to a unified cyber jihad against Israel produced little tangible effect. Similar calls to a unified jihad have been made among real-world Islamist activists for years, yet they have never managed to overcome fragmentation and in-fighting. This is in part because the jihadist hacktivist community, like its kinetic counterpart, is prone to decentralization, which explains why it has been unable to consistently mount high impact cyber attacks, whether through DDoS tools or exploiting code vulnerabilities and performing data breaches.

Conclusion

To date, jihadist hacktivists and hacking collectives sympathetic to jihadist causes have largely used unsophisticated attack methods. These include brute force cracking to hack into e-mail and web servers and other basic techniques such as using open source hacking software that scans for vulnerabilities, or programs that run pre-programmed exploits. Pro-jihadist hackers have also used DDoS attacks (with occasional effectiveness, notably in the case of the Saudi hacker 0xOmar) and limited spear-phishing[41] and other social-engineering-based network intrusion techniques. These attacks have resulted in low- to medium- level data and privacy loss, but a significant distance remains between jihadist hacktivists' demonstrated abilities and the capability to conduct an effective cyber attack on critical infrastructure, or even those with significant financial cost.

A number of "hacking units" or "cyber armies" incorporating al-Qa`ida's name or identifying as aiding militant jihad have emerged (or at least announced their formation) in recent years, yet so far none have managed to gain traction, garner much media attention, or carry out a significant attack. They also appear to lack any skilled membership. While some on jihadist forums have called for attacks on critical infrastructure targets, no specific or viable plots have emerged against them from any jihadist-affiliated actors.[42] Nonetheless, the interest in such high profile, high impact attacks remains among jihadist hacktivists and proponents of Islamist militancy. As the pool of jihadist hacktivists continues to grow, and some advance to more sophisticated attack tools and methods, the possibility of an effective cyber attack emerging from among these actors becomes more likely.

The continuance of vulnerable attack targets and the likely increase in Islamist hacking activity in the near term combine to form a potentially challenging security environment for U.S. and other Western governments and private companies. Many of these potential targets, however, can mitigate the impact of cyber-terrorists—whether they are jihadist hacktivists or hackers from a collective like Anonymous—by taking additional steps to safeguard the integrity of their data and their customers' information, thereby reducing the media attention such attackers seek to exploit in pursuit of their agenda and message campaigns.

Christopher Heffelfinger is President of AVH, LLC. Based in Washington, D.C., he is also author of Radical Islam in America: Salafism's Journey from Arabia to the West *and a former Fellow of the Combating Terrorism Center's FBI program.*

41 "Spear-phishing" is a phishing attack against a specific target, rather than a general population, typically with the aim of gaining access to a secured network.

42 On June 11, 2011, in a leading jihadist forum, "Yaman" posted a highly detailed message calling for a "Center for Electronic Terrorism." A top priority for this center, he described, is the targeting of "SCADA [supervisory control and data acquisition] systems to distort the companies of electricity, gas, water, airports, trains, subway trains and central control systems" in the United States, United Kingdom and France. Yaman claimed the endeavor was a "new center for Qa`idat al-Jihad."

Fire as a Weapon in Terrorist Attacks

By Joseph W. Pfeifer

This article is adapted from the author's testimony on "Protecting the Homeland Against Mumbai-Style Attacks and the Threat from Lashkar-e-Taiba" presented to the U.S. House Committee on Homeland Security's Subcommittee on Counterterrorism and Intelligence on June 12, 2013.

THE USE OF fire for criminal, gang, and terrorist activities, as well as targeting first responders, is not new. During the past four decades, the New York City Fire Department (FDNY) has faced hundreds of intentionally set fires that would often target people. On March 25, 1990, however, the unthinkable happened. An arsonist, with a plastic container of gasoline, spread the fuel on the exit stairs of the "Happy Land Night Club" in the Bronx intentionally killing 87 people, foreshadowing even larger events to come.

The attacks of September 11, 2001, are remembered as the first to employ airplanes as weapons of mass destruction, resulting in the deaths of almost 3,000 people. It was the resultant fires, however, that brought down Towers 1 and 2 of the World Trade Center in the deadliest attack on U.S. soil. Seven years later, in what is described as a "paradigm shift," 10 terrorist operatives from Lashkar-i-Tayyiba (LeT) carried out attacks over three days in Mumbai, India, in November 2008, using a mix of automatic weapons, explosives and fire.[1] Each of these attacks is remembered for something other than fire, yet in each it was the fire that complicated rescue operations and drastically increased the lethality of the attacks.

A full understanding of fire as a weapon and implications for response are essential for homeland security, as it requires new policies and partnerships to address this emerging threat. Fire is an attractive weapon for terrorists for several reasons. Igniting a fire requires little to no training. Fire and associated

1 Michael R. Bloomberg et al., *FDNY Counterterrorism and Risk Management Strategy* (New York: New York City Fire Department, 2011).

Appendix HEFFELFINGER02

smoke can penetrate defenses with alarming lethality. Fire makes tactical response more difficult. The images of fire also increase media coverage, capturing world attention.[2] The FDNY has been studying this terrorist trend closely and, as a result of those efforts, is leading the national fire service on this issue.

Security personnel and emergency responders must rethink the way that they prepare and respond to incidents and anticipate the use of fire as a weapon, especially when combined with other attack methods. This article examines the terrorist use of fire as a weapon, the complexities of responding to multi-modality attacks involving fire, and the role the FDNY can play in national homeland security efforts.

Understanding Fire as a Weapon

The devastating 2008 attacks in Mumbai, India, represented a game-changer. Over three days, a city of nearly 14 million was held hostage while 166 people were murdered in multiple locations across the city, introducing a new model for terrorist attacks. The nature of the Mumbai attack confused those providing tactical response, rescue operations, fire extinguishment and mass casualty care. The attackers employed multiple means of attack, including: improvised explosive devices, assassination, hostage barricade, building takeover, active shooter, kidnapping and fire. Despite all of the violence, the most iconic images from that event remain the fire at Taj Mahal Hotel. The pictures of people hanging out of the windows of the hotel to escape the fire are reminiscent of 9/11.

Brian Jenkins notably stated in 1974 that "terrorist attacks are often carefully choreographed to attract the attention of the electronic media and the international press…Terrorism is theater." Directing the Mumbai attack from Pakistan, the mastermind asked the terrorists, "Are you setting the fire or not?"[3] He understood that the fire would capture the attention of the television cameras outside the hotel and would create an image the world would watch. In this case, fire was used as a strategic weapon. Yet it also created a condition that complicated the rescue planning and challenged the first responders to deal with not only an active shooter threat inside a hostage barricade situation, but also one where fire and smoke created a second layer of obstacles to the rescue force—one for which they were not prepared.

On September 11, 2012, the first murder of an American ambassador since 1988 took place in Benghazi, Libya. Although firearms, IEDs and military ordinance were used, it was not bullets or explosives that killed the U.S. ambassador, but rather smoke from an arson fire. During the attack on the U.S. mission in Benghazi, which killed four Americans,[4] terrorists reportedly linked to Ansar al-Shari`a and al-Qa`ida in the Islamic Maghreb (AQIM) used fuel from jerry cans to start a fire in the main villa, where Ambassador Christopher Stevens was sheltering in the designated location with two members of his diplomatic security detail. As the three men attempted to escape the untenable atmosphere—filled with choking, blinding smoke—the ambassador was separated from the one member of the detail who was able to escape through a window. Unfortunately, Ambassador Stevens and the other agent did not follow. Similar to 9/11 and Mumbai, the world was left with another image of a building ablaze during a terrorist attack. Following this incident, similar arson attacks took place days after Benghazi against the UN Multinational Force in the Sinai Peninsula as well as at the U.S. Embassy in Tunis, Tunisia.

"A full understanding of fire as a weapon and implications for response are essential for homeland security, as it requires new policies and partnerships to address this emerging threat."

While successful attacks are instructive, it is equally important to study unrealized terrorist plots as they reveal a great deal about adversary intentions, motivations, target selection and desired tactics.

- Arriving in the United States from the United Kingdom, al-Qa`ida operative Dhiren Barot carried out reconnaissance for terrorist attacks in New York City and Washington, D.C. Part of his research focused on exploiting building vulnerabilities, including gaps in fire protection. He determined that he could cause significant damage to the Prudential Building in Newark, New Jersey, and the Citi Corp Building in New York by ramming a loaded gas tanker truck into the lobby and then igniting the fuel.

- Another al-Qa`ida operative, Brooklyn-born Jose Padilla, determined that a "dirty bomb" attack might be too difficult to execute, so instead he planned to set wildfires, as well as ignite high-rise buildings by damaging the gas lines in apartments.

- An al-Qa`ida cell in the United Kingdom researched means to disable fire suppression systems to increase the impact of a plot that was ultimately disrupted by authorities.

These failed plots point to a strong interest in the use of fire as a weapon by terrorist groups and those they influence. In its widely disseminated English-language *Inspire* magazine, al-Qa`ida in the Arabian Peninsula (AQAP) has repeatedly urged aspiring homegrown violent extremists to carry out low tech, high impact attacks in the United States or other Western countries. In one issue of *Inspire*, the readers were introduced to various methods of conducting an attack, including the use of simple "ember bombs" to ignite forest fires.[5] Equally important, the images from attacks like Mumbai serve as a model for others to follow.

2 The images of buildings on fire with people trapped at the windows captured the world's attention and provided a dramatic backdrop to the terrorist actions.

3 "Keep Fighting But Don't Be Taken Alive: Handlers to 26/11 Attackers," *Times of India*, July 6, 2012.

4 Ambassador Christopher Stevens and information management officer Sean Smith were killed inside the main villa by smoke and fire. Former Navy SEALs Tyrone Woods and Glen Doherty were killed elsewhere during the attacks.

5 "Inspire Magazine: The 'Vanity Fair' of Terrorism," CBS, May 4, 2012.

Appendix HEFFELFINGER02

These events reveal that a group does not need a great deal of training to conduct a dramatic terrorist attack. In April 2013, two men at the Boston Marathon killed three people, injured 275 others and paralyzed the city. The Boston attacks serve as an important reminder that attacks need not be sophisticated to be deadly. Indeed, a survey of terrorists' attack plots in the United States over the past decade reveals a trend remarkable for the simplicity of attack plans. Fire as a weapon, by itself or along with other tactics, presents significant challenges that first responders and security forces must contend with in planning, preparation and drills.

Complexities in Responding to Multi-Modality Attacks Involving Fire

FDNY research and preparedness efforts on fire as a weapon have centered on what is now known as the "Mumbai-style attack method." The salient features of a Mumbai-style attack include:

- multiple attackers,
- multiple targets and
- multiple weapon types (guns, explosives and fire)
- deployed over a prolonged operational period leveraging media attention to amplify the effects of the attack.[6]

These factors create unique challenges for first responders beginning with the ability to quickly and accurately gain situational awareness of the nature and extent of the attack, the need for several command posts to address multiple attack sites and tactics, and techniques and procedures to deal with attacks deploying both fire and other attack modalities (e.g., active shooter).

Fire presents a qualitatively different type of attack when used in conjunction with other attack means. Fire, and its associated smoke, can prove disorienting to a responding force, inhibit ingress to the target, create structural dangers and potentially increase the number of casualties that the security forces will encounter while trying to resolve the situation. These factors present significant challenges to counterterrorism operations.

To address these complex challenges, the FDNY has reaffirmed its relationships with established partners like the NYPD, and forged new partnerships that add essential expertise to develop effective techniques, tactics and procedures. The results of these initiatives are jointly published intelligence bulletins, forward-looking joint exercises and information exchanges that are pushing response models forward.

Several partnerships are worthy of mention: FDNY began meetings with FBI's New York SWAT team to explore the idea of joint tactical teams simultaneously facing armed terrorists, fire and smoke, victims and mass casualties. Discussions and tabletop exercises led to two full-scale exercises that tested this concept. The insights gained from this one-year collaboration with the FBI culminated in the *Interagency Tactical Response Model* released in June 2012.

In May 2012, FDNY began collaboration with a group from the U.S. Army that specialize in rapid solutions to current and anticipated problems on the battlefield. As with the FBI, a series of meetings, training modules and tabletop exercises led to the group's February 2013 "Red Team" paper on *Fire and Smoke as a Weapon*, envisioning a Mumbai-style attack in a hypothetical Manhattan office building in an attempt to gauge emergency responder preparedness related to this novel attack method.

After the Benghazi attacks, the Department of State's Diplomatic Security Service leveraged the FDNY to provide advice to its high-threat response team—the Mobile Security Deployment. Diplomatic Security Service agents were briefed on the most critical features of fire as a weapon. Agents were then put through firefighting training at the FDNY training academy, including extrication of fortified vehicles and a walk-through exercise of a Mumbai-style scenario.

Finally, the FDNY has worked closely with the London Fire Brigade on counterterrorism measures since the 7/7 bombings in 2005. In preparation for the 2012 Olympics, FDNY discussed with London's fire service and the Metropolitan Police Service possible response scenarios to active shooter attacks involving fire in multiple locations.

Leading Role of FDNY in National Homeland Security Efforts

As consumers of intelligence, and the first line of defense when terrorist attacks occur, emergency responders require the best intelligence to carry out their duties across all mission areas. The understanding of the threat environment drives training initiatives, general awareness, safety protocols, operating procedures and risk management.

The fire service, however, is more than a consumer of intelligence. It is also a producer of intelligence as a non-traditional intelligence partner to the intelligence community. Firefighters and emergency medical personnel offer unique perspectives to more established intelligence partners and law enforcement, adding richness and insights in the understanding of the vulnerabilities and consequences related to varying threat streams. For more than five years, the FDNY has produced a weekly intelligence product called the *Watchline*, balancing a strategic focus with operational relevance to its primary readership: emergency responders. Fire service intelligence serves not only the response community but its intelligence partners with the delivery of tailored intelligence on the latest threats, trends, events and innovations that affect these groups, including the use of fire as a weapon on the world stage.

FDNY has also sent one of its officers to the National Counterterrorism Center (NCTC) on a one-year detail where the officer not only receives the latest intelligence and threat data, but also provides the intelligence community with fire service subject matter expertise on a broad range of issues related to emergency responders. NCTC has committed to providing first responders with the best threat intelligence so they can operate safely in performing their life saving mission, and recognizes the intrinsic value of this non-traditional partnership.

In addition, the FDNY collaborates with other partners throughout the intelligence community on the production of intelligence products. In

6 Fire Department in the City of New York, "Defining a Mumbai-style Attack," *Fireguard*, April, 2011.

Appendix HEFFELFINGER02

May 2012, the Department of Homeland Security Office of Intelligence and Analysis released *Terrorist Interest in Using Fire as a Weapon*, written in close consultation with the FDNY. Key findings centered on the advantages of using fire over other terrorist tactics, potential for mass casualties, economic damage and emergency resource depletion.

Working with the Department of Defense's Combating Terrorism Technical Support Office and New Mexico Tech's Energetic Materials Research and Testing Center, the FDNY wants to examine the vulnerability of high-rise building fire suppression systems. This interagency group hopes to construct a fire protection system and building mock-up for the purpose of testing blast effects on standpipes and sprinklers. Test results could then be used to inform first responders, Homeland Security and the State Department of the level of vulnerability of a combination attack of IEDs and fire.

Conclusion

This type of interagency and international collaboration by the FDNY demonstrates the importance of multi-agency solutions to these complex problems. In an era of ever constraining resources, it is critical that organizations such as the FDNY leverage their expertise to support broader audiences in the face of a dynamic and resilient enemy. The recognition of terrorists' interest in the use of fire as a weapon and the resulting complexities are important considerations for all first responders and security forces.

Assistant Chief Joseph W. Pfeifer is the Chief of Counterterrorism and Emergency Preparedness at the Fire Department of the City of New York. He is a non-resident senior fellow at the Combating Terrorism Center at West Point.

The Syrian Spillover and Salafist Radicalization in Lebanon

By Bilal Y. Saab

CONFLICT WITH THE Sunnis, who constitute the majority of all Muslims in the Middle East and around the world, is a scenario that severely undermines the long-term interests of Lebanese Hizb Allah. Not only would such a religious war be a costly distraction from the military struggle against Israel, but it would also be a strategic blunder because of its likely effects of endangering Hizb Allah's Shi`a support base and consequently threatening the organization's existence.[1] For the past decade, however, Hizb Allah has failed to fend off the specter of sectarian war, provoking and alienating the Lebanese Sunni community.[2] Yet it is Hizb Allah's military intervention in Syria, designed to prevent the collapse of an allied Syrian government and to maintain vital supply lines, that constitutes the most serious and immediate action that could precipitate Sunni-Shi`a conflict in Lebanon.

While it is evident that Sunni-Shi`a tensions in Lebanon and the Middle East have risen partly because of the Syrian crisis, it is less clear how the Syrian spillover has radicalized larger parts of Lebanon's Sunni Islamist community.[3] Sunni militancy is no stranger to Lebanon's landscape, having existed in several forms since the 1970s. Over the years, it has sporadically caused bouts of violence and criminal activity in the northern region and in the Palestinian refugee camps.[4]

The deadliest encounter between Lebanese authorities and Sunni militancy occurred in the summer of 2007 when the Lebanese army was forced to destroy the Palestinian refugee camp of Nahr al-Bared to eradicate the threat of Fatah al-Islam, a Salafi-jihadi group that was growing in size and influence in the northern part of the country.[5] The most recent violent clash between the Lebanese army and Salafi-jihadis, however, occurred in Abra, a neighborhood in the southern city of Sidon, in late June 2013.[6] Lebanese Salafist Shaykh Ahmad al-Assir, along with several of his armed followers, ambushed a Lebanese army checkpoint on June 23, causing a two-day battle that killed 17 Lebanese soldiers and dozens of al-Assir's gunmen.[7] Al-Assir's fighting force was ultimately defeated and the Lebanese army took control of his security compound, but the shaykh himself supposedly managed to escape. His whereabouts are still unknown.[8]

This article provides an assessment of the effects of the war in Syria on the growth trajectory of Lebanon's Salafi-jihadis, clarifying the old and new actors, their clout in the north and in the Palestinian refugee camps, and their military involvement in Syria. It also looks specifically at the case of Ahmad al-Assir, assessing the implications of his recent rise and fall in Abra. The article finds that despite increased

1 Regional demographics have worked against Hizb Allah and the Shi`a. Therefore, even the staunchest Lebanese Shi`a supporters of Hizb Allah would prefer to be at peace with Sunni Lebanese and their broader environment, which is predominantly Sunni.

2 These actions include removing the Lebanese Sunni community's leader, Saad Hariri, from power in a so-called coup in 2011, and allegedly participating in a spate of assassinations of its most influential figures including former Prime Minister Rafik Hariri (the father of Saad) in February 2005 and Brigadier General Wissam al-Hassan, chief of the Intelligence Bureau of the Lebanese Internal Security Forces, in October 2012. See "Hariri Supporters Accuse Hezbollah of Lebanon Coup," Reuters, January 24, 2011; "Hezbollah Suspects to be Tried Over Rafik Hariri," BBC, August 17, 2011; "Hezbollah-led Govt Blamed for Murder of Top Lebanese Security Official," al-Arabiya, October 20, 2012.

3 All officially recognized sects that compose Lebanese society are considered minorities. Indeed, there is no religious or communal majority in Lebanon. It is difficult to determine more precise numbers on sectarian representation in Lebanon because the last time the country had an official census was in 1932 due to political sensitivities.

4 Bilal Y. Saab, "Lebanon at Risk from Salafi-Jihadi Terrorist Cells," *CTC Sentinel* 3:2 (2010); Bilal Y. Saab, "The Failure of Salafi-Jihadi Insurgent Movements in the Levant," *CTC Sentinel* 2:9 (2009).

5 Saab, "Lebanon at Risk from Salafi-Jihadi Terrorist Cells"; Thanassis Cambanis, "Desolation Awaits Returning Palestinian Refugees," *New York Times*, October 11, 2007.

6 Mohammed Zaatari and Antoine Amrieh, "Sidon Leaders Demand Post-Abra Probe," *Daily Star* [Beirut], June 29, 2013.

7 Ibid.; Anne Barnard, "Calm Returns to Lebanese City, but Tensions Remain," *New York Times*, June 25, 2013.

8 Ibid.

Appendix HEFFELFINGER02

Sunni-Shi`a tensions and political polarization in Lebanon due to the Syrian crisis and sectarian violence in the Middle East, and despite a leadership void in the Lebanese Sunni political class, Salafi-jihadis still do not enjoy a popular following among Lebanese Sunnis. Indeed, their presence consists of cells and small groups, not a large and armed social movement that has attained insurgent status. Nevertheless, Hizb Allah's intervention in Syria is a powerful catalyst for the radicalization of larger parts of the Sunni Islamist and specifically Salafist communities in Lebanon. This radicalization process could threaten Hizb Allah and potentially bring about Sunni-Shi`a conflict in Lebanon.

Evidence of Jabhat al-Nusra Activity in Lebanon

As the war in Syria rages, there is growing evidence of Salafi-jihadi activity in Lebanon. *Al-Monitor* reported in April 2013 that "it can no longer be denied that Jabhat al-Nusra[9] has found fertile ground in the Palestinian refugee camps in [Lebanon], among the nearly one million Syrian refugees there, as well as in Lebanese Sunni areas, especially in northern Lebanon near the Syrian border."[10] In July 2013, Lebanon's military charged six alleged members of Jabhat al-Nusra with forming an armed gang with the intent to conduct terrorist attacks in Lebanon.[11]

Five months earlier, in February 2013, Lebanon's *al-Akhbar* newspaper published a report claiming that Salafi-jihadis in Lebanon could be close to establishing a Jabhat al-Nusra branch in the country.[12] *Al-Akhbar* journalist Nasser Sharara claimed that "H.A. Dargham," a Lebanese man, proclaimed to his followers in February that the creation of the Lebanese branch of Jabhat al-Nusra was drawing near.[13] According to the report, the site of Dargham's pronouncement was in the town of Arsal, located in Lebanon's Bekaa Valley near the Syrian border, a Hizb Allah stronghold.[14] According to the Lebanese military intelligence services, who were cited in the *al-Akhbar* report, Dargham is also an associate of Khaled Hamid who is reported to be a major "logistical facilitator" between Jabhat al-Nusra in Syria and its allies in Lebanon's north and in the Palestinian refugee camp of Ain al-Hilwah in Sidon, including Fatah al-Islam[15] and Jund al-Sham.[16]

The idea of a Lebanese branch of Jabhat al-Nusra reportedly started with the Saudi national Majid bin Muhammad al-Majid, the *amir* of the Abdullah Azzam Brigades.[17] In June 2012, the Abdullah Azzam Brigades announced that al-Majid was their leader, and that the group supports the revolution in Syria.[18] Yet the *al-Akhbar* report in February 2013—which has not been corroborated by other sources—claimed that al-Majid traveled from Ain al-Hilwah to Syria in late 2011 along with cadres from Fatah al-Islam and Jund al-Sham for the purpose of reportedly dethroning Abu Muhammad al-Julani, the head of Jabhat al-Nusra in Syria, and replacing him as the *amir* (leader).[19] A sizeable but unspecified number of his comrades, however, defected from his ranks during travel for unknown reasons, and joined al-Julani instead, forcing al-Majid to return to Ain al-Hilwah.[20] According to the report, these skilled fighters later trained al-Julani's forces, ultimately helping to make his Jabhat al-Nusra the most effective and resourceful actor in the Syrian militant opposition.[21]

If a Lebanese branch of Jabhat al-Nusra were to be formed today, it would likely be composed of two main battalions: an Ain al-Hilwah battalion and a Tripoli battalion.[22]

The Ain al-Hilwah battalion would probably include the remaining cadres and fighters of Fatah al-Islam and Jund al-Sham, the Abdullah Azzam Brigades and the Platoons of Ziad al-Jarrah.[23] The gradual weakening of the secular Palestinian party Fatah in Ain al-Hilwah has allowed these Salafi-jihadi entities to increase their presence and mobilize greater numbers of men from the al-Ta'mir area of the camp, also at the expense of older and larger Sunni Islamist Palestinian factions such as Asbat al-Ansar,[24]

9 Jabhat al-Nusra is a Salafi-jihadi insurgent group operating in Syria that has established ties to al-Qa`ida and has been designated a "terrorist" group by the U.S. State Department. The U.S. government views Jabhat al-Nusra as a "front group" for al-Qa`ida in Iraq.

10 "Hezbollah Prepares for Attacks by Jabhat al-Nusra in Lebanon," *al-Monitor*, April 26, 2013.

11 "Six Men Charged for Plotting Terror Acts in Lebanon," *Daily Star*, July 19, 2013. For other evidence of possible Jabhat al-Nusra activity in Lebanon, see Jean Aziz, "Jabhat al-Nusra Reportedly in Lebanon," *al-Monitor*, December 24, 2012; Mitchell Prothero, "Al Qaida-linked Nusra Front Rebels Blamed for Bloody Fight Against Lebanese Army in Sidon," McClatchy Newspapers, June 26, 2013.

12 Nasser Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon," *al-Akhbar*, February 25, 2013; Nasser Charara, "Coming Soon: Al-Nusra Front in Lebanon," *al-Akhbar*, February 25, 2013; Kassem Kassem and Amal Khalil, "Ein el Helweh fi Intithar wiladat fir' li 'Jabhat al Nusra,'" *al-Akhbar*, February 4, 2013; Ja`far al-Attar, "Al Assir...Al Nusra wal Qa'ida," *as-Safir*, June 24, 2013.

13 Ibid.

14 Ibid.

15 Fatah al-Islam is a militant Sunni Islamist group that is inspired by al-Qa`ida's ideology. Its members are mostly Arabs from various Middle Eastern countries. It emerged in the Nahr al-Bared Palestinian refugee camp in Lebanon in November 2006. Its goals are unclear but include the establishment of an Islamic state in northern Lebanon.

16 Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon." Jund al-Sham is a title claimed by several Sunni Islamic extremist entities, all or none of which may be connected. These entities mostly operate in Lebanon, Syria and Jordan, and their goals include the establishment of an Islamic caliphate throughout the Levant.

17 Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon"; Bill Roggio, "Abdullah Azzam Brigades Names Leader, Advises Against Attacks in Syria's Cities," *The Long War Journal*, June 27, 2012.

18 Roggio.

19 Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon."

20 Ibid.

21 Ibid.

22 These two battalions do not exist in name. The author contends that if a Lebanese branch of Jabhat al-Nusra—or some other coordinated, jihadist entity—were to be formed in Lebanon, then its fighters would likely coalesce into two separate factions/battalions.

23 The "Platoons of Ziad al-Jarrah" is often referred to as a brigade that operates under the command of the Abdullah Azzam Brigades. See, for example, "Who Are The Abdullah Azzam Brigades?" Reuters, August 4, 2010.

24 Asbat al-Ansar is a Palestinian Salafi-jihadi group that was involved in a number of terrorist operations against Lebanese official targets in the past, including the killings of four judges in a courtroom in Sidon in June 1999. Recently, it reached a permanent truce with the Lebanese authorities in return for its intelligence cooperation on al-Qa`ida elements in the Ain al-Hilwah camp. For details of the incident in 1999, see Audrey Kurth Cronin et al., "Foreign Terrorist Organizations," CRS Report for Congress, February 6, 2004.

Appendix HEFFELFINGER02

Ansar Allah,[25] Hamas, and al-Haraka al-Islamiyya al-Mujahida,[26] which are described by these extremist groups as more "moderate" and insufficiently committed to supporting the jihad in Syria.[27]

As for the "Tripoli battalion," it would likely comprise militants who follow Lebanese Sunni Shaykh Hussam Sabbagh.[28] Sabbagh has reportedly helped smuggle militants across the border into Syria, and he has fought with Jabhat al-Nusra rebels against the al-Assad regime in Syria.[29] According to Lebanon's *Daily Star*, "The Nusra Front and other Syrian Islamist fighting groups now use him [Sabbagh] as their foremost representative in Lebanon, and [Sabbagh] coordinates between various groups who hope to establish an Islamic emirate in Tripoli."[30] He has approximately 250 followers.[31]

Fatah al-Islam Still Active

Since Jund al-Sham and Fatah al-Islam would likely be incorporated into a Lebanese branch of Jabhat al-Nusra—or another Salafi-jihadi entity in Lebanon—it is worth assessing their current capabilities. Other than being accurately labeled by Lebanon commentators as freelance jihadist fighters lacking any organizational structure or *modus vivendi*, not much else is known about the current state of Jund al-Sham. The group used to be part of the larger and much more influential Asbat al-Ansar until several members defected due to differences over ideology, politics, and jihadist strategy and priorities.[32] Today, Jund al-Sham likely consists of a small group of fighters who are "looking for action and trouble and the right opportunity to merge with a like-minded movement," as one Lebanese internal security officer told this author over the telephone in May 2013.[33]

It is Fatah al-Islam that may have to do the heavy lifting should Salafi-jihadis in Lebanon coordinate their activities. Knowing the devastating losses it suffered in the battle of Nahr al-Bared in the summer of 2007, one would think that Fatah al-Islam is defunct or poses a marginal security threat.[34] That would be a false conclusion. Although the group's organizational structure was crushed, most of its members killed or jailed (some fled), and its size and influence much reduced, it still has a presence in Lebanon in the form of scattered cells, most notably in the north, as well as in Ain al-Hilwah and ironically Roumieh, Lebanon's largest prison.[35]

"As the war in Syria rages, there is growing evidence of Salafi-jihadi activity in Lebanon."

There are a number of ways that the group has managed to survive and stay active during the past couple of years—even from behind bars in Roumieh. Of the 480 people suspected of involvement in the Nahr al-Bared battle, more than 200 are being held in Roumieh.[36] Members of Fatah al-Islam, who number 88 according to court reports and the Lebanese Judicial Council, are among the most problematic and powerful prisoners at Roumieh.[37] Small in number compared to others, they are the strongest physically, and the most influential and resourceful.[38]

Perhaps more consequential than Fatah al-Islam's presence inside Roumieh is the group's ability to communicate with Salafi-jihadi fighters operating in Lebanon and Syria, as well as manage its remaining cadre. Although prisoners in general have mobile telephones at their disposal in Roumieh, Islamists, and Fatah al-Islam members particularly, carry laptops, allowing them to publish statements on Islamist online forums.[39]

Several escapes and foiled attempts by ordinary criminals and Salafi-jihadis from Roumieh have been reported over the years. Perhaps the single most important incident, other than the case of Walid al-Bustani,[40] occurred in mid-2012, when five prisoners—four of whom were Fatah al-Islam members—escaped using sheets to scale the prison wall. Muhammad Abdullah al-Dousari, also known as "Abu Talha al-Kuwaiti," led the group of four and was described by Lebanese media in December 2012 as "the ambassador of al-Qa`ida in Lebanon."[41] The three others were Abdullah al-Shukri and Abdul Aziz al-Masri from Syria and Abdul Nasser Sanjar from Lebanon.[42] Al-Dousari's group may have been ambushed by the Syrian military, with one member killed and the others reportedly arrested.[43]

25 Ansar Allah was founded in southern Gaza in November 2008 as an armed Sunni Islamist group with strong Salafi-jihadi credentials. Its goals include the establishment of an Islamic state in Gaza.

26 Al-Haraka al-Islamiyya al-Mujahida is a Salafi-jihadi group that was founded in 1984 by Jamal Khattab and Abdallah Hallaq, allegedly with the help of Iran. Its base is the Palestinian refugee camp of Ain al-Hilwah. Khattab has good relations with Hizb Allah and Hamas.

27 Kassem and Khalil.

28 Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon"; Misbah al-Ali, "Who is Sabbagh? A Look into the Life of the Sheikh and Fighter," *Daily Star*, January 15, 2013. Sabbagh has called for an Islamic emirate that unites all Salafi-jihadi entities operating in Lebanon and Syria under one umbrella and covers geographical areas from northern Lebanon and the outskirts of the Syrian city of Homs.

29 Al-Ali.

30 Ibid.; Sharara, "Jabhat al-Nusra From Northern Syria to Southern Lebanon."

31 Ibid.

32 Roger Hardy, "Profile: Jund al-Sham," BBC, June 4, 2007. For more on Jund al-Sham, see Bilal Y. Saab and Magnus Ranstorp, "Securing Lebanon from the Threat of Salafist Jihadism," *Studies in Conflict & Terrorism* 30:10 (2007).

33 Personal interview, Lebanese internal security officer, May 2013.

34 Saab, "Lebanon at Risk from Salafi-Jihadi Terrorist Cells."

35 Ibid.; Ja`far al-Attar, "Lebanon Prison 'Controlled by Islamists,'" *as-Safir*, January 30, 2013.

36 Kassem Kassem, "Ein el Helweh: Al Lino ya'oud li sahk al Nusra," *al-Akhbar*, February 20, 2013.

37 Kassem and Khalil.

38 Ibid.; Al-Attar, "Lebanon Prison 'Controlled by Islamists.'"

39 Ibid.

40 One notable example of communication and coordination between imprisoned Fatah al-Islam members and like-minded extremists from the outside is the case of Walid al-Bustani. Al-Bustani, a Salafi-jihadi who fought in the Nahr al-Bared battle, escaped from Roumieh in mid-2012, but was later executed by Syrian rebels in the town of Tel Kalakh. For details, see "Fatah al-Islam Militants Killed in Syria," *Daily Star*, April 24, 2012; Al-Attar, "Lebanon Prison 'Controlled by Islamists'"; Radwan Mortada, "Roumieh Prison: Rule by the Sword," *al-Akhbar*, October 14, 2012.

41 "Kuwaiti Tried to Set Up Al-Qaeda camp – Report," *Daily Star*, May 27, 2009.

42 Ibid.

43 Doha Shams, "Abu Talha wal mashrou` al Qa'idi fi Lubnan," *al-Akhbar*, December 27, 2012.

Appendix HEFFELFINGER02

Another significant escape from Roumieh by Fatah al-Islam members occurred in September-October 2012. The Lebanese press reported that Omar Othman from Syria, Faisal Aqla from Algeria, and Mahmoud Fallah from Palestine managed to escape from the prison and join their comrades in Syria.[44] A couple of months later, Roumieh's administration reported that prison guards had foiled another major group escape attempt by more than 20 Fatah al-Islam detainees.[45]

In addition to those who have escaped, some of the Islamist detainees in Roumieh who were eventually freed have also opted to commit violent acts against the Syrian government in Syria. The case of Khaled Mahmoud, once one of the leaders of Fatah al-Islam, is well known in Lebanese circles.[46] Mahmoud was released in June 2012 despite his militant activities against the Lebanese army in Nahr al-Bared. Six months after his discharge, he appeared in an online video surrounded by militants and declared the establishment of the Syrian version of Jund al-Sham.[47]

Mahmoud is believed to be responsible for dispatching Lebanese youths to Syria for militant aims.[48] According to reports from *al-Akhbar*, Lebanese security reports indicate that "Yahia J.," a close associate of Mahmoud, is in charge of recruiting and deploying Salafi-jihadis to Syria, including the group that was reportedly ambushed by the Syrian army in Tel Kalakh in November 2012.[49] Such reports suggest that Yahia, who is based in Tripoli, along with "Nader H." and "Bashir M.," are actively involved in recruiting Salafi-jihadi cells.[50] The latter two were previously detained on charges of belonging to Fatah al-Islam and spent one and three years, respectively, in prison.[51]

According to the Lebanese media, Mahmoud, the *amir*, reportedly entered Syrian territory through the Mashari al-Qaa region of eastern Lebanon.[52] From there, he traveled to Syria where he became the commander of an armed group stationed at a historic crusader castle in the Krak des Chevaliers area.[53] Shortly afterward, more than 20 Lebanese Salafi-jihadis joined Mahmoud there from Tripoli, swearing allegiance to him.[54] According to the *Washington Post*, Mahmoud is linked to the Lebanese Shaykh Sabbagh, who has been reported by Lebanese media to be recruiting youths in Tripoli to fight in Syria.[55]

The Rapid Rise and Fall of Ahmad al-Assir

The case of Ahmad al-Assir sheds further light on the evolving state of Sunni militancy in Lebanon, and the outcome of the battle he waged against the Lebanese army carries important implications for the growth trajectory of Salafi-jihadism in Lebanon.[56]

The fact that al-Assir's men lost the battle of Abra in Sidon in June 2013 may suggest that Salafi-jihadism in Lebanon has suffered a major blow.[57] Unfortunately, that is only partially true because al-Assir emerged under crisis conditions that are still very much relevant in today's domestic and regional environment: a raging civil war in Syria pitting Sunnis against Alawites and spilling over into Lebanon, extremely tense Sunni-Shi`a relations in Lebanon, severe political instability in Beirut, and a leadership void within the Lebanese Sunni community that has allowed radical Sunni elements to assume, by default, more prominent sociopolitical roles at the expense of secular figures. As long as these conditions exist, another al-Assir could arise.

"There is a sense of anxiety in Lebanon about the increased potential for Sunni-Shi`a conflict."

Three important observations on the outcome of the battle of Abra, however, support the argument that Salafi-jihadism, despite all the external factors enabling its growth and expansion, still faces major challenges in finding a solid popular base and a permanent home in Lebanon.

First, despite two public calls by al-Assir during the battle for Sunnis to join him in the fight and for Sunni soldiers to defect from the Lebanese army, the Salafist shaykh's requests fell on deaf ears.[58] This same scenario happened in the battle in Nahr al-Bared in 2007 when Fatah al-Islam was forced to fight the Lebanese army alone. Militants from Ain al-Hilwah were supposed to lend their full support to Fatah al-Islam in the past and al-Assir at present, but militants in the camp remained relatively calm, distancing themselves from the fighting, despite a few minor attacks against the Lebanese army in the Ta'amir area by elements of Jund al-Sham and remnants of

44 "Judge Orders Roumieh Warden Detained," *Daily Star*, October 15, 2012.

45 "Police Thwart Roumieh Escape Attempt by Suspected Islamists," *Daily Star*, December 13, 2012.

46 Babak Dehghanpisheh and Suzan Haidamous, "More Lebanese Sunnis are Crossing into Syria to Aid Rebellion, Officials Say," *Washington Post*, January 26, 2013.

47 Kassem; Radwan Mortada, "Khaled Mahmoud: A Prisoner in Lebanon Turns Emir in Syria," *al-Akhbar*, January 4, 2013; Dehghanpisheh and Haidamous.

48 Ibid.

49 Ibid.

50 Mortada, "Khaled Mahmoud: A Prisoner in Lebanon Turns Emir in Syria."

51 Ibid.

52 Doha Shams, "Layali al Shamal al Hazina Tatarakkab al-`Arka al Kabira," *al-Akhbar*, February 25, 2013; Dehghanpisheh and Haidamous.

53 Shams, "Layali al Shamal al Hazina Tatarakkab al-`Arka al Kabira."

54 Ibid.

55 Ibid.; Dehghanpisheh and Haidamous.

56 Ahmad al-Assir rose to prominence a year ago when he and his followers blocked a main road leading to southern Lebanon for a month and a half, a move that allegedly was part of a broader strategy for creating a Sunni militia that would ultimately face Hizb Allah. The Salafist shaykh began his mobilization campaign by recruiting young Lebanese using a highly sectarian and anti-Hizb Allah speech. He attracted Lebanese nationals, Palestinian refugees, Syrians, and other Arabs to his anti-Shi`a cause. Once established in Sidon, he visited other areas including Arsal and Tripoli, hoping to expand his movement and reach the northern Sunni part of the country. From June 2012 to June 2013, he reportedly benefited from funding provided by Gulf and local, unknown sources. He allegedly used the money to buy weapons from Ain al-Hilwah. For details, see Amal Khalil, "Al Sheikh Assir Wal Thalath Sa'at Al 'Ro'b Fi Janoub Loubnan," *al-Akhbar*, June 20, 2013; Dana Moukhallati, "Breaking Down Ahmad al-Assir: The Man Behind the Beard," al-Arabiya, June 25, 2013; "What Happened to Fadel Shaker?" *al-Monitor*, July 4, 2013.

57 Ibid.

58 Mohammad Saleh, "Al Assir You'lin 'Harbahou' 'ala al Jaysh," *as-Safir*, June 24, 2013.

Appendix HEFFELFINGER02

Fatah al-Islam led by Bilal bin Badr.[59] Less radical Islamist actors who are in disagreement with Fatah al-Islam and Jund al-Sham—including the influential Asbat al-Ansar, Hamas, and al-Jihad al-Islami—assisted the Lebanese army and worked toward preserving the relative peace in Ain al-Hilwah.[60] These actors, along with the secular Fatah, played the same role in 2007 as a rational attempt to protect their own interests.

Second, the battle of Abra would have been the perfect opportunity for the so-called Lebanese branch of Jabhat al-Nusra to be established. Such a development, however, did not occur, and the reasons are unclear. Perhaps the project was not ready for implementation, or it was too ambitious. Regardless, it raises questions about the prospects, or even existence, of this potentially new Salafi-jihadi enterprise in Lebanon.

Third, despite their anger at Hizb Allah over its military involvement in Syria, major secular political figures in the Lebanese Sunni community—including Saad Hariri and Fouad Siniora—condemned the attack in Abra and sided with the Lebanese army and state institutions.[61] Equally important, the most influential Salafist leader in Lebanon, Shaykh Da'i al-Islam al-Shahal (son of Salem al-Shahal, the founder of the Salafist movement in Lebanon), remained silent and did not lend his full support (at least publicly) to al-Assir, making it clear that the Salafist community in Lebanon is not united in its stance against the Lebanese army. Indeed, neither Beirut's nor the north's Salafists erupted during or after the fighting, and al-Assir's supporters in Tripoli refrained from starting a diversionary battle and opening another front against the Lebanese army to potentially come to the rescue of their "champion."

59 Ibid.

60 Ibid.

61 Ibid.

Conclusion

From Arsal to Wadi Khaled, from Tripoli to Akkar, and from Sidon to the heart of Beirut, black Salafi-jihadi flags and banners have been spotted in increasing numbers, a picture unseen before in Lebanon's history.[62] The Lebanese people used to take lightly the tirades of al-Assir against Hizb Allah and the Shi`a. This is no longer true, given the Salafist shaykh's growing but still relatively small support base. There is a sense of anxiety in Lebanon about the increased potential for Sunni-Shi`a conflict. Hizb Allah is on edge and high alert, and despite its strategic interest in avoiding civil unrest in Lebanon, the organization, if it feels threatened, could lash out against fellow Lebanese, as it did in May 2008.[63]

This tense and sectarian environment in Lebanon, should it explode, is a development that the Lebanese army, due to its modest resources and a divided political leadership in Beirut, is likely incapable of pacifying.

Bilal Y. Saab is Executive Director and Head of Research of the Institute for Near East and Gulf Military Analysis (INEGMA) North America. He is also a nonresident scholar at the Center for Nonproliferation Studies at the Monterey Institute of International Studies.

62 During the past year, Lebanese news agencies including al-Jadeed, MTV, and LBC have reported in their evening news on the Sunni Islamist and Salafist demonstrations and rallies in Lebanon's northern region and Beirut, showing footage of bearded men waving black flags of al-Qa`ida. Also see Dehghanpisheh and Haidamous.

63 In response to the Lebanese government's decision to shut down Hizb Allah's telecommunications network and sack airport security chief Wafic Choucair (a man close to Hizb Allah), the Shi`a group took over parts of West Beirut, a move described by its rivals in the pro-Western March 14 coalition as an "armed coup." See Robert F. Worth and Nada Bakri, "Hezbollah Seizes Swath of Beirut From U.S.-Backed Lebanon Government," *New York Times*, May 10, 2008.

Anti-Muslim Attacks in Myanmar Threaten Uptick in Regional Violence and Islamist Activism

By James Brandon

DURING THE LAST 18 months, a surge of communal violence between Muslims and Buddhists in Myanmar (formerly known as Burma) has killed at least 180 people and displaced more than 120,000, the vast majority of them members of the country's one million-strong Muslim minority.[1] These events, accompanied by instances of official government discrimination against Muslims, has prompted a range of attacks on Buddhists throughout the region, including an attack in July 2013 on the Bodh Gaya, India's most important Buddhist temple, and in Bangladesh, Indonesia and Malaysia. Myanmar's perceived mistreatment of Muslims has also galvanized a wide range of Islamist and jihadist groups from the Middle East to Australia and the United Kingdom.

This article provides background on militancy among Rohingya Muslims, highlights recent regional incidents of related anti-Buddhist violence, and examines the potential for a wider uptick in Islamist militancy over the issue. It finds that Islamist organizations, as well as jihadist groups, are increasingly incorporating developments in Myanmar into their propaganda campaigns, as well as including Myanmar's interests in their targeting selection.

Background

There are approximately one million Muslims living in Myanmar, a country of 55 million people.[2] Around 800,000 of these are from the Rohingya ethnic minority, a group concentrated in the country's southwestern Rahkine State (formerly known as Arakan State).[3] The Rohingya typically argue that

1 "Unpunished Crimes Against Humanity, Humanitarian Crisis in Arakan State," Human Rights Watch, April 22, 2013. Rahkine State was known as Arakan State until 1989.

2 Thomas Fuller, "Extremism Rises Among Myanmar Buddhists," *New York Times*, June 20, 2013.

3 Araminta Wordsworth, "Burmese Sleepwalking Into 'Ethnic Cleansing,'" *National Post*, July 13, 2013.

Appendix HEFFELFINGER02

they are indigenous to the area, but many Burmese claim that they are relatively recent immigrants from Bangladesh, where many Rohingya also live. Regardless, tensions between Buddhists and Muslims in Myanmar are long-standing.[4] The recent unrest, however, is the worst in at least a decade—involving widespread attacks on Muslim homes and businesses.

The violence has primarily been conducted by Buddhist civilians, although there are credible reports of the authorities turning a blind eye or tacitly supporting the violence.[5] There are examples of Buddhist monks (notably Ashin Wirathu from the Buddhist 969 movement) inciting hatred against Muslims; this intolerance can easily lead to violence over local issues. There are also allegations of some monks directly taking part in the violence.[6] The Myanmar government itself has imposed discriminatory measures against Muslims, notably a two-child limit on some Muslim families.[7] The unrest has been concentrated in the country's southwestern Rahkine State, as well as in central towns such as Meiktila and Shan State in the west. In contrast to previous outbreaks of conflict, however, news of the latest violence has impacted Muslims outside of the country. This is largely due to the growth of social media, which has dramatically raised awareness of the plight of Myanmar's Muslims. The country may also have gained a greater significance to the global Muslim community due to high profile U.S. and Western engagement with Myanmar during the last two years.[8]

"The recent uptick both in intra-communal violence against Muslims and the government's latest intolerance toward Muslims could prompt Islamist radicalization."

Rohingya Armed Groups

Myanmar has been challenged by a range of armed separatist groups representing Rohingya Muslims since its independence in 1948 when the self-described "Mujahidin," a term that embraced activists from a range of groups, began fighting for some form of self-government, ranging from autonomy to full independence. Since then, various armed groups have been periodically active, such as the Rohingya Liberation Party in the 1970s. All of these groups have had a limited impact, with most being broken up by Myanmar's efficient security apparatus or forced abroad.

The most active current group is the Rohingya Solidarity Organization (RSO) based largely in neighboring Bangladesh. The RSO, which has a limited capacity in Myanmar, is widely accused of working with radical Bangladeshi groups, including the two banned jihadist groups Jamaatul Mujahidin Bangladesh (JMB) and Harkat-ul-Jihad-al-Islam-Bangladesh (HuJI-B). The RSO has denied these links.[9] Separately, a range of media outlets and commentators have alleged that the RSO, along with the Arakan Rohingya National Organization and individual Rohingya, have had contact with foreign militant groups such as the Taliban.[10] These claims, however, are difficult to substantiate. Likewise, in July 2013, radical Islamist websites reported that a group of jihadists from Bangladesh, Indonesia and Myanmar had entered Myanmar and killed 17 Burmese soldiers and claimed that Rohingya Muslims were being trained by militants in Bangladesh; the claim was denied by Myanmar's government.[11]

More credibly, the RSO is accused of stoking anti-Buddhist violence in Bangladesh and of involvement with radical Bangladeshi groups.[12] For example, the group was accused of involvement in the September 2012 attacks in Ramu in Cox's Bazaar district in Bangladesh, in which mobs attacked and destroyed 12 Buddhist temples and more than 50 houses, in apparent revenge for attacks by Buddhists in Myanmar, although the unrest also seems to have been stirred up through social media.[13] In March 2013, Hafez Sanaul Islam, a senior RSO leader, was arrested by Bangladeshi police over the violence, along with several other Rohingya refugees, illustrating the potential for hard line Rohingya activists to use the issue to stoke violence abroad.[14]

While the RSO stages periodic attacks in Myanmar—for example, the government blamed the group for a minor cross-border attack from Bangladesh in November 2012—there is no evidence that militant Islamism has enjoyed significant traction among Rohingya. This may be because Islamist groups such as Jamaat-i-Islami,[15] which can create environments in which jihadist

4 Although the precise origin of the Rohingya is contested, it seems likely that while significant numbers of Rohingya have lived in the area for centuries, others arrived from Bangladesh during the British colonial period in 1826-1948.

5 "Unpunished Crimes Against Humanity, Humanitarian Crisis in Arakan State"; Kate Hodal, "Video Shows Burmese Police Standing by as Buddhists Attack Muslims," *Guardian*, April 22, 2013.

6 Hodal; Jared Ferrie, "Buddhist Mobs Attack Muslim Homes in Myanmar, One Dead," Reuters, May 29, 2013.

7 "What is Behind Burma's Wave of Religious Violence?" BBC, April 4, 2013; "One Region in Myanmar Limits Births of Muslims," Associated Press, May 25, 2013.

8 Previously, there was little awareness of Myanmar even from regional Islamist groups.

9 "RSO Press Release: Statement in Response of the News Reports of the Daily Star, Dhaka, Bangladesh and Indonesia," Rohingya Solidarity Organization, May 21, 2009. The Australian government has categorized the RSO with groups such as the Moro Islamic Liberation Front (MILF) in the Philippines that are "not primarily anti-Western" and are instead largely motivated by "local socio-political and economic grievances." See "Transnational Terrorism: The Threat to Australia," Australian Department of Foreign Affairs and Trade, 2004.

10 For example, in 2002 CNN said it had found an al-Qa`ida tape in Afghanistan that showed jihadists training in Burma in 1990. See "Exclusive Tapes Reveal al Qaeda's Capabilities," CNN, August 23, 2002.

11 "Jihad Started in Burma," Kavkaz Center, July 12, 2012; "Myanmar Bins Reports of Rohingyas' Military Training," *Daily Star* [Dhaka], July 13, 2013.

12 "Rohingya Rebels Trained JMB Men," *Daily Star*, May 19, 2009.

13 "Man Held for Link with Ramu Attacks," *Daily Star*, October 31, 2012; "Rohingya Accused of Sparking Bangladesh Riots," *The Irrawaddy*, October 1, 2012.

14 "Rohingya Groups Under Scanner," *Daily Star*, October 7, 2012; "Remand of Ramu Violence 'Plotter' Sought," BDNews24, March 23, 2013.

15 Jamaat-i-Islami was founded in 1941 in British-run India by Abu al-A`la Mawdudi, which broke in different, but closely-associated, national movements following the creation of India, Pakistan and Bangladesh as separate states.

Appendix HEFFELFINGER02

groups can flourish, have gained little purchase in the country. Similarly, Wahhabism and other forms of Salafi Islam have made few visible inroads among the Rohingya, and Islamic practices remain largely traditional, syncretistic, and Sufi-influenced. Additionally, the minimal numbers of Rohingya studying in the West have meant that radicalism has not been imported into the country by returning students.

Current dynamics, however, could change. Although the Myanmar government's harsh rule has historically quashed radical tendencies, the recent uptick both in intra-communal violence against Muslims and the government's latest intolerance toward Muslims could prompt Islamist radicalization. Conversely, however, political liberalization by Myanmar could give Islamists, and particularly non-violent political Islamists, more room to maneuver.[16]

In a possible sign of incidents to come, on July 21, 2013, a small bomb exploded near a Buddhist temple in Mandalay, Myanmar's second-largest city, only meters away from where the 969 leader Ashin Wirathu was speaking.[17] It was not immediately clear who was responsible for the blast.

Regional Islamist Reaction

In addition to the spillover violence in Bangladesh, Myanmar's treatment of Muslim civilians has prompted attacks further afield in both South and Southeast Asia, providing Islamist and jihadist groups with fresh grievances and a new cause with which to rally their followers.

In Indonesia, although overall support for jihadist groups is much reduced from the 1990s, recent events in Myanmar have reenergized the country's small and fragmented radical groups. In May 2013, for example, Indonesian police broke up a terrorist cell that was allegedly planning a bomb attack on the Myanmar Embassy in Jakarta, arresting several people and seizing bomb-making materials.[18] The police added that the suspects were believed to be linked to a previous December 2012 plot to attack U.S. diplomatic and commercial interests in Jakarta and Surabaya.[19] The arrests illustrated both the Indonesian government's ability to contain jihadist groups and an uptick in Islamist activism over the Myanmar issue.

Similarly, Abu Bakar Bashir, the jihadist cleric imprisoned for his role in the Bali bombings, issued an open letter to Myanmar President Thein Sein from prison in July 2012.[20] The letter characterized the actions of Burmese Buddhists as "barbarous" and described how "they burn the homes of Muslims, forbade [Islamic] worship and slaughter them like animals."[21] Bashir also warned that if Myanmar did not improve its treatment of Muslims, "the destruction of the lands [of Myanmar] at the hands of the mujahidin (with the permission of Allah) will take place."[22]

Echoing these sentiments, hundreds of hard line Islamists gathered outside the Myanmar Embassy in Jakarta in May 2013 to call for "jihad in Myanmar," carrying banners declaring "we want to kill Myanmar Buddhists" and "stop genocide in Myanmar."[23] Then, in August 2012, around 1,000 members of the Islamist group Hizb al-Tahrir marched in Jakarta "to show solidarity with the Rohingya who the extremist Buddhists are slaughtering, raping and torturing."[24]

Sporadic violence has also occurred. On April 5, 2013, Rohingya refugees from Myanmar killed eight Burmese Buddhists in a clash at an immigration detention center in Indonesia.[25] The Indonesian authorities said the attack began after Muslim detainees circulated photographs of the violence in Myanmar.[26]

A comparable uptick in tensions and related violence is also visible in normally peaceful Malaysia, with at least four Burmese Buddhists killed in a spate of attacks in the capital Kuala Lumpur in early June 2013.[27] Some reports blamed local Malay Muslims; others said that Rohingya refugees had carried out the attacks. Members of the 969 movement, a hard line Buddhist nationalist group blamed for much of the Myanmar violence, had reportedly visited Malaysia in the last year, which local activists said had contributed to an increase in tensions.[28] Following the June clashes, Malaysia arrested 900 Burmese nationals to preempt further violence, with a police spokesman saying the arrests were intended to "send a clear message to stop this nonsense and not bring the violence over to Malaysia."[29]

Meanwhile, in Pakistan, Tehrik-i-Taliban Pakistan (TTP) threatened to take "revenge" on Myanmar in July 2012. Although its spokesman Ehsanullah Ehsan warned that they would "attack Burmese interests anywhere," the group rarely carries out attacks abroad.[30] The statement was more likely an attempt to pressure the Pakistani government and to justify fresh attacks against it; Ehsan warned the government to halt relations with Myanmar and close Myanmar's embassy in Islamabad or face the consequences.[31] Likewise, on June 14, 2013, Jamaat-i-Islami held rallies in Lahore where the group's leaders reportedly told followers that thousands of Burmese Muslims were being "martyred" and that their mosques were being demolished.[32]

16 That being said, given the government's reluctance to even accept Rohingya claims to citizenship, this appears unlikely at present.

17 "Blast During Sermon by Radical Myanmar Monk Wounds 4," Reuters, July 22, 2013.

18 "Myanmar Embassy Bomb Plotter Used Facebook to Recruit Followers," *Jakarta Globe*, May 4, 2013.

19 "Bomb Plot Linked Known Cells," *Jakarta Post*, May 4, 2013.

20 "Ustadz Abu Bakar Ba'asyir Letter to the President of Myanmar," Arrahmah.com, August 2, 2012.

21 Ibid.

22 Ibid.

23 "Radical Rallies in Jakarta Call for Deadly Jihad in Myanmar," *Jakarta Globe*, May 3, 2012.

24 "Solidarity with the Rohingya Muslims by Hizb ut-Tahrir Indonesia," Khilafah.com, August 6, 2012.

25 "Myanmar Muslims Kill 8 Buddhists in Indonesian Centre," Agence France-Presse, April 5, 2013.

26 Ibid.

27 "Burmese Migrant Community in Malaysia Simmers after Attacks," *The Irrawaddy*, June 13, 2013.

28 Ibid.

29 "Malaysia Arrests 900 Myanmar Nationals in Wake of Violence," Agence France-Presse, June 6, 2012.

30 "Tehreek-e-Taliban Pakistan Threaten Myanmar over Rohingya," *Express Tribune*, July 26, 2012.

31 Ibid.

32 "Religious Parties Protest Against Muslim Killings Today," *The News International*, June 14, 2013.

Appendix HEFFELFINGER02

Violent protests have also occurred in India, notably in Mumbai in August 2012 when two protesters were killed and 50 injured.[33] Indian police said the violence was prompted by inflammatory SMS (cellular phone text messaging) videos of violence in Myanmar—an example of how social media has dramatically raised awareness of events in the country.[34] Separately, on July 7, 2013, nine small bombs exploded at India's most significant Buddhist temple, Bodh Gaya in Bihar Province, injuring two.[35] Although no official statements have yet been made on the motive, Indian media speculated that the attack was revenge for events in Myanmar and cited past threats by Lashkar-i-Tayyiba (LeT) founder Hafiz Saeed to target India over its relations with Myanmar.[36]

Global Islamist Reaction

Events in Myanmar have also provided a rallying cry for Islamist groups worldwide. Some groups seem to have eagerly seized on the issue as it offers a clear narrative of Muslims being victimized by non-Muslims. This is particularly true of the Muslim Brotherhood which, at least until the recent military coup in Egypt, was eager to deflect attention from its growing image problems in Egypt and elsewhere. For example, the Freedom and Justice Party (FYP), the Muslim Brotherhood's Egyptian wing, has repeatedly seized on the issue, describing it as "genocide against Burmese Muslims" and the "barbaric cleansing of more than a million Muslims."[37] In April 2013, the FYP's Ezzedin al-Komi, chairman of the Human Rights Committee of Egypt's Shura Council, described Myanmar's actions as "crimes against humanity, gross injustice and brutal genocide."[38] Gamal Heshmat, another senior FJP leader, suggested a wider conspiracy: "Why do some parties—certain world powers—not consider the tragedy of Muslims as they scrutinize most minority issues in the Muslim world?"[39] In other countries, there have been calls for violence. In Lebanon, for example, a prominent Salafist was quoted as calling for attacks on Buddhists: "Every person who can get to a Buddhist should kill him because they are killing our people."[40]

In Western countries, the same emotive propaganda intended to foment anger and a narrative of Muslim victimhood is also visible. In September 2012, a grouping of British Islamists linked to the Cordoba Foundation, a London-based Muslim Brotherhood "front" group,[41] organized an event entitled "Burma Bloodbath," at the East London Mosque, the UK headquarters of Jamaat-i-Islami.[42] Anas al-Tikriti gave a speech describing Buddhists as being "far more capable of bloodshed and violence than we ever gave them credit for" and urging the audience to "remember Bosnia."[43] Another speaker, Ufuk Secgin, a representative of the Turkish Islamist group Milli Gorus, also compared the violence to Bosnia, saying: "Will we learn the names of the towns of Burma and Arakhan as we learned the names of the towns of Bosnia?"[44] Separately, Alyas Karmani, a Bradford-based preacher widely employed by the UK government through the counterterrorism "Prevent" program, gave an emotive talk in Bradford describing Rohingya Muslims "crying tears of blood" and "being clubbed to death."[45] He blamed the root cause of the strife on British colonialism and on Muslims "rejecting their core values" and "abandoning their *din* [religion]," closing his talk by calling on Allah to provide *nusra* (victory), although he was careful to urge his audience to work peacefully through democratic channels.[46]

Also in the United Kingdom, Hizb al-Tahrir held a protest outside the Bangladeshi High Commission in London in August 2012. The event was, however, mainly directed at Bangladesh's secular government, a long-standing opponent of Hizb al-Tahrir's UK branch, and the West. "The West doesn't care about Muslims—they care about business deals," read one placard. Hizb al-Tahrir's central office issued a notably more hard line statement in May 2013, advocating the reestablishment of the caliphate as this "will cause nations to quiver at the mere thought of harming a single Muslim under their rule."[47]

"It is clear that Myanmar's treatment of its Muslim minority is rapidly becoming a mobilizing issue for Islamist and jihadist groups—both regionally and globally."

Further along the Islamist spectrum, groups such as the one run by Anjem Choudary (his group is currently nameless after its previous incarnations were banned) also agitated against Myanmar. On April 5, 2013, the group protested outside the Myanmar Embassy in London (as well as the Sri Lankan one[48]) against the "continuous

33 "Inflammatory SMSs, Pictures Behind Rioting?" *Times of India*, August 13, 2012.

34 Ibid.

35 "Blasts at Indian Buddhist Shrines of Bodh Gaya in Bihar," BBC, July 7, 2013.

36 "Buddhist-Rohingya Conflict in Myanmar Spells Trouble for India," *Times of India*, July 8, 2013; "Terror Strikes Bodhgaya, Two Monks Injured," *Indian Express*, July 7, 2013.

37 "Freedom and Justice Party Demands Immediate Halt of Genocide Against Burmese Muslims," Ikhwanweb.com, August 5, 2012.

38 "Freedom and Justice Party Demands Immediate Halt of Genocide Against Burmese Muslims," Ikwanonline.info, August 15, 2012; "Egypt Parliament's Human Rights Committee Urges International Protection for Burma Muslims," Ikhwanweb.com, April 16, 2013.

39 "Gamal Heshmat: Muslims of Myanmar Victims of Discrimination, Racism and Genocide," fjponline.com, March 28, 2013.

40 "Syrian Islamists Set Sight on Myanmar," Now Lebanon, April 12, 2013.

41 In 2008, David Cameron, who led the British parliament at the time, called the Cordoba Foundation a "political front for the Muslim Brotherhood."

42 "Burma Bloodbath: Symposium on the Human Rights Abuses Inflicted on the Rohingya Muslims," available at www.youtube.com/watch?v=m57w86J5Oik.

43 Ibid.

44 "Ethnic Cleansing of a Forgotten People – New Coalition Supports Burma's Rohingya Muslims," Rohingya Minority Crisis Group, September 7, 2012.

45 Alyas Karmani, "Tears of Blood (Massacre of Muslims in Burma)," August 15, 2012, available at www.youtube.com/watch?v=fo1g_Pa5TCY.

46 Ibid.

47 "The Fascist Burmese Regime Presents Birth Control of Rohingya Muslim Women as its latest Weapon of Ethnic Cleansing of Rakhine's Muslims," Hizb al-Tahrir, May 10, 2013.

48 There has been a similar uptick in Buddhist intolerance against Muslims in Sri Lanka, but at a much lower

Appendix HEFFELFINGER02

atrocities being committed against Muslim men, women and children…at the hands of Buddhist individuals and monks."[49] In Australia, Shaykh Shady Alsuleiman, a Salafist preacher at the controversial Lakemba Mosque, gave a sermon that "Muslims in Burma are being slaughtered."[50] In Myanmar, he continued, Muslim "blood was so cheap, it was like stepping on a cockroach" or "slaughtering a chicken."[51] He warned his audience, however, not to attack Buddhists in Australia as this "would not please Allah" and would "bring harm upon the Muslims."[52]

Conclusion

Although there is little evidence of Rohingya Muslims being involved in international terrorist groups, or any recent record of Rohingyas carrying out any sustained terrorist campaign within Myanmar, it is clear that Myanmar's treatment of its Muslim minority is rapidly becoming a mobilizing issue for Islamist and jihadist groups—both regionally and globally. Indeed, in the past year the issue has already generated violence overseas, including in Bangladesh, India, Indonesia and even in normally peaceful Malaysia. It has also attracted the attention of Islamist and jihadist groups from Afghanistan to Australia to the United Kingdom—a notable development given that the Rohingya issue had rarely featured on the radar of such groups previously.

There are several reasons for this. Western engagement with Myanmar has raised the country's profile and enabled groups such as the Muslim Brotherhood and Hizb al-Tahrir to link Western engagement with anti-Muslim violence. In addition, many such groups have a vested interest in raising the profile of the violence in Myanmar. For some beleaguered organizations such as the Muslim Brotherhood, Myanmar has been an opportunity to recast themselves as defenders of Islam. The conflict—which is typically presented by Islamists as a straightforward story of Buddhists victimizing innocent Muslims—is also much easier to market to core Islamist audiences than complex intra-Muslim conflicts in Syria, Iraq and Bahrain. In addition, anger at Myanmar (as evidenced by riots in Mumbai and the Indonesian detention center killings) has been repeatedly stoked by the deliberate circulation of forged or mislabeled photographs—as well as by Islamists' routine inflation of casualty figures.[53] Likewise, the characterization of the violence in Myanmar by a wide range of Islamist groups as a "slaughter," "massacre" and "genocide" has resonated in many Muslim communities, including among relatively more moderate groups.[54]

It is unlikely that Myanmar's treatment of the Rohingya will ever become a core issue for Islamist or jihadist groups globally. Nevertheless, one should not discount the power of a compelling and simple "good vs. evil" narrative—promoted by a broad range of Islamist groups and based on genuine instances of anti-Muslim discrimination and violence—to cause an uptick in Islamist violence and radicalization, within Myanmar, regionally, and further afield.

James Brandon is an associate fellow at the International Centre for the Study of Radicalisation and Political Violence (ICSR) at Kings College London.

level.

49 "Demonstration Outside the Sri Lankan and Burmese Embassies Tomorrow," anjem-choudary.com, April 2013.

50 "Burma: Where are the Muslims," August 3, 2012, available at www.youtube.com/watch?v=HAw4xFG5xS4.

51 Ibid.

52 Ibid.

53 For instance, some of the most widely shared photographs on social media sites which purport to show dead Rohingya actually show Indonesian victims of the 2004 tsunami.

54 In June 2013, for example, a Maryland representative of the Council on American Islamic Relations (CAIR), while making many accurate points about Myanmar's denial of equal rights to Rohingyas, himself described the killings of Muslims "as an ongoing genocide." See Saqib Ali, "In Burma, Let's Call it What It Is: Genocide," Council on American Islamic Relations, June 5, 2013.

Zaynab's Guardians: The Emergence of Shi`a Militias in Syria

By Christopher Anzalone

THE PUBLIC EMERGENCE of Twelver Shi`a[1] foreign fighter militias operating with Syrian government forces loyal to President Bashar al-Assad, together with the recent public admission by Lebanese Hizb Allah that it is also operating alongside them, is the latest in the increasing sectarianization of Syria's civil war. Sectarianism has long been a tool used for social and political mobilization by a variety of actors, and has historically been employed as much in struggles within the same community as in struggles between different communities.[2] Historically, sectarianism has been driven by politics and competition over group identity, and has been a part of social processes to mobilize large numbers of people against other groups. This remains true today. As conflicts break down along sectarian or ethnic lines, identifiers (differentiating one group from another)—such as religious affiliation, nationality, or tribe—become increasingly salient. Mobilization frames, which draw upon cultural

1 Shi`a Islam is divided into several different groups: 1) "twelver" or Imami Shi`a who believe in a line of twelve divinely-guided religious and temporal leaders, the imams; 2) Ismaili Shi`a, who believe in the same line of imams as the "twelvers" up until the sixth, Ja`far al-Sadiq, who died in the eighth century CE; and 3) Zaydi Shi`a, whose beliefs lie between Sunni and Shi`a. This article focuses exclusively on Twelver Shi`a, hereafter referred to simply as "Shi`a." There is no evidence of pro-regime Ismaili or Zaydi Shi`a militias operating in Syria. Yemen is the only country in the Arab world with a significant Zaydi Shi`a population, and Syria's Ismaili Shi`a have actually participated in peaceful anti-government protests. See Omar Hossino, "Salamiyeh Bombings Strike the Heart of Syria's Peaceful Revolt," Syria Deeply, February 14, 2013.

2 See Ussama Makdisi's seminal study on the political and social utilization of sectarianism in Ottoman Lebanon: Ussama Makdisi, *The Culture of Sectarianism: Community, History, and Violence in Nineteenth Century Ottoman Lebanon* (Berkeley: University of California Press, 2000); Marc Lynch, "The War for the Arab World," *Foreign Policy*, May 23, 2013; Geneive Abdo, "The New Sectarianism: The Arab Uprisings and the Rebirth of the Shi`a-Sunni Divide," Brookings Institution, 2013; Mariz Tadros, "Sectarianism and its Discontents in Post-Mubarak Egypt," *Middle East Report*, Summer 2011.

Appendix HEFFELFINGER02

idioms, are created and utilized to drive social mobilization.[3]

This article examines the gradual sectarianization of the Syrian civil war with a particular focus on the emergence, composition, mobilization frames and media campaigns of pro-Assad Shi`a militias.[4] Close attention is paid to the historical and cultural significance of the mobilization frames and idioms used to inspire support from a broad public, particularly Shi`a, for these groups' participation. Understanding these frames and their historical and cultural resonance, referred to as "frame resonance" in social movement theory literature,[5] is vital to comprehending the drivers of the mobilization and recruitment of Shi`a foreign fighters in Syria. It finds that these frames, in turn, are the central element at play in the formulation of a sectarian counter-narrative aimed at delegitimizing the Syrian opposition and Sunni rebel groups as well as attracting Shi`a foreign fighters from abroad to fight for al-Assad.

Gradual Sectarianization

Reports of Hizb Allah and other Shi`a militias' involvement in the war in Syria have led to new calls from influential Sunni religious leaders—such as the Qatar-based, Egyptian Muslim Brotherhood-affiliated Yusuf al-Qaradawi—for able-bodied Sunnis to travel to Syria to fight a military jihad against the Syrian government and its allies, chiefly Iran, Hizb Allah, and other Shi`a militias.[6] Prominent political actors and religious officials—including Egypt's Muslim Brotherhood and the grand imam of al-Azhar seminary in Egypt, Ahmed al-Tayeb—have also become increasingly blunt in their rhetoric toward Shi`a, who they accuse of supporting sectarianism in Syria.[7] The creeping sectarianization of the Syrian conflict is a windfall for the Syrian Islamic Front, an umbrella organization for a number of Syrian Salafist rebel groups.[8]

Iran and its regional allies, including Hizb Allah and Shi`a militias in Syria, are actively trying to sway Shi`a worldwide toward viewing the conflict in black-and-white terms, portraying all Syrian rebel groups as being virulently anti-Shi`a and in league with the "neo-imperialists," namely the United States, European powers, and Israel.[9] These attempts are meant to frame the Syrian conflict as one having dire repercussions for Shi`a worldwide, a messaging tool that has historically proven both vital and successful in the recruitment of foreign fighters and "international brigades."[10] Reports of attacks on Shi`a in Syria, the destruction of some Shi`a mosques and shrines or the tombs inside them, and claims that major Shi`a shrines such as those of Sayyida Zaynab and Sakina bint Husayn have been damaged are mobilizing Shi`a around the world to support the al-Assad regime.[11] These events have been continuously highlighted and spun in the news coverage in the hopes of further inflaming the public opinion of Shi`a generally.[12] Such attacks or even reports of them are issues of concern for many Shi`a, not only those supportive of Iran and its allies, and this has been a boon for the latter, making it more likely that their mobilization frames will resonate with their intended audience.[13]

The ongoing production of pro-Shi`a militia artwork and media such as poetry recitations (*anashid*) set to song as well as the increasing numbers of recruits and volunteers to fight in Syria are evidence of the resonance of these groups' mobilization frames.[14] While focusing on the Shi`a fighters' "defense" of shrines such as Sayyida Zaynab's, militia media operations conceal these fighters' reported involvement in military operations in other parts of the country.

3 For a succinct discussion of mobilization frames and their role in social movement mobilization, see Quintan Wiktorowicz ed., *Islamic Activism: A Social Movement Theory Approach* (Indianapolis: Indiana University Press, 2004).

4 This article focuses primarily on Shi`a militias excluding Lebanese Hizb Allah.

5 Robert D. Benford and David A. Snow, "Framing Processes and Social Movements: An Overview and Assessment," *Annual Review of Sociology* 26 (2000).

6 "Al-Qaradawi Calls Upon the Able-Bodied to Fight in Syria," Qaradawi.net, June 1, 2013; Griff Witte, "New Wave of Foreigners in Syrian Fight," *Washington Post*, June 21, 2013; "Sending of Egyptian Mujahideen to Syria Stirs Debate among Religious Scholars and Jihadi Organizations," *Azzaman*, May 31, 2013.

7 "Top Egypt Cleric Condemns 'Sectarian' Foes in Syria," Reuters, June 11, 2013; Lee Keath, "Hezbollah Entry in Syria Fans Shiite-Sunni Fires," Associated Press, June 7, 2013; Maggie Fick, "Egypt Brotherhood Backs Syria Jihad, Denounces Shi'ites," Reuters, June 14, 2013.

8 Aaron Y. Zelin and Charles Lister, "The Crowning of the Syrian Islamic Front," *Foreign Policy*, June 24, 2013.

9 Qays al-Khaz`ali, "Remarks of the Secretary General of the Ahl al-Haqq Islamic Resistance Movement on the Tenth Anniversary of its Founding," Asaib Ahl al-Haq, May 5, 2013.

10 David Malet, *Foreign Fighters: Transnational Identity in Civil Conflicts* (Oxford: Oxford University Press, 2013); David Malet, "Why Foreign Fighters? Historical Perspectives and Solutions," Foreign Policy Research Institute, 2010.

11 Ruth Sherlock, "Syrian Shias Flee to Lebanon to Escape Sunni Militias," *Telegraph*, May 1, 2013; Patrick J. McDonnell and Nabih Bulos, "Syria's Shiites Offer Different Picture of War," *Los Angeles Times*, March 26, 2013.

12 "Senior Shia Cleric Seyyed Naser al-Alawi Assassinated in Syria," AhlulBayt News Agency, April 14, 2013; "Holy Shrine of Hazrat [Holiness] Sakina (AS) Damaged by Terrorists in Syria," AhlulBayt News Agency, February 12, 2013; "The Islamic Resistance, `Asa'ib Ahl al-Haqq, Condemns the Heinous Crime Against the Shrine of the Prophet's Companion Hujr bin `Adi, may God bless him, in Syria," Asaib Ahl al-Haq, May 3, 2013; "Hizb Allah: Where are the Syrian Opposition [Members] Who Said They Wanted to Defend the Holy Sites?" al-Manar, May 2, 2013; "After Threats to Shrine, Iraqi Shiite Fighters Prepare for Sectarian Strife at Home, in Syria," Associated Press, October 25, 2012.

13 Attacks on Shi`a Muslims and shrines have been picked up by other Shi`a news sources. See, for example, "Head of Hawza E Zainabia Martyred by Target Killing in Syria," Jafria News, April 16, 2013; "Desecration of Shrine of Hujr bin Adi al Kindi," May 3, 2013, available at www.coej.org/secretariat/statements/2639-hujr-bin-adi-al-kindi-statement.

14 Suadad al-Salhy, "Iraqi Shi'ites Flock to Assad's Side as Sectarian Split Widens," Reuters, June 19, 2013. Popular reciters of *anashid* and *latmiyat* (mourning recitations used in Shi`a ceremonies to commemorate the martyred imams and other Shi`a historical heroes and heroines) include Lebanese *munshid* (reciter) Ali Barakat and Iraqi *munshids* Ali Abu Kiyan al-Muwali and Muhammad Abu `Izra'il al-Karbala'i. Barakat has recorded a *latmiyya* entitled "In God's Protection," dedicated to the "martyr-guardians of Sayyida Zaynab," in which "the Awaited One's (Imam al-Mahdi) soldiers" are said to be "God's party (Hizb Allah)," who are congratulated for "achieving God's gardens [of Paradise]." Al-Muwali's *nashid* "O' Zaynab" has become a staple song used in many pro-Shi`a militia videos posted to video-sharing websites such as YouTube. Al-Karbala'i is identified as a "fighter [of Imam Husayn]" and he has recorded a martial anthem, based on a praise poem (*qasida*) by poet Muzaffar al-Jabari, dedicated to Liwa Abu al-Fadl al-Abbas.

Appendix HEFFELFINGER02

The Emergence and Composition of the Shi`a Militias

Media reports of Shi`a foreign fighters in Syria first surfaced in the autumn of 2012 and were largely based on interviews with militiamen participating in the fighting as well as Iraqi government officials.[15] Videos supporting Shi`a fighters began to appear on websites such as YouTube around the same time.[16] These included videos dedicated to those slain fighting "for Sayyida Zaynab," the daughter of Ali ibn Abi Talib, the first Shi`a imam and cousin and son-in-law of the Prophet Muhammad, as well as footage of funeral prayers for them inside the Sayyida Zaynab shrine.[17]

The fighters are primarily affiliated with Liwa Abu al-Fadl al-Abbas (hereafter the al-Abbas Brigade), a militia composed of fighters from a variety of nationalities including Iraqis, Syrians, Lebanese, Iraqi refugees in Syria, other Arabs, and Afghans.[18] Exact numbers of fighters are impossible to verify, but in an interview in late October 2012 an unnamed militiaman said that about 200 Iraqi recruits had traveled to Syria.[19] Most were drawn from groups that splintered from the mainstream Sadr Movement (*Tayyar al-Sadr*) led by Moqtada al-Sadr, as well as Iranian-supported militias originally formed to fight U.S. and coalition military forces in Iraq.[20] The latter include Asaib Ahl al-Haq (League of the Righteous/People of Truth) and Kataib Hizb Allah (Party of God Brigades).[21] The involvement of these groups can be verified through martyrdom statements, reports of funerals of Iraqi fighters slain in Syria, and reports on pro-militia websites.[22] Photographs of martyrs are often emblazoned with the emblems of the groups to which they belonged.[23] The exact breakdown of the affiliation of al-Abbas Brigade fighters is unclear, although a video commemorating some of its slain fighters purported to show each individual's group affiliation.[24]

"Recruitment of fighters has escalated in Iraq, with different Shi`a groups, including the Sadr Movement, forming committees to seek volunteers."

An unnamed official in the Badr Organization, a powerful Iraqi Shi`a political movement originally trained and supplied by Iran in the 1980s to fight Saddam Hussein, claimed that Shi`a militias were acquiring "new and advanced" heavy weapons for use in case the conflict escalates.[25] In an early October 2012 interview, Abu Hajer, an al-Abbas Brigade commander who had been living in Syria as an Iraqi refugee, said that the al-Abbas Brigade comprised 500 fighters.[26] According to interviews with unnamed Iraqi Shi`a leaders, the Syrian and Iranian governments have provided the militias with weapons, and the latter has organized the travel of Iraqi recruits.[27] Training camps for recruits are reportedly located in Iran, and fighters are aided by local contacts inside Syria.[28]

In a March 2013 interview with Russia Today, a state-funded satellite television network, Abu `Ajib, who is identified in media reports and militia media materials as the leader or secretary general of the al-Abbas Brigade, said that the militia was formed "months before" to defend Sayyida Zaynab's shrine from rebel attacks.[29] Brigade commander Abu Hajer has said that the brigade has carried out joint military operations against rebel groups when fighting erupted in the Sayyida Zaynab district near the shrine.[30] Although its leaders have said that they are not concerned with other internal fighting except with regards to the Sayyida Zaynab shrine and other Shi`a shrines, a brigade fighter, Abu Mujahid, said that the militia also carried out joint attacks with the Syrian military against bases of militias belonging to the Free Syrian Army (FSA) umbrella group.[31]

Photographs and artwork released on pro-brigade websites feature fighters in front of posters and portraits of Bashar and Hafiz al-Assad and other pro-al-Assad pieces of artwork, including a comparison of Hizb Allah Secretary General Hassan Nasrallah and Bashar al-Assad, who are both shown reading the Qur'an with divine

15 Al-Salhy, "Iraqi Shi'ite Militants Fight for Syria's Assad."

16 This is based on the author's collection and cataloguing of videos and other media materials produced by or for these militias. It is possible, however, that earlier media uploads on websites such as YouTube were taken down by the uploader or the website administrators.

17 These videos, which were uploaded online by unknown individuals whose affiliation with the brigade seems likely but is unconfirmed, include: *The Prayer over the Martyrs in the Shrine of Sayyida Zaynab, Upon Her Be Peace*, September 2012; *Presentation to the Souls of the Martyrs of Sayyida Zaynab, Upon Her be Peace, and to the Souls of the Martyrs of al-Asad's Syria*, September 2012; and *Presentation to the Souls of the Martyrs of Sayyida Zaynab*, September 2012.

18 This is based on the militia's media materials, pro-militia media materials, and media reports, including: "Syria..National Defense Brigades' Specific Tasks," Russia Today, March 26, 2013; Figures 12 and 34 in "Visual References," Views from the Occident blog, June 23, 2013, available at http://occidentblog.wordpress.com/2013/06/24/visual-references/. Afghan Shi`a seminary students expelled from Iraq in the 1970s settled in the Sayyida Zaynab district where many of them restarted their studies at the Zaynabiyya Hawza. See Laurence Louër, *Transnational Shia Politics: Religious Networks in the Gulf* (New York: Columbia University Press, 2008).

19 "After Threats to Shrine, Iraqi Shiite Fighters Prepare for Sectarian Strife at Home, in Syria."

20 Ibid. This is also based on martyrdom statements released by Iraqi Shi`a militias.

21 Ibid.

22 Representative statements include: "Statement: The Martyr, Ahmad Mahdi al-Shuwayli," Kataib Hizb Allah, April 15, 2013; "Statement: The Martyr Riza Khudayr al-Khalidi," Kataib Hizb Allah, May 6, 2013.

23 See Figures 13-15 in "Visual References."

24 *Martyrs of Liwa' Abu al-Fadl al-`Abbas*, June 2013. In addition to the Sadr Movement, Asaib Ahl al-Haq, and Kataib Hizb Allah, the video also purports to show martyred fighters from two other militias, the God's Soldiers Brigade and Brigade of the Force of Haydar, an honorific title meaning "lion" in Arabic used by Shi`a for Imam `Ali.

25 Ibid.

26 Al-Salhy, "Iraqi Shi'ite Militants Fight for Syria's Assad."

27 Yasir Ghazi and Tim Arango, "Iraqi Sects Join Battle in Syria on Both Sides," *New York Times*, October 27, 2012.

28 Mona Mahmood and Martin Chulov, "Syrian War Widens Sunni-Shia Schism as Foreign Jihadis Join Fight for Shrines," *Guardian*, June 4, 2013.

29 "Syria: National Defense Brigades' Specific Tasks," Russia Today, March 26, 2013.

30 Al-Salhy, "Iraqi Shi'ite Militants Fight for Syria's Assad."

31 Ibid.

Appendix HEFFELFINGER02

light emanating from the pages onto their faces.[32]

The al-Abbas Brigade is composed of several smaller fighting units named after figures from Islamic history who are particularly revered by Shi`a, such as the Twelve Imams recognized by Shi`a as the world's legitimate religious and temporal authorities.[33] These units include the Ali Akbar Brigade, Brigade of the Awaited One (named after an honorific title of the twelfth imam, Muhammad bin Hassan), al-Qasim Brigade, and the Brigade of Malik al-Ashtar.[34] Units likely affiliated with the al-Abbas Brigade include the Brigade of Zaynab's Protector and the Zulfiqar Brigade, which share members based on photographs released online on pro-brigade websites, primarily on Facebook.[35] At least one unit is named after an al-Abbas Brigade martyr, Ahmad Kayara, of which some photographs have been released on pro-brigade websites. Members of Iran's Islamic Revolutionary Guard Corps and other Iraqi Shi`a groups—such as the Brigades of the Prince of Martyrs (Kataib Sayyid al-Shuhada),the Imam Husayn Brigade, the God's Soldiers Brigade, and the Brigade of Ammar bin Yasir—are also reportedly active militarily inside Syria.[36] Full details, however, regarding the internal organization of the al-Abbas Brigade, its individual fighting units, and other Shi`a militias are unclear.

"By portraying all Syrian rebels as Salafists, the Shi`a militias and their supporters are essentially arguing that dialogue is hopeless."

Recruitment of fighters has escalated in Iraq, with different Shi`a groups, including the Sadr Movement, forming committees to seek volunteers.[37] Convoys of buses said to be carrying pilgrims have instead been filled with fighters and military supplies bound for the front in Syria.[38] In October 2012, a recruitment committee in Iraq's Diyala Province, the site of much sectarian strife between Iraqi Sunnis and Shi`a and a stronghold of the Islamic State of Iraq, a jihadist/insurgent umbrella group dominated by al-Qa`ida in Iraq (AQI), claimed that it had sent 70 fighters to Syria.[39]

Although the majority of fighters seems to be drawn from Shi`a communities, al-Abbas Brigade-affiliated sources, including leader Abu `Ajib, have claimed that the group's fighters comprise members of multiple groups including Sunni and Shi`a Muslims and Druze.[40] Determining the truth of these claims is not possible. The mobilization frames and historical and cultural repertoire from which the al-Abbas Brigade draws and the way in which they are deployed is distinctly Shi`a, which brings into question whether it can successfully recruit outside its Shi`a base.

Unity within the brigade has also reportedly come under increasing strain.[41] Emerging internal divisions show that while communal and sectarian identities can be used to initiate social and militant mobilization, they are not sufficient enough by themselves to ensure long-term unity.

Maintaining social movement cohesion and group unity in the longer term is difficult. Sectarian, ethnic, and other group identities are often not enough by themselves to maintain unity, and political and economic interests often come into play. Divergent political and economic interests can then trump shared sectarian and group identity. Localism is also often a factor in militant recruitment and the formulation of strategic goals and ideological positions.

Mobilizing Historical Memory and Popular Piety

The mobilization frames used by Shi`a actors in Syria play an integral role in their media operations, which are aimed at both attracting recruits as well as more non-military and more passive types of support from Shi`a communities globally. This includes the creation of pro-militia artwork, video montages, and other media as well as legitimizing, in the public sphere, the militias' involvement in Syria.

In formulating its mobilization frames, the al-Abbas Brigade and other Shi`a militias draw upon the deep reservoir of historical and cultural memory of Shi`a Islam. The heroic figure of Zaynab, the sister of Imam Husayn, and the other historical persons referenced by the brigade and its supporters evoke the tragedy of Karbala in 680 CE when Imam Husayn and many of his closest supporters and male family members were slain in battle against a larger force sent by the Umayyad Caliph Yazid I.[42] Brigade and pro-brigade media output contains numerous Shi`a historical and cultural symbols, including paintings of Abu al-Fadl al-Abbas[43]

32 See Figures 16-18 in "Visual References."

33 "News Report: Formation of a Shi'ite Brigade Composed of Iraqis and Lebanese for Defending the Shrine of Sayyida Zaynab South of Damascus," Russia Today, March 4, 2013.

34 See Figures 2-4 in "Visual References." Also see *Clashes of the Al-Qasim Brigade of Liwa' Abu al-Fadl al-`Abbas*, April 2013.

35 See Figure 5 and 6 in "Visual References." It is also possible that members of the latter two groups broke away from the al-Abbas Brigade after recent infighting.

36 "Formation [Unit] Calling Itself the 'Prince of Martyrs Brigades' Rejects Sectarianism and Calls for the Unity of the Iraqi People," al-Baghdadiyya News Agency, May 8, 2013; "Defection of the Secretary General of the Hizb Allah Brigades [in] Iraq and the Formation of the 'Prince of Martyrs [Brigade],'" *al-Masalah News*, April 14, 2013; Marcus George, "Iran's Revolutionary Guards Commander Says its Troops in Syria," Reuters, September 16, 2012; "Kazimiyya: The Funeral of a Retired Army Officer who was Killed One Month after His Participation [Began] Fighting in Syria," *al-Mustaqbal News*, July 10, 2013; *Clash of the Heroes of the `Ammar bin Yasir Brigade in the Countryside of Aleppo*, July 19, 2013; Figure 36 in "Visual References." Ammar bin Yasir was a key supporter and military commander of the first Shi`a imam, Ali ibn Abi Talib, who was martyred fighting for him at the Battle of Siffin in 657 CE.

37 Ghazi and Arango. Videos of trucks reportedly carrying Iraqi foreign fighter volunteers have also been posted to the internet.

38 Ibid.

39 Ibid.

40 "News Report: Formation of a Shi'ite Brigade Composed of Iraqis and Lebanese for Defending the Shrine of Sayyida Zaynab South of Damascus." Also see Figure 1 in "Visual References."

41 Al-Salhy, "Iraqi Shi'ites Flock to Assad's Side as Sectarian Split Widens."

42 Their martyrdoms are commemorated annually during the first 10 days of the Islamic lunar month of Muharram, which culminates on the 10th day, called the Day of Ashura, when Imam Husayn himself was slain.

43 "Abu al-Fadl" is an honorific name meaning "father

Appendix HEFFELFINGER02

and Zaynab, Ashura artwork, and flags and standards (*`alam*) representing that of al-Abbas.[44] Popular religious songs based on poetry and videos produced by brigade supporters and possibly the brigade itself feature frequent mentions of Zaynab and other Shi`a holy figures such as al-Abbas.

Shi`a consider Zaynab to have been the individual most responsible for keeping Imam Husayn's message, and thus the purest form of Islam, alive after his death.[45] The pledge of loyalty, "*labbayk ya Zaynab*," ("we are at your service, O' Zaynab") and vow from fighters that "we sacrifice our souls for you, O' Zaynab" are frequent features in pro-brigade artwork and videos.[46] In a bid to counter anti-Assad activists, pro-brigade activists have even modified a slogan used previously by an anti-Assad Facebook campaign, "From [location], here is Damascus," followed by a declaration of loyalty to Zaynab.[47]

In the media operations of the al-Abbas Brigade and its supporters, Shi`a militiamen fighting in Syria are portrayed as "Zaynab's guardians" (*hurras Zaynab*) who are defending her shrine and, in turn, her honor, through self-sacrifice (*fida'*) from Salafist/Wahhabi hordes, which is how Syrian rebels are described in the discourse of the militias and their supporters; they are equated with historical villains such as the Umayyads.[48] Salafist creedal beliefs are virulently anti-Shi`a, and thus there is a long history of polemics between the two groups. By portraying all Syrian rebels as Salafists, the Shi`a militias and their supporters are essentially arguing that dialogue is hopeless. Brigade and pro-brigade media materials frequently include photographs of slain members and commanders from various rebel factions including the AQI-affiliated Jabhat al-Nusra.[49]

The al-Abbas Brigade takes its name from Abu al-Fadl al-Abbas, who was Imam Husayn's half-brother and standard bearer at Karbala.[50] Al-Abbas is the quintessential selfless hero of the Karbala story, and his sense of duty is emphasized in Shi`a accounts of his role as the defender of Husayn's family and his refusal to allow Husayn's standard to fall to the ground until al-Abbas could no longer stand.[51] A popular slogan in al-Abbas Brigade artwork is "*kuluna `Abbasak, ya Zaynab*" ("we are all your Abbas, O' Zaynab").[52]

The names of units within or affiliated with the brigade are taken from other heroes and symbols prominent in Shi`a Islam. These include units named after Imam `Ali's famous two-pronged sword, Zulfiqar, and the Brigade of Zaynab's Protector (Kafil Zaynab), the latter of which again emphasizes the defense of Zaynab's honor by defending her shrine.[53] The 12th Shi`a imam, Muhammad bin Hassan al-Mahdi, who is a messianic figure in Shi`a eschatology, also has a militia unit named after him, the Brigade of the Awaited Savior (Katibat al-Mahdi al-Muntazar).[54]

Other Shi`a heroes after whom Shi`a units fighting in Syria are named include Ali Akbar, the teenage son of Imam Husayn, and al-Qasim, his nephew and son of his predecessor, Imam Hassan, both of whom were martyred at Karbala.[55] Malik al-Ashtar and Ammar bin Yasir, two of the supporters of the Prophet Muhammad who became ardent followers of the first Shi`a imam, `Ali, also have fighting units named after them.[56]

For their part, pro-Assad, Sunni rebel groups have drawn upon historical narratives and polemics to counter the

of al-Fadl," the name of one of his sons.

44 See Figures 24-27 in "Visual References."

45 She is given the honorific *sayyida*, which in this context translates to "Lady."

46 See, for example, this video uploaded to YouTube and other video sharing websites: *Presentation O' Zaynab: Operations of Liwa' Abu al-Fadl al-`Abbas, Designated to Protect the Shrine of Sayyida Zaynab, Upon her be Peace*, December 2012. Another example is `Ali Abu Kiyan al-Muwali's song "O' Zaynab." One pro-brigade video montage included a part of a scene from a popular Arabic-language film about Karbala, *The Caravan of Pride*.

47 See Figures 20-23 in "Visual References." For background of the slogan's origins and use in a major Facebook campaign by opposition supporters, see "Here is Damascus," The Syrian, May 29, 2013, available at www.english.the-syrian.com/2013/05/29/here-is-damascus.

48 See Figures 7, 8, 30, and 31 in "Visual References." Also see *Letter from Hezbollah to the Wahhabis*, June 2013, available at www.youtube.com/watch?v=vstAoKzA7nA. One internet poster from a pro-brigade website compared Syrian rebels, by association with rebel commander-turned-symbolic cannibal Abu Sakkar, with Hind bint `Utba, one of the Prophet Muhammad's most virulent enemies before his conquest of Mecca in 630 CE. For Abu Sakkar's reported cannibalism, see Drishya Nair, "Syria's Cannibal Commander Abu Sakkar: Why I Ate My Enemy's Heart," *International Business Times*, May 15, 2013. For the poster, see Figure 35 in "Visual References."

49 See Figures 28-30 in "Visual References."

50 An alternative name for the brigade is Katibat Liwa Abu al-Fadl al-Abbas (Brigade of the Standard of Abu al-Fadl al-Abbas), which emphasizes his role as Imam Husayn's standard bearer at Karbala. See Figure 19 in "Visual References."

51 Sayyid Najmulhasan Karrarvi, *Biography of Hazrat Abbas* (Karachi: Peermahomed Ebrahim Trust, 1974), particularly pp. 49-51 and 78-90. Also see the film *Karbala: When the Skies Wept Blood*.

52 See Figure 9 in "Visual References."

53 "'Zulfiqar Brigade,' Second Militia for the Defense of the Shrine of Sayyida Zaynab!" Orient News, June 9, 2013; *'Zulfiqar Brigade'...Formation of a New Shi'i [Militia]*, June 9, 2013, available at www.youtube.com/watch?v=WH7O4hm-P8A. As with many Islamic historical figures and symbols, Zulfiqar is a contested symbol and some anti-Assad Syrian rebel groups have also taken it as part of their name. See *Statement on the Formation of the Zulfiqar Brigade of the Brigade of the Martyrs of Rusafa following the Military Revolutionary Council [in] Deir al-Zur*, Zulfiqar Brigade, June 2013; *With God's Help, the Formation of the Zulfiqar Brigade in the Free City of Jarablus, following the `Amr bin al-`As Brigade*, Zulfiqar Brigade, September 2012; *Zulfiqar Brigade of Colonel Muhammad Ziyad Qasimu*, Zulfiqar Brigade, April 2013; *Shield of Zulfiqar Brigade Responding to the Hordes of al-Asad in the Towns of Karnaz and the Countryside of Hama*, Shield of Zulfiqar Brigade, February 2013.

54 See Figure 3 in "Visual References." Shi`a Muslims believe that the 12th imam, Muhammad bin Hassan, is the last imam in a line of divinely-appointed successors to the Prophet Muhammad who possess both religious and temporal authority over the Muslim community.

55 *Clashes of the Qasim Brigade of Liwa' Abu al-Fadl al-`Abbas*, April 2013, available at www.youtube.com/watch?v=SZoSKvyW4KI. See Figure 2 in "Visual References." Al-Qasim is known among Shi`a as the "Bridegroom of Karbala" because it is said he was married on the morning of the day he was martyred at Karbala. This account first appears, it seems, in the 16th century, Persian-language elegiac martyrology of Husayn Va`iz Kashifi, *Garden of the Martyrs*. See Syed Akbar Hyder, *Reliving Karbala: Martyrdom in South Asian Memory* (Oxford: Oxford University Press, 2006); David Pinault, *Horse of Karbala: Muslim Devotional Life in India* (New York: Palgrave, 2001). It has been published numerous times in various languages. See Husayn Va`iz Kashifi, *Garden of the Martyrs* (Qum: Daftar-i Nashr-i Navid-i Islam, 2000 or 2001).

56 See Figure 4 in "Visual References." Malik was appointed governor of Egypt. Malik was poisoned while traveling to Egypt on the orders of Ali's chief rival for leadership of the Muslim state, Mu`awiya ibn Sufyan, who is reviled by Shi`a but considered one of the Prophet Muhammad's companions by Sunnis.

Appendix HEFFELFINGER02

information operations and messaging of the al-Abbas Brigade, Hizb Allah, and other Shi`a actors in Syria. Rebel discourses refer to them and sometimes Shi`a generally as being guilty of polytheism (*shirk*) and unbelief (*kufr*) because of their belief in the line of the Twelve Imams. They also refer to Shi`a as the "fire worshippers" or Zoroastrians since they allege that Shi`a Islam is nothing more than an offshoot of pre-Islamic Iranian religions. Rebel videos showing military operations against the brigade often feature captured Shi`a cultural artifacts such as paintings of the imams and Abu al-Fadl al-Abbas, Shi`a religious writings and multimedia, and tablets of clay used by Shi`a during prayer.[57]

Conclusion

Information on Shi`a militias operating in Syria is limited, despite the relative wealth of information on pro- and quasi-militia websites. In particular, information on the individual motivations of recruits remains unknown due to the paucity of available primary sources.[58] Even when such sources are available, the declared intentions of fighters need to be evaluated carefully since these declarations are not necessarily truthful or accurate.[59] Martyr biographies are also usually highly and posthumously hagiographical in their depiction of the deceased and their motivations.[60]

The rise of Shi`a militias and their political backers in Iran and Iraq also present a potential challenge to the authority of Shi`a *maraji` al-taqlid* (Shi`a religious authorities), such as Grand Ayatollah Ali al-Sistani in Iraq. Although the most recognized senior Shi`a jurists, such as al-Sistani, have not yet addressed the appearance of these militias or related questions regarding the religious legitimacy of their claims to be engaged in military jihad, it is important not to exaggerate these jurists' ability to guide the public's behavior, even that of individuals who ostensibly follow their religious opinions and rulings. In 2006, for example, following the first bombing of the al-Askari Mosque in the Iraqi city of Samarra, where two of the Twelve Imams are buried, Grand Ayatollah al-Sistani issued several statements calling for patience and forbidding acts of revenge and retaliatory violence, explicitly mentioning sectarian attacks against Sunnis.[61] He and Najaf's three other resident grand ayatollahs were ultimately trumped by powerful Iraqi Shi`a political actors who, together with Iraqi and foreign Sunni militants, led Iraq down the path of sectarian conflict.

Although the influence of Shi`a religious leaders is significant, it is not absolute, and recent history has shown that in the midst of sectarian violence they are often overshadowed by virulent sectarian political voices.

Christopher Anzalone is a Ph.D. student in the Institute of Islamic Studies at McGill University where he studies modern Muslim socio-political movements, contemporary jihadi movements, Shi`a Islam, and Islamist visual cultures. He is also an adjunct research fellow at the Ali Vural Ak Center for Global Islamic Studies at George Mason University and managing editor for the center's new online Islamic Studies platform, Islamium, currently under development.

57 *The Free [Syrian] Army Raids One of the Headquarters of Liwa' Abu al-Fadl al-`Abbas in the Damascus Countryside*, April 2013.

58 Christopher Anzalone, "A Few Notes on Shi'ism in Syria and the Emergence of a Pro-Asad Shi'i Militia, Liwa' Abu'l Fadl al-`Abbas," al-Wasat blog, May 22, 2013.

59 Thomas Hegghammer, "Should I Stay or Should I Go? Explaining Variation in Western Jihadists' Choice between Domestic and Foreign Fighting," *American Political Science Review* 107:1 (2013).

60 This is based on the author's analysis of contemporary written and audiovisual Sunni and Shi`a martyrologies.

61 `Ali al-Sistani, statements, February 2006 and February 2007.

Al-Qa`ida in the Islamic Maghreb's Tunisia Strategy

By Aaron Y. Zelin, Daveed Gartenstein-Ross & Andrew Lebovich

DURING THE LAST two years, Salafist activity has escalated in Tunisia. Much of this activity—primarily *da`wa* (religious outreach) designed to expand the Salafist movement—has taken place under the auspices of Ansar al-Shari`a in Tunisia (AST), headed by veteran jihadist Saifullah bin Hassine (also known as Abu Iyadh al-Tunisi).[1] A series of security incidents in and around Tunisia, however, have been attributed to al-Qa`ida in the Islamic Maghreb (AQIM), and more recently to an opaque group known as the Uqba ibn Nafi Brigade.[2] Regional security officials have described the Uqba ibn Nafi Brigade as an attempt to establish a Tunisian jihadist group linked to AQIM, one that purportedly combines local recruits from western Tunisia's Kasserine area and some members of AST under the guidance and leadership

1 Abu Iyadh was part of the United Kingdom's "Londonistan" scene in the 1990s, where he became associated with jihadist figures such as Hani al-Siba`i and Abu Qatada al-Filistini, even being described as Abu Qatada's "disciple" on some jihadist forums. See "Tunisian Salafi Ansar al-Shari'ah Gaining Hard-line Jihadist Support," BBC Monitoring, May 24, 2012; Fabio Merone, "Salafism in Tunisia: An Interview with a Member of Ansar al-Shari`a," *Jadaliyya*, April 11, 2013. He also became associated with Algeria's Salafist Group for Preaching and Combat (GSPC) during this period and facilitated Tunisians' travel to Algeria so they could receive militant training with the GSPC. In Afghanistan, he co-founded the Tunisian Combatant Group, which the United Nations designated an al-Qa`ida-affiliated terrorist organization in 2002. For details, see "QE.T.90.02. Tunisian Combatant Group," UN Security Council, April 7, 2011; Aaron Y. Zelin, "Missionary at Home, Jihadist Abroad: A Profile of Tunisia's Abu Ayyad the Amir of Ansar al-Shari`a," *Militant Leadership Monitor* 3:4 (2012); Lorenzo Vidino, *Al Qaeda in Europe: The New Battleground of International Jihad* (Amherst, NY: Prometheus Books, 2005).

2 Uqba ibn Nafi was a seventh century Arab general under the Umayyad Caliphate who founded the Tunisian city of Qayrawan. Interestingly, the Uqba ibn Nafi mosque in Qayrawan, a city renowned for its longstanding tradition of Islamic scholarship, has become a bastion of Salafist preaching since 2011. See "Tunisie – Mohamed Hassen à la Mosquée Okba Ibn Nafaâ: Une foule, survoltée, réclame la Chariâa (vidéo)," *Business News* [Tunis], May 3, 2013.

Appendix HEFFELFINGER02

of figures reputedly close to AQIM leader Abdelmalek Droukdel (also known as Abu Mus`ab al-Wadud).[3] Regional security officials, therefore, perceive the incidents on Tunisia's border with Algeria beginning in late April 2013 as highlighting AQIM's increased focus on Tunisia.

This article analyzes how AQIM, viewing events in Tunisia through its strategic lens, has responded to that country's revolution. It finds that AQIM has tried to insert itself into AST's relationship with the Tunisian state. AQIM has urged AST to be patient to prevent the Tunisian government from cracking down on its activities. At the same time, AQIM's rhetoric toward the Tunisian state has become sharper, opening the possibility of an increase in AQIM-related violence in Tunisia.

AQIM's Strategic Outlook

Understanding AQIM's policies toward Tunisia requires an awareness of the group's perceptions of the changes brought by the Arab uprisings, and how these changes in turn influence its strategy toward the region. Salafi-jihadi thinkers and strategists who are influential to AQIM quickly reached consensus about what the Arab uprisings meant, formulating their ideas about the revolutionary events even while Hosni Mubarak's regime in Egypt drew its final breaths. They agreed that the changes gripping the region were positive, primarily because they created unprecedented opportunities to undertake *da`wa*.[4] At the same time, the movement would not be satisfied constraining itself to *da`wa*, and this early theoretical work emphasized a phased approach wherein Salafi-jihadis—even while undertaking *da`wa* peacefully, in ways they could not under the old regimes—would prepare to later engage in violence.[5]

Tunisia-based researcher Monica Marks identified three major divisions of Salafism within Tunisia: *Salafiyya `Almiyya* (usually translated as scientific Salafism, but which Marks believes is better understood as scripturalist Salafism), political Salafism, and *Salafiyya Jihadiyya* (Salafi-jihadism).[6] To those who can be categorized as *Salafiyya `Almiyya*, democracy is "a tempting, but ultimately dead-end street," and instead of engaging in party politics they choose "apolitical lives of quietist piety."[7] Political Salafists have much in common with *Salafiyya `Almiyya*, but believe participation in democratic politics is justified despite its flaws because it "could serve as a vehicle to attain a more caliphate-like, Shari`a-based polity."[8] Those who can be considered *Salafiyya Jihadiyya* reject both democratic participation and also the quietism of *Salafiyya `Almiyya*.

Although AST has distanced itself from violence at this time, it has shown a devotion to Salafi-jihadi ideology, a fact that can be seen from both their own explanations and threats,[9] as well as the significant and long-standing connections that leaders and key members of the organization have to transnational jihadists.[10] Indeed, because of these factors, AST has won the praise of key leaders and intellectuals in the Salafi-jihadi movement.[11]

Its current activities within Tunisia are mostly concentrated on *da`wa*, which has allowed the movement to grow in size and give it influence beyond its relatively small numbers.[12]

The bulk of the jihadist activities in which the group might be involved have been focused abroad, such as promoting the flow of foreign fighters to Syria, although the group has also been accused of orchestrating attacks on perceived cultural opponents within Tunisia.[13] Although the extent of AST's

3 See, for example, Benjamin Roger, "Tunisie: sur la trace des djihadistes de mont Chaambi," *Jeune Afrique*, May 7, 2013; Mourad Sellami, "Les précisions du ministre de l'Intérieur tunisien," *El Watan* [Algiers], December 22, 2012; "Tunisie-Alqaida-Aqmi: Groupes armés, Larayedh sonne l'alert," Tunis Afrique Presse, December 21, 2012; Andrew Lebovich, "Confronting Tunisia's Jihadists," *Foreign Policy*, May 16, 2013.

4 See Ayman al-Zawahiri, "And be Neither Weakened nor Saddened," al-Sahab, August 15, 2011; Atiyatallah `Abd al-Rahman, "The People's Revolt...The Fall of Corrupt Arab Regimes...The Demolition of the Idol of Stability...and the New Beginning," distributed by the Global Islamic Media Front, February 16, 2011. For an in-depth discussion of early Salafi-jihadi perceptions of the Arab uprisings, see Daveed Gartenstein-Ross and Tara Vassefi, "Perceptions of the 'Arab Spring' Within the Salafi-Jihadi Movement," *Studies in Conflict & Terrorism* 35:12 (2012).

5 Prominent Mauritanian ideologue Abu al-Mundhir al-Shinqiti, who is well known for his belief in the necessity of violence, emphasized a phased approach by speaking to the need for leniency while *da`wa* is predominant. Referring specifically to Tunisia, he said that "present circumstances...require that the people of monotheism concentrate on the aspect of preaching," due to ignorance of Islam caused by Zine El Abidine Ben Ali's fallen regime. See Abu al-Mundhir al-Shinqiti, "Answers to Questions from Our Tunisian Brothers," Minbar al-Tawhid wal-Jihad, December 15, 2011. AST leaders have also articulated a phased approach. Hassan Ben Brik, who heads AST's *da`wa* committee, has described jihad as "certainly part of our political project," but said that AST has "no interest currently in embarking on violent initiatives, or acts of terrorism." See Sergio Galasso, "Intervista ad Hassan Ben Brik: 'Non crediamo nella democrazia, ma senza appoggio del popolo niente jihad,'" *Limes*, October 11, 2012. Another young AST leader told an Italian researcher that AST has not "eliminated the idea of jihad from our philosophy," but that the group is not currently engaged in revolutionary violence because it is focused on *da`wa*. See Merone, "Salafism in Tunisia."

6 Monica Marks, "Youth Politics and Tunisian Salafism: Understanding the Jihadi Current," *Mediterranean Politics* 18:1 (2013). This tripartite division is consistent with the strains of Tunisian Salafism identified by other researchers. See Stefano M. Torelli, Fabio Merone & Francesco Cavatorta, "Salafism in Tunisia: Challenges and Opportunities for Democratization," *Middle East Policy* 19:4 (2012).

7 Marks, "Youth Politics and Tunisian Salafism: Understanding the Jihadi Current."

8 Ibid.

9 "Abu Iyadh Threatens to 'Afghanise' Tunisia in Most Serious Salafist Menace," Middle East Online, May 13, 2013; Aaron Zelin, "Who's Who in Tunisia's Salafi-Jihadi Community: A Look at Key Leaders of Ansar al-Shari`a in Tunisia," *Militant Leadership Monitor*, April 2013.

10 Abu Iyadh has said that AST "shared in the same *manhaj*," or religious methodology, as al-Qa`ida. See Zelin, "Missionary at Home, Jihadist Abroad." For AST leaders' explanations of their devotion to jihadist violence, see Galasso, "Intervista ad Hassan Ben Brik"; Merone, "Salafism in Tunisia."

11 For praise of AST by key Salafi-jihadi figures, see "Journalistic Encounter with the Director of Al-Andalus Media Foundation," al-Qa`ida in the Islamic Maghreb, April 18, 2013; "Web: Jihadists Send Mixed Messages to Tunisia's Salafis," BBC Monitoring, November 8, 2012.

12 Daveed Gartenstein-Ross, *Ansar al-Shari`a Tunisia's Long Game: Dawa, Hisba and Jihad* (The Hague: International Centre for Counter-Terrorism-The Hague, 2013).

13 Thomas Joscelyn, "Tunisian 'Martyrs' Celebrated by Ansar al Sharia," *The Long War Journal*, March 13, 2013; Ahmed Maher, "Tunisia's Radical Divide over Salafi Agenda," BBC, June 6, 2013; Thomas A. Bass, "How Tunisia is Turning Into a Salafist Battleground," *The Atlantic*, June 20, 2013. Aaron Y. Zelin has kept a comprehensive database of martyrdom announcements issued by rebel forces in Syria that demonstrates both the presence of AST fighters and also the fact that AST subsequently

Appendix HEFFELFINGER02

relationship with AQIM is unknown, AQIM leaders have repeatedly offered advice to the group on how to proceed in Tunisia.

Advising Tunisians, Ennahda, and AST

In its public statements, AQIM has attempted to act as a wise outsider providing counsel on the best course for Tunisia's future—offering advice to Ennahda and AST. Ennahda, Tunisia's Islamist political party, contested democratic elections after the fall of the Zine El Abidine Ben Ali government and won a majority of seats. Yet Ennahda's more moderate policies have placed it at odds with more hard line groups such as AST and other Salafists.

AQIM has a powerful message in the form of its warnings that Tunisians' hard-won victory over the former regime could be stolen from them. This message especially resonates with Tunisians who are already sympathetic to AQIM's worldview.

The first statement AQIM released on the Tunisian revolution came on January 13, 2011, when the group's *amir*, Abdelmalek Droukdel, justified the uprising by painting the regime as "exceeding its boundaries in fighting Islam like no other."[14] He told Tunisians that their movement must be broad and led by the wisdom of the *ulama*.[15] At the same time, he offered Tunisians the opportunity to train with AQIM for the final battle against the Jews and Christians, saying that the overthrow of the local regime was but one stage in a broader war.[16] He framed Tunisians' struggle as part of AQIM's larger fight: "Your mujahidin brothers are with you, your problem is our problem, and your happening is ours."[17]

A couple of weeks later, AQIM released another statement highlighting the opportunities and dangers of the post-Ben Ali era.[18] AQIM advised Tunisians "to take advantage of this historical opportunity to spread the *manhaj* (methodology) of *tawhid* (monotheism) and jihad."[19] At the same time, Tunisians must "prepare and be ready as the days are pregnant and the Crusade war against Islam and Muslims remains utmost intense."[20] This statement is consistent with the theoretical work produced by Salafi-jihadi thinkers counseling an initial stage of preaching that would culminate in violence when the movement is ready (or when its hand is forced).

AQIM did not release any new statements on Tunisia until October 2012. The return of its public rhetoric coincided with AST's ascendance, as well as Tunisia's leading Islamist party, Ennahda, distancing itself from Salafist movements following the September 2012 attack on the U.S. Embassy in Tunis.[21] The rhetorical and later street confrontation between Tunisian security services and AST gave AQIM the opportunity to attempt to insert itself as an arbiter among different Islamist movements. Later, AQIM offered support and advice to AST, as a look at a series of AQIM's subsequent statements demonstrates.

In October 2012, AQIM took a soft tone in a public address, noting that it was only giving advice so as not to "fall in the traps of the enemies of Islam."[22] The message to Ennahda was that while there might be differences in tactics, both groups want the same end result: implementation of Shari`a.[23] Therefore, AQIM said that it is better to work together than to engage in intra-Islamist infighting, which would only help those with a secular or liberal agenda.[24] AQIM warned against working with the United States, and also France and the old regime—an allusion to the upstart political party Nida' Tunis (the Tunisian Call), which has a number of leaders from the former regime.[25] AQIM stated that the "Islamic movement" should focus on spreading its message through *da`wa* while avoiding *takfir* (excommunication).[26]

In a March 2013 statement, AQIM further emphasized the importance of *da`wa* by exhorting Tunisians to stay in their own country rather than joining the jihad abroad.[27] Without a critical mass of Salafi-jihadis at home, AQIM warned, Tunisia would be ceded to secularists who "spread corruption."[28]

Yet the tone of AQIM's statements shifted less than two months later, as a confrontation heated up with Ennahda and the Tunisian state on one side, and AQIM and AST on the other.[29] At the time, Tunisia was undertaking military operations against the Uqba ibn Nafi Brigade on its border with Algeria, and there were also low-level clashes in Tunisian cities due to the state's cancellation of AST's annual conference in May 2013. AQIM was evolving to a more hostile stance toward Ennahda—one that mirrors the increasing anger of Salafists and Salafi-jihadis toward Ennahda, and the worsening AST-Ennahda relationship.

For the first time, AQIM explicitly criticized the Islamist Ennahda party for its perceived wavering on the issue of Shari`a, and for selling democracy as legitimate under Islam. "Don't be deceived by the people who call to democracy by decorating it under the name of the truth and give it a religious garment," the statement implored.[30] Therefore, AQIM advised Tunisians to shield themselves from these provocations by joining "your sincere preachers of Ansar al-Shari`a, who stood next to you."[31] Yet while these statements were hostile toward Ennahda, Mauritanian AQIM member Abu Yahya al-Shinqiti also counseled AST to be

issued laudatory announcements when they were killed.

14 Abu Mus`ab al-Wadud, "In Support of the Intifadah of our People in Tunisia," al-Andalus Media, January 13, 2011.

15 Ibid.

16 Ibid.

17 Ibid.

18 "To Our People in Tunisia: The Tyrant has Fled but the Infidel and Tyrannical System Remains," al-Andalus Media, January 28, 2011.

19 Ibid.

20 Ibid.

21 Camille Tawil, "Ennahda Reconsiders Ties With Tunisian Salafist Groups," *al-Hayat*, May 29, 2013.

22 "To the Wise Men of the Islamist Movement in Tunisia," al-Andalus Media, October 22, 2012.

23 Ibid.

24 Ibid.

25 Ibid.

26 Ibid.

27 "Call to the Youth of Islam: To Those Who Aspire to Hijrah in the Way of God in the Islamic Maghreb in General and Tunisia in Particular," al-Andalus Media, March 17, 2013.

28 Ibid.

29 Abu Yahya al-Shinqiti, "O People of Tunisia of al-Qayrawan Be Supporters of Shari`a al-Rahman," al-Andalus Media, May 18, 2013; Shaykh Ahmad Abu `Abd al-Ilah, "Letter of Advice to Ennahda in Tunisia of al-Qayrawan," al-Andalus Media, May 21, 2013.

30 Al-Ilah.

31 Ibid.

Appendix HEFFELFINGER02

cautious of the Tunisian government's "provocations," and to demonstrate "patience and wisdom" rather than being dragged into "random actions."[32] AQIM's public stance has been to counsel AST not to escalate its conflict with the government, so that it can maintain its ability to undertake *da`wa* openly. At the same time, AQIM serves as an ominous presence in the background, a dark cloud hanging over Tunisian society.

AQIM Operations in Tunisia

When the Algeria-based Salafist Group for Preaching and Combat (GSPC) became AQIM in 2007, one goal behind its founding was to integrate into one structure the various jihadist groups in North Africa, including the Tunisian Combatant Group (TCG). The TCG's involvement with the GSPC/AQIM, however, was limited for a number of years. Early Tunisian global jihadist activity, such as the arrangement for two Belgian-Tunisians to kill Afghan mujahidin commander Ahmad Shah Massoud in 2001, involved direct links to al-Qa`ida's core organization rather than to GSPC/AQIM.[33] The GSPC and later AQIM were regarded as largely focused on Algeria.

"AQIM's public stance is to counsel AST not to escalate its conflict with the government, so that it can maintain its ability to undertake da`wa openly. At the same time, AQIM serves as an ominous presence in the background, a dark cloud hanging over Tunisian society."

This is not to say that Tunisian jihadists were not involved with their Algerian counterparts during this period. TCG co-founder Tarek Maaroufi worked with GSPC networks in Europe, Tunisians trained in GSPC/AQIM camps, and some Tunisiasn radicals were captured or killed by Algerian security services.[34] Still, this cooperation appears limited.

Yet with the Arab uprisings and the fall of Ben Ali, AQIM's reported presence in Tunisia became more visible. In May 2011, a gunfight that killed two Tunisian security force members (including a colonel) broke out in the southern Tunisian town of Rouhia after security forces discovered men trying to transport weapons.[35] The militants, reportedly AQIM members, included Algerians, Libyans, and Tunisians, two of whom were, according to Tunisian authorities, part of a group of fighters who tried to infiltrate and recruit fighters in Tunisia in 2006 and 2007 known as the Suleiman Group.[36] The next year, in Bir Ali Ben Khalifa, fighting between suspected militants and security forces broke out after authorities tried to stop arms smugglers.[37] The clashes resulted in the deaths of two militants and the recovery of 34 automatic weapons and nearly $55,000 in cash.[38] According to then-Tunisian Interior Minister (now prime minister) Ali Laarayedh, some of the 12 men arrested after the incident had links to AQIM.[39]

"It is likely that the tenor of the relationship between AST and the Ennahda-led government will continue to deteriorate, as will the relationship between AQIM and the Tunisian state."

During this period, according to Algerian officials, the flow of arms increased through and within Tunisia (from both Libya and Algeria), a trend that became noticeable in the Algerian provinces of El Oued and Tebessa, which border Tunisia, as well as in close inland provinces like Khenchela and Bouira.[40]

In December 2012, the situation escalated with the emergence of what Tunisian officials termed AQIM's attempt to set up a branch in Tunisia, the Uqba ibn Nafi Brigade.[41] Following security incidents that month in Kasserine and the arrests of 16 militants reputedly linked to AQIM, Laarayedh publicly identified the group, saying that it was led by Algerians close to AQIM leader

32 Al-Shinqiti, "O People of Tunisia of al-Qayrawan Be Supporters of Shari`a al-Rahman."

33 Benjamin Roger, "Tunisie: Abou Iyadh, au nez et à la barbe des autorités," *Jeune Afrique*, September 18, 2012.

34 For information on Maaroufi, see "The United States and Italy Designate Twenty-Five New Financiers of Terror," U.S. Treasury Department, August 29, 2002; Christophe Dubois and Frédéric Vézard, "Le Tunisien d'Al-Qaïda fréquentait un cybercafé de Menton," *Le Parisien*, October 22, 2002. For information on Tunisians training in GSPC/AQIM camps, see Anneli Botha, "Terrorism in the Maghreb: The Transnationalisation of Domestic Terrorism," Institute for Security Studies, July 2008. For information on Tunisians killed or captured by Algerian authorities, see Mounir B., "Six terroristes tunisiens arrêtés à Annaba," *Le Quotidien d'Oran*, April 17, 2005.

35 Frida Dahmani, "Tunisie: échanges de tir meurtriers entre militaires et membres présumés d'Aqmi," *Jeune Afrique*, May 18, 2011.

36 Contemporary reporting identified these two as Sofiane Ben Amor and Abdelwaheb Hmaied. See Frida Dahmani, "Tunisie: échanges de tir meurtriers entre militaires et membres présumés d'Aqmi," *Jeune Afrique*, May 18, 2011. The Suleiman Group was built around several jihadists, led by a former Tunisian national guardsman named Lassad Sassi. Sassi had trained with the GSPC in Algeria, then slipped into Tunisia and recruited and trained as many as 40 men before the Suleiman Group was broken up by Tunisian security forces in early 2007. The group, however, had difficulties with funding and equipment, and later analysis has theorized that the group posed somewhat less of a threat to Tunisian security than initial reporting suggested. See, for example, Alison Pargeter, "The Suleiman Affair: Radicalism and Jihad in Tunisia," *Jane's Intelligence Review* 23:1 (2011).

37 "Tunisie: violences et défi salafiste," International Crisis Group, February 13, 2013.

38 Zohra Abid, "Tunisie. La groupe de Bir Ali Ben Khalifa est lié à Al Qaïda," *Kapitalis*, February 13, 2012.

39 Ibid.

40 For one notable incident involving an arms convoy that fled from Algeria into Tunisia, see Georges Malbrunot, "Un mystérieux convoi armé stoppé en Tunisie," *Le Figaro*, September 23, 2011. Also see "À Bir El Ater, les contrebandiers préfèrent le trafic d'armes," *El Watan*, October 7, 2011.

41 Mourad Sellami, "Cellule d'Al Qaîda démantelée aux frontières: Les precisions du ministre de l'intérieur tunisien," *El Watan*, December 22, 2012.

Appendix HEFFELFINGER02

Abdelmalek Droukdel.[42] Laarayedh added that the Uqba ibn Nafi Brigade was composed largely of recruits from the region of Kasserine, which abuts Algeria, and that the group's goal was to provide initial training before sending fighters along to "real" AQIM camps in Algeria or Libya.[43]

After several months searching for militants in the region of Jebel Chaambi, Tunisia's highest mountain, and further north toward El Kef, Tunisian security forces struck a homemade mine on the mountain on April 29, 2013.[44] Between then and early June, at least 20 Tunisian soldiers were wounded, some seriously, and two killed by improvised explosive devices on the mountain.[45] Again, Tunisian officials identified those behind the bombs as the Uqba ibn Nafi Brigade, which they alleged included fighters who had spent time in northern Mali.[46] While Tunisian officials have claimed that the group has left the mountain, Algerian security forces have stepped up surveillance and border security operations over the last two months.[47] According to *El Watan*, Algerian forces have even conducted limited counterterrorism operations in Tunisian territory.[48]

Interestingly, at a time when Tunisian security services were hunting fighters in Jebel Chaambi and El Kef, militant-linked violence in the regions of Algeria bordering Tunisia appears to have increased. While it is unclear to what extent this violence may have been connected to developments in Tunisia or to the Uqba ibn Nafi Brigade, at least one incident in February 2013—a coordinated assault on an Algerian army base at Khenchela—was reportedly conducted by a group of fighters bearing a striking resemblance to the descriptions of the Uqba ibn Nafi Brigade.[49] This could bolster suggestions that the group is using Tunisia largely for recruitment and training, while focusing on combat operations elsewhere, such as in Algeria or Mali.

While Tunisian officials have charged that members of AST may have been involved in the Uqba ibn Nafi Brigade and the incidents around Jebel Chaambi, and that AST leader Abu Iyadh al-Tunisi has visited the mountain, this information is impossible to confirm and could be related to the tensions between the Ennahda-led government and AST.[50]

Conclusion

It is likely that the tenor of the relationship between AST and the Ennahda-led government will continue to deteriorate, as will the relationship between AQIM and the Tunisian state. AST has something to lose, however, if it moves toward direct confrontation with the state: the group has been able to provide social services to areas neglected by the government, and has positioned itself as a leading voice opposed to a system widely seen as failing. Although AST remains numerically small, it enjoys influence beyond its numbers. Going to war with the state would mean sacrificing its ability to engage in *da`wa* openly.[51]

Thus, while it is not clear that this is the result of coordination between the two groups, AQIM has served as a safety valve for AST. It has urged AST to deescalate its confrontations with the state, even while AQIM itself has adopted an increasingly harsh tone toward the Tunisian government and appears to be escalating its activities in the Algeria-Tunisia border region.[52]

Aaron Y. Zelin is the Richard Borow fellow at The Washington Institute for Near East Policy, and maintains the website Jihadology.net.

Daveed Gartenstein-Ross is a senior fellow at the Foundation for Defense of Democracies and an adjunct assistant professor at Georgetown University. In 2013, he was a visiting research fellow at the International Centre for Counterterrorism—The Hague, where he produced a major report on Ansar al-Shari`a in Tunisia.

Andrew Lebovich is a Washington, D.C.-based researcher and analyst focused on political and security issues in North Africa and the Sahel.

42 Dorra Megdiche Meziou, "Al Qaïda, menace-t-elle vraiment la Tunisie?" *Business News*, December 21, 2012.

43 Ibid.; Roger, "Tunisie: sur la trace des djihadistes du mont Chaambi." Other reports indicated the presence of purported training camps in Tunisia as early as last fall. See Malik Aït-Aoudia, "Tunisie: qui protège les camps jihadistes?" *Marianne*, October 23, 2012.

44 "Fifth Mine Explodes in Mount Chaambi, No Victims," Tunis Afrique Press, May 20, 2013.

45 Antoine Lambroschini, "Jihadists Hunted in Tunisia 'Former Mali Fighters,'" Agence France-Presse, May 8, 2013.

46 Ibid.

47 Amira Masrour, "Government Says Chaambi Attacks Planted by Actors Outside the Mountain," Tunisia Live, June 7, 2013; "Tunisie – l'armée algérienne soupçonne les terroristes de se déplacer sous la couverture de contrebandiers," *Business News*, June 10, 2013.

48 Aziz M., "Frontière algéro-libyenne: Le boom du trafic d'armes," *El Watan*, June 21, 2013; "Algérie: surveillance renforcé à la frontière tunisienne," Agence France-Presse, May 4, 2013. If true, the reports about Algerian deployments in Tunisia would mark a shift from Algeria's public stance on the deployment of Algerian forces outside of Algerian borders.

49 The attack was conducted by approximately 50 fighters, matching some estimates of the Uqba ibn Nafi Brigade's size, and reportedly involved a mix of Algerians, Tunisians, and Libyans. See "Algérie: un groupe armé attaque une caserne militaire," Radio France Internationale, February 7, 2013.

50 According to an *El Watan* article, "Concerning the origins of the people captured, Ali Laarayedh indicated that the majority of these terrorists are originally from the Kasserine governorate and were known for their activities in protests organized by Ansar al-Shari`a." See Mourad Sellami, "Les precisions du ministre de l'Interieur tunisien," *El Watan*, December 22, 2012. Also see Imed O., "Selon les services de renseignement spéciaux tunisiens: Abou Iyadh aurait visité les terroristes au mont Chaambi," *Liberté-Algérie*, June 15, 2013; Lebovich.

51 For its part, AQIM might lose traction if it found itself in an actual fight with the Tunisian state, as that could darken many Tunisians' perceptions of it.

52 Al-Ilah.

Appendix HEFFELFINGER02

Recent Highlights in Terrorist Activity

June 1, 2013 (UNITED STATES): Al-Qa`ida in the Arabian Peninsula (AQAP) released a new six-minute, English-subtitled audio message addressed "to the American People." The message said, "The Boston events…and the poisoned letters [sent to the White House], regardless of who is behind them, show that your security is no longer under control, and that attacks on you have taken off and cannot be stopped. Everyday you will be hit by the unexpected and your leaders will not be able to defend you." Qasim al-Raymi, who read the statement, is AQAP's military chief. – *Sky News, June 2; CBS/AP, June 2*

June 1, 2013 (IRAQ): The Iraqi Defense Ministry said that they foiled a plot by an al-Qa`ida cell to execute poison gas attacks in the Middle East, Europe and North America. The militants may have been planning to use remote-controlled toy planes to distribute the gas. Iraqi officials said that they arrested five suspects who had been using three facilities in Iraq to try to produce sarin and mustard gas using instructions from another al-Qa`ida group. The cell, however, had not yet been able to produce any lethal chemical agents. – *Fox News, June 1; BBC, June 1*

June1, 2013 (MAURITANIA): Mauritanian officials announced that U.S. authorities recently transferred Yunis al-Mauritani, a senior al-Qa`ida operative previously held in Afghanistan, to their custody. – *Reuters, June 1*

June 2, 2013 (SYRIA): A car bomb exploded near a police station in Jubar district of Damascus, killing at least eight Syrian security forces. It was not clear which group was behind the attack, but the Syrian Observatory for Human Rights suspected that it had been carried out by Jabhat al-Nusra, an al-Qa`ida-linked Salafi-jihadi militant group. – *France24, June 2*

June 2, 2013 (YEMEN): A suicide bomber tried to drive an explosives-laden vehicle into a gas pumping station in the southern port city of Balhaf, but gunfire from Yemeni guards detonated the vehicle prematurely. Besides the bomber, there were no casualties and the gas facility was unaffected. – *Reuters, June 2*

June 3, 2013 (GLOBAL): The U.S. State Department offered a $7 million reward for information leading to the location of Abubakar Shekau, the leader of Boko Haram in Nigeria. The U.S. government also authorized rewards up to $5 million each for information leading to the location of Yahya Abu al-Hammam, a leader in al-Qa`ida in the Islamic Maghreb (AQIM), and Mokhtar Belmokhtar, a former leader in AQIM who now heads his own militia. It also offered $3 million each for information leading to the location of AQIM leader Malik Abu Abdelkarim and Movement for Oneness and Jihad in West Africa (MUJAO) spokesperson Oumar Ould Hamaha. – *Daily Post, June 4*

June 3, 2013 (AFGHANISTAN): A suicide bomber on a motorbike attacked a passing U.S. military patrol in Paktia Province, killing two U.S.-led coalition service members. The explosion also killed at least 10 schoolchildren and a police officer. – *Guardian, June 3; Voice of America, June 3*

June 3, 2013 (INDONESIA): A suicide bomber on a motorbike detonated his explosives at a police compound in Poso in Central Sulawesi, killing only himself. A police source told Agence France-Presse that the bomber was suspected of being part of an Islamist terrorist group led by Santoso, Indonesia's most wanted militant. – *Australian Broadcasting Corporation, June 3; AFP, June 3*

June 4, 2013 (MALI): A suicide bomber blew himself up in the northern Mali city of Kidal after being chased by authorities. The bomber was the only casualty. – *Voice of America, June 4; Reuters, June 4*

June 5, 2013 (IRAQ): Gunmen ambushed a bus and executed 15 passengers on a desert road between Anbar Province and Karbala. The dead included 10 border police and five local residents. Authorities blamed al-Qa`ida in Iraq for the attack, saying, "Those terrorists hunt people along this road and kill according to religious sect." – *Reuters, June 5*

June 5, 2013 (SYRIA): Syrian troops and Hizb Allah fighters captured the strategic border down of Qusayr from anti-government rebels. According to Voice of America, "the rebels had held Qusayr for more than a year in their overall struggle to topple the government of President Bashar al-Assad. Its loss to government forces represented a major blow to rebels, whose most effective fighting units, the al-Qa`ida-linked al-Nusra Front [Jabhat al-Nusra], were defending the town." – *Voice of America, June 5*

June 6, 2013 (GLOBAL): Al-Qa`ida chief Ayman al-Zawahiri released a new audio message, urging Syrians to unite to overthrow President Bashar al-Assad and prevent what he claimed was a U.S. plot to establish a client state in Syria to protect Israel's security. – *Reuters, June 6*

June 6, 2013 (AFGHANISTAN): A suicide bomber drove a truck filled with explosives into a military base in Helmand Province, killing at least six Georgian soldiers. The Afghan Taliban claimed responsibility. – *Reuters, June 6*

June 6, 2013 (RUSSIA/SYRIA): The head of the Russian security service, the FSB, said that approximately 200 Russian Islamists who support North Caucasus insurgents are fighting in Syria for al-Qa`ida. "The fact that around 200 fighters from Russia are fighting on the side of the Caucasus Emirate under the flag of al-Qa`ida and other related structures provokes concern in Russia," FSB chief Alexander Bortnikov said. – *AFP, June 6*

June 7, 2013 (IRAQ): A suicide bomber in an explosives-laden vehicle targeted a convoy of buses carrying Iranian Shi`a pilgrims near Muqdadiya, killing at least nine of them. – *RFE/RL, June 7; BBC, June 7*

June 7, 2013 (PAKISTAN): A U.S. drone reportedly killed six suspected militants in the Shawal area of North Waziristan Agency in the Federally Administered Tribal Areas. Pakistan's *Dawn* newspaper said that it was the "first such attack since Nawaz Sharif was sworn in as prime minister this week." – *Dawn, June 8*

June 9, 2013 (IRAQ/SYRIA): Al-Jazira claimed to have obtained a copy of a letter from al-Qa`ida chief Ayman al-Zawahiri addressed to the leaders of the Syria-based Jabhat al-Nusra and the Iraq-based Islamic State of Iraq (ISI). In the letter, al-Zawahiri said that ISI chief Abu Bakr al-Baghdadi was "wrong" to declare a merger between the ISI and Jabhat al-Nusra, adding that al-Baghdadi's rule was limited to Iraq. Al-Baghdadi had appeared to unilaterally announce the merger of the two groups in April 2013, but was rebuffed by Jabhat al-Nusra, which said that they pledge allegiance to

Appendix HEFFELFINGER02

Ayman al-Zawahiri but that they would not merge with the ISI. – *al-Jazira, June 9; Voice of America, June 10*

June 9, 2013 (IRAQ): A suicide bomber detonated an explosives-laden vehicle at a police checkpoint outside a Shi`a district in northwest Baghdad, killing at least seven people. – *Reuters, June 9*

June 10, 2013 (AFGHANISTAN): Militants tried to attack NATO's headquarters at Kabul international airport. The attack, which involved suicide bombers, left seven militants dead, with Afghan authorities saying that the attack was thwarted. Two civilians were wounded. – *Guardian, June 10*

June 10, 2013 (IRAQ): Two car bombs exploded and a suicide bomber in another vehicle detonated his explosives at a produce market in Jadidat al-Shatt in Diyala Province. The three bombs killed at least 13 people. – *RFE/RL, June 10*

June 11, 2013 (SYRIA): Two suicide bombers detonated their explosives at a square in central Damascus, killing at least 14 people. – *AFP, June 11; Los Angeles Times, June 11*

June 11, 2013 (AFGHANISTAN): An Afghan Taliban suicide bomber in a vehicle attacked the entrance to Afghanistan's Supreme Court in Kabul, killing at least 17 people. – *AFP, June 11; al-Jazira, June 12*

June 15, 2013 (IRAQ/SYRIA): The head of the Islamic State of Iraq (ISI), Abu Bakr al-Baghdadi, released a new audio message appearing to defy the orders of al-Qa`ida chief Ayman al-Zawahiri not to merge with Jabhat al-Nusra, saying that "the Islamic State of Iraq and the Levant remains [in existence], as long as there is blood in our veins and our eyes are blinking…It remains, and we will not compromise; we will not give up…until we die." Al-Baghdadi also said, "I chose the command of God over the command that runs against it in the [al-Zawahiri] letter." – *AFP, June 15; AP, June 15*

June 15, 2013 (PAKISTAN): A female suicide bomber detonated explosives on a bus carrying students of the Sardar Bahadur Khan Women's University in Quetta. After the explosion, a male suicide bomber was waiting inside the BMC Hospital ward for the arrival of high level officials, and he blew himself up when they arrived. In total, 25 people were killed, including 14 students of the women's university, four nurses and Deputy Commissioner of Quetta Abdul Mansoor Kakar. The sectarian militant group Lashkar-i-Jhangvi claimed responsibility. – *Dawn, June 16*

June 16, 2013 (IRAQ): A suicide bomber detonated explosives inside a crowded café in a largely Shi`a neighborhood of Baghdad, killing 11 people. – *AP, June 16*

June 16, 2013 (PAKISTAN): Two anti-polio vaccination volunteers were shot and killed while visiting homes in Khyber Pakhtunkhwa Province. – *CNN, June 17*

June 17, 2013 (KENYA): Kenyan police shot to death two suspected militants after they refused to surrender during a raid on a house in Mombasa. Police accused the men of plotting attacks in Kenya on behalf of Somalia's al-Shabab militant group. – *Reuters, June 17*

June 18, 2013 (IRAQ): Two suicide bombers shot to death the guards outside the Habib ibn Mathaher Shi`a mosque in Baghdad, and then detonated their explosives inside the building. The double suicide attack killed at least 31 people. – *BBC, June 18*

June 18, 2013 (PAKISTAN): A suicide bomber detonated explosives in a crowd of hundreds of mourners attending a funeral in the village of Sher Garh in Mardan District of Khyber Pakhtunkhwa Province, killing 29 people. Among the dead was a recently elected lawmaker who may have been the target. The lawmaker, Imran Khan Mohmand, campaigned as an independent, but later supported the party of cricketer-turned-politician Imran Khan. – *AP, June 18*

June 19, 2013 (IRAQ): A suicide bombing killed the leader of a provincial political party and four of his relatives in Ninawa Province. The leader, Yunus al-Ramah, was the head of the United Iraq Party. The incident occurred one day before elections in Ninawa and Anbar provinces, two Sunni-majority provinces where the April 20 polls were delayed due to security concerns. – *AFP, June 19*

June 19, 2013 (YEMEN): A suicide bomber on a motorbike detonated his explosives in the middle of a market in Yemen's mainly Shi`a city of Sa`da, killing two civilians. – *AFP, June 19*

June 19, 2013 (SOMALIA): Al-Shabab militants launched a major attack on the UN Development Program compound in Mogadishu, killing at least 15 people, including four foreigners. – *BBC, June 19; Bloomberg, June 19*

June 20, 2013 (UNITED STATES): The U.S. federal government charged Eric Harroun with conspiring to provide material support to a designated terrorist organization and with conspiring to use weapons of mass destruction overseas. Harroun, a former U.S. Army medic from Arizona, allegedly fought in Syria against the Bashar al-Assad regime in coordination with the al-Qa`ida-linked Jabhat al-Nusra militant group. He was arrested near Dulles International Airport in Virginia in March 2013. – *Politico, June 21*

June 20, 2013 (AFGHANISTAN): The Afghan Taliban announced that they are ready to free U.S. Army soldier Bowe Bergdahl in exchange for five of their senior operatives held at Guantanamo Bay. Bergdahl, the only known American captive from the war in Afghanistan, has been held by the Afghan Taliban since 2009. – *AP, June 20*

June 20, 2013 (IRAQ): A suicide bomber killed three people at a vote counting center in Ramadi, Anbar Province. – *AFP, June 21*

June 21, 2013 (SPAIN): Spanish security forces broke up an al-Qa`ida-linked network accused of sending fighters to Syria, arresting eight people. The raids occurred in Ceuta, a Spanish territory in North Africa. – *AFP, June 21*

June 21, 2013 (PAKISTAN): A bomb tore through a Shi`a Muslim religious center in Peshawar, killing at least 15 people. – *Voice of America, June 21*

June 22, 2013 (IRAQ): A suicide bomber detonated his explosives inside a Shi`a mosque in northern Baghdad, killing at least 12 people. – *Reuters, June 22*

June 23, 2013 (PAKISTAN): Tehrik-i-Taliban Pakistan (TTP) militants killed nine foreign mountain climbers, including a U.S. citizen, at a camp at Nanga Parbat, the world's ninth-tallest mountain. The

Appendix HEFFELFINGER02

CTC Sentinel Staff

Editor-in-Chief
Erich Marquardt
Senior Editor, CTC

Editorial Board
COL Cindy R. Jebb, Ph.D.
Department Head
Department of Social Sciences (West Point)

COL Suzanne Nielsen, Ph.D.
Deputy Department Head
Department of Social Sciences (West Point)

MAJ Bryan Price, Ph.D.
Director, CTC

CONTACT
Combating Terrorism Center
U.S. Military Academy
607 Cullum Road, Lincoln Hall
West Point, NY 10996
Phone: (845) 667-6383
Email: sentinel@usma.edu
Web: www.ctc.usma.edu/sentinel/

* For Press Inquiries: (845) 667-6383

SUPPORT
The Combating Terrorism Center would like to express its gratitude to its financial supporters, for without their support and shared vision of the Center products like the CTC Sentinel could not be produced. If you are interested in learning more about how to support the Combating Terrorism Center, please visit http://www.ctc.usma.edu or call Kristin Sorenson at West Point's Association of Graduates at 845-446-1561.

The views expressed in this report are those of the authors and not of the U.S. Military Academy, the Department of the Army, or any other agency of the U.S. Government.

TTP said that the attack was in retaliation for the U.S. drone strike that killed Waliur Rahman, the second-in-command of the TTP, on May 29, 2013. – *Washington Post, June 23*

June 23, 2013 (SYRIA): A car bomb exploded in Aleppo in northern Syria, killing 12 soldiers. Jabhat al-Nusra, a Salafi-jihadi group linked to al-Qa`ida, claimed responsibility. – *AP, June 23; AP, June 25*

June 23, 2013 (SYRIA): Three suicide bombers detonated their explosives while trying to break into the Rukn al-Din police station in northern Damascus, killing five people. Jabhat al-Nusra, a Salafi-jihadi group linked to al-Qa`ida, claimed responsibility. – *AP, June 23; AFP, June 23; AP, June 25*

June 24, 2013 (IRAQ): A suicide bomber killed an Iraqi counterterrorism official as he left a university in Tikrit. The official, Colonel Ghazi Ali al-Jubouri, was the head of counterterrorism for Salah al-Din Province. – *Reuters, June 24*

June 25, 2013 (AFGHANISTAN): Five Afghan Taliban gunmen and suicide bombers targeted the Ariana Hotel, which the Taliban claim is the location for a CIA base, and the presidential palace in Kabul. The bombers drove at least two land cruisers similar to those used by NATO soldiers, as well as fake badges and vehicle passes, which allowed at least one of the bombers to gain access to a heavily guarded area. The Afghan Interior Ministry said that three Afghan security guards and all five assailants were killed. – *New York Times, June 24; AFP, June 25*

June 25, 2013 (IRAQ): Two suicide bombers detonated explosives at a protest camp near Tuz Khurmato, Salah al-Din Province, killing at least 10 people. The protesters were reportedly ethnic Turkmen demanding better security for the area. The dead included Ali Hashim Muhtaroglu, the deputy head of the Torkoman Front, along with two of his sons. – *BBC, June 25; RFE/RL, June 26*

June 26, 2013 (UNITED STATES): U.S. federal authorities announced that Justin Kaliebe, an 18-year-old Long Island high school student, pleaded guilty to providing material support to terrorists in a secret hearing on February 8, 2013. Kaliebe was allegedly inspired by Usama bin Ladin and attempted to travel to Yemen to join al-Qa`ida operatives waging "violent jihad." According to NBC News, "Investigators said Kaliebe approached an undercover officer about joining a terror group and admitted he followed teachings of terrorist Yemeni cleric Anwar al-Awlaki and Sheik Omar Abdel-Rahman—the blind sheik tied to the 1993 World Trade Center bombing." He was arrested on January 21, 2013, at Kennedy Airport trying to fly to Yemen. – *NBC News, June 26*

June 26, 2013 (PAKISTAN): A bomb exploded on a busy street during rush hour in Karachi. The bomb targeted Judge Maqbool Baqir, who has worked on a number of terrorism cases. The judge and eight members of his security detail were killed in the blast. – *AFP, June 26*

June 27, 2013 (UNITED STATES): A federal grand jury in Boston indicted Dzhokhar Tsarnaev in the Boston Marathon bombings, charging him with detonating a weapon of mass destruction. Tsarnaev could face the death penalty or life in prison. – *Los Angeles Times, June 27*

June 27, 2013 (SYRIA): A suicide bomber killed at least four people in a Christian area of Damascus. – *AAP, June 28*

June 27, 2013 (AFGHANISTAN): A suicide bomber attacked a NATO convoy in Farah Province, killing two civilians. – *AP, June 29*

June 29, 2013 (THAILAND): A roadside bomb killed eight Thai soldiers in Yala Province in southern Thailand. Authorities suspected that Muslim separatist insurgents were to blame for the attack. – *Bangkok Post, June 29*

June 30, 2013 (PAKISTAN): A suicide bomber killed at least 28 people in a Shi`a area of Quetta, Baluchistan Province. The sectarian militant group Lashkar-i-Jhangvi claimed responsibility. – *Express Tribune, June 30; BBC, June 30; Dawn, July 1*

June 30, 2013 (PAKISTAN): A remotely-detonated bomb targeted security forces in Peshawar, leaving 17 people dead. – *BBC, June 30; UPI, June 30*

Appendix HEFFELFINGER02

MARC View

LEADER 01778nam a22002652 4500			
001 22613706			
003 SIRSI			
005 20180314130447.0			
006 m d			
007 cr n			
008 160125e20150825nju s||||||| 2|eng|d			
020			a| 9781509020942
020			a| 1509020942 (Spiral) c| USD 417.00 Retail Price (Publisher) 9| Active Record
037			b| 00066573
040			a| BIP US d| WaSeSS
110	2		a| IEEE Staff e| Author
245	1	0	a| 2015 IEEE ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) h| [electronic resource]
260			a| Piscataway : b| IEEE c| Aug. 2015
506			a| License restrictions may limit access.
520	8		a| Annotation b| The international conference on Advances in Social Network Analysis and Mining (ASONAM 2015) will primarily provide an interdisciplinary venue that will bring together practitioners and researchers from a variety of SNAM fields to promote collaborations and exchange of ideas and practices ASONAM 2015 is intended to address important aspects with a specific focus on the emerging trends and industry needs associated with social networking analysis and mining The conference solicits experimental and theoretical works on social network analysis and mining along with their application to real life situations
521			a| Scholarly & Professional b| IEEE
655		0	a| Electronic books.
856	4	0	u| http://SK8ES4MC2L.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001947865&title=2015%20IEEE%20ACM%20International%20Conference%20on%20Advances%20in%20Social%20Networks%20Analysis%20and%20Mining%20%28ASONAM%29
955			a| Books In Print, (c) 2017 R.R. Bowker LLC
949			a| Electronic resource w| ASIS c| 1 i| 22613706-1001 l| ONLINE m| ONLINE r| Y s| Y t| ONLINE u| 3/20/2018 o| .STAFF. 360 MARC

ASK A LIBRARIAN

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368

Libraries Intranet (Staff Only)

Accessibility Help

Website Feedback

Policies and Guidelines

Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

FaceBook

Twitter

Instagram

2015 IEEE ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) [electronic resource]

Corporate Author IEEE Staff

Published Piscataway : IEEE Aug. 2015

ASK A LIBRARIAN

Access Online

serialssolutions.com

Restrictions on Access

License restrictions may limit access.

Summary

Annotation The international conference on Advances in Social Network Analysis and Mining (ASONAM 2015) will primarily provide an interdisciplinary venue that will bring together practitioners and researchers from a variety of SNAM fields to promote collaborations and exchange of ideas and practices ASONAM 2015 is intended to address important aspects with a specific focus on the emerging trends and industry needs associated with social networking analysis and mining The conference solicits experimental and theoretical works on social network analysis and mining along with their application to real life situations

Genre(s)

Electronic books

ISBN

9781509020942
1509020942 (Spiral)

Audience Notes

Scholarly & Professional IEEE

View MARC record | catkey: 22613706

PennState

Copyright @2019 The Pennsylvania State University. All rights reserved. Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 license. Details and exceptions.

Legal Statements | PSU Hotlines

PENN STATE UNIVERSITY LIBRARIES

Libraries Home (814) 865-6368
Libraries Intranet (Staff Only)
Accessibility Help
Website Feedback
Policies and Guidelines
Staff Directory

CONNECT WITH PENN STATE UNIVERSITY LIBRARIES

Facebook
Twitter
Instagram

2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Identifying Digital Threats in a Hacker Web Forum

Mitch Macdonald, Richard Frank*, Joseph Mei, and Bryan Monk
School of Criminology, Simon Fraser University
British Columbia, Canada
{ mcmacdon, rfrank, jmmei, bmm8 } @sfu.ca

Abstract — **Information threatening the security of critical infrastructures are exchanged over the Internet through communication platforms, such as online discussion forums. This information can be used by malicious hackers to attack critical computer networks and data systems. Much of the literature on the hacking of critical infrastructure has focused on developing typologies of cyber-attacks, but has not examined the communication activities of the actors involved. To address this gap in the literature, the language of hackers was analyzed to identify potential threats against critical infrastructures using automated analysis tools. First, discussion posts were collected from a selected hacker forum using a customized web-crawler. Posts were analyzed using a parts of speech tagger, which helped determine a list of keywords used to query the data. Next, a sentiment analysis tool scored these keywords, which were then analyzed to determine the effectiveness of this method.**

Keywords—hacking; critical infrastructure; natural language processing; sentiment analysis

I. INTRODUCTION

Cyber-attacks targeting the computer and control systems of critical infrastructures are increasing in both scope and prevalence. There have been reports of dramatic global increases in attacks targeting power grids, financial institutions, and transportation networks [17]. For example, approximately one quarter of organizations operating critical infrastructure report malware attacks [21]. Due to the connected nature of the Internet and the lack of capable guardianship in the form of computer security, malware is easily spread over the Internet. Virtual infrastructures are frequently targeted in these attacks to gain access to massive databases of information. In 2014 alone, three of the ten largest data breaches in history occurred and more than 822 million records were exposed worldwide, exceeding 2013's statistics [22]. The estimated proceeds of cyber-attacks surpass $US 1 trillion per year, affecting millions of individuals, corporations, and governments worldwide [32].

One strategy to combat this problem is to monitor and analyze the communal networks of the cyber criminals who commit these attacks. More specifically, their communication activity can be analyzed to determine potential threats. Chapter II provides background on critical infrastructure, before implicating hacker communities as threats to these systems. The characteristics of hacker communities are then discussed and the motivations for the current study are contextualized; however, analyzing virtual communities, such as online discussion forums (or simply, forums), poses serious methodological challenges due to the large volume of posts made to these websites. These issues are addressed in Chapter III. First, we use a customized web-crawler specifically designed to collect data from discussion forums. Second, we use two analytic tools to analyze the language of forum members: 1) a parts of speech (POS) tagger to extract nouns from the data (of which 16 were selected as keywords); 2) identify interesting posts by cross-referencing keywords that potentially discuss critical infrastructure with those discussing hacking tools and hacker jargon; and 3) assign sentiment scores to the previously identified posts to highlight those that are the most relevant. Results are presented in Chapter IV, which indicate the most negative sentiment is directed toward the financial and government sectors of the critical infrastructure system. Finally, in Chapter V, the implications of the findings are discussed and contextualized within the current literature. The limitations associated with this study and directions for future research are also discussed. Chapter VI concludes the study.

II. BACKGROUND

A. Critical Infrastructure

Critical infrastructures are the physical and virtual systems, facilities, technologies, networks, assets, and services that support the health, safety, security, and economic well-being of both individuals and communities, as well as the functioning of government [26, 27, 31]. These include, but are not limited to, government, healthcare, energy, information and communication technology, and transportation systems. Critical infrastructures are often interconnected and interdependent within and across provincial, state, or territorial boundaries, as well as international borders [23]. These interconnected systems are massive, geographically dispersed, and are controlled by many different players—mostly within the private sector [23]. However, critical infrastructures are increasingly being connected to the Internet, making them vulnerable to cyber-attacks that cause physical damage, disrupt vital services, or threaten the economy [1, 24].

As critical infrastructures become increasingly digitalized, cyber-attacks targeting these systems are predicted to continually increase. The most industrialized countries might also be the most vulnerable to these attacks, because they are the most dependent on digital networks [2]. Diverse groups of actors, including nation-states, terrorist groups, and criminal organizations, methodically target and plan attacks against critical infrastructures [25]. Cyber-attacks target the computer networks, industrial control systems, or supervisory control and data acquisition systems that operate and monitor their physical systems and are carried out to shut down these networks, delete files, manipulate equipment, or steal sensitive information [5]. However, the expansion of the Internet has shifted global power dynamics such that other non-state actors, including lone wolf or small groups of hackers, are now capable of attacking critical infrastructures [5, 28]. For this reason, it becomes pertinent to analyze hacker communities.

© 2015 ACM. ISBN 978-1-4503-3854-7/15/08 $15.00
DOI: http://dx.doi.org/10.1145/2808797.2808878
Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

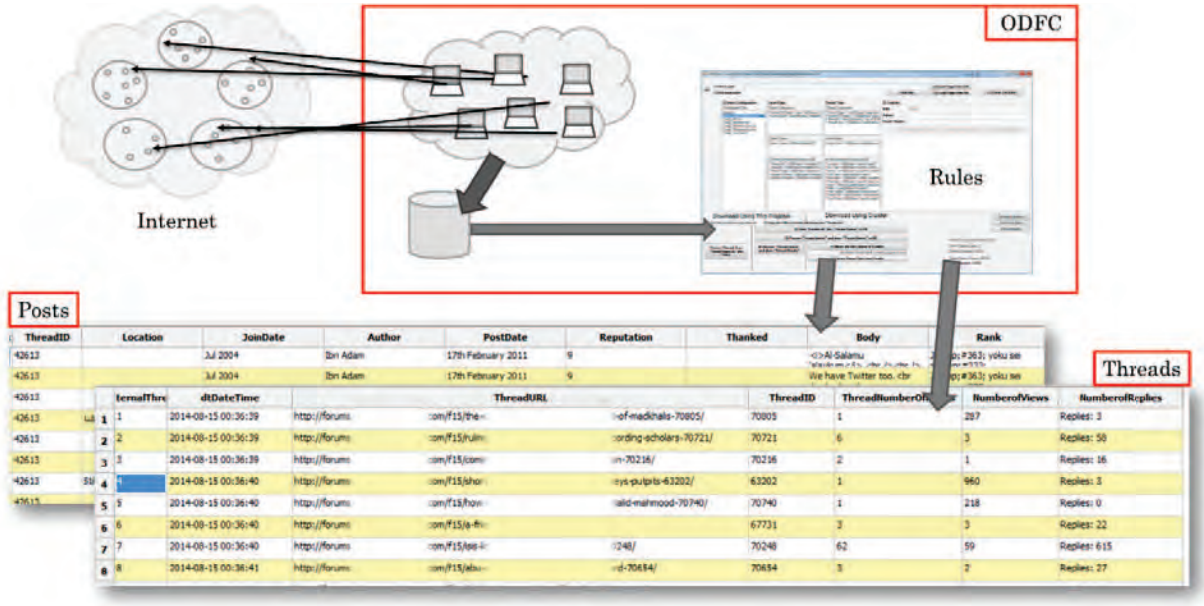


Figure 1 - A summary of the data capture process

B. Hacker Communities

Hacker communities are loosely constructed social networks that facilitate the exchange and distribution of information [4, 6]. Hacker communities are dark networks, so members explicitly protect their anonymity, which serves to counter detection from law enforcement [29]. One way to protect anonymity is by communicating over the Internet. For this reason, discussion forums are a popular method of communication amongst hacker communities [6]. Approximately 140 hacker forums have been identified on the public Internet that cater to English-speakers, although this figure is likely a gross underestimation of the prevalence of these online communities [30]. Many of these forums are freely accessible, which means that membership is not required to access content. Simply put, anyone who has access to the public Internet has the ability to view information exchanged through open forums. As noted by Eric Filol, this concept is important, because, "the proliferation of open data, not cyber warfare, poses the larger threat to [a country's] critical infrastructures" [8]. It is then relevant to not only examine the extent to which critical infrastructures are discussed in hacker forums, but also the language and sentiment used by forum members when referring to critical infrastructures to better understand the threats these communities pose [3].

C. The Current Study

Much of the current literature has focused on describing or developing typologies of cyber-attacks, but has not directly examined the individuals implicated in these crimes. Through a dual approach of POS tagging and sentiment analysis, this study makes two primary contributions: 1) to the cyber security detection methods and analyses used to identify open source data (exchanged freely over the internet) and potential threats to critical infrastructure; and 2) to determine the extent to which and how critical infrastructures are expressed in hacker forums. The end goal of this research is to better understand the context around the use of selected keywords; therefore, we believe these methods are complementary to one another. Following keyword selection, the sentiment of posts cross-referencing 'critical infrastructure' with 'hacking tools' and 'hacker jargon' were measured. These posts were then analyzed to determine the context in which they were used. In other words, we used sentiment analysis as a mechanism to identify potential threats cross-referencing critical infrastructure and the tools or strategies used to attack them. Ultimately, our goal is to develop automated tools that combine these methods to streamline data analysis and facilitate real-time monitoring to identify potential threats as they emerge.

III. Methods

A. Data collection

Data was collected using software called the Open Discussion Forum Crawler (ODFC). ODFC is a customized web-crawler designed to capture the content posted to discussion forums; the software captures information from a user-selected forum by downloading its webpages, parsing the page apart with the use of forum-specific 'rules' to capture all the useful information present on the forum, which is then stored in a corresponding database. For an overview, see Figure 1. The database is designed to resemble the structure inherent to discussion forums and is navigated in the same way: each forum has many sub-forums, which in turn have many threads of discussion, each with at least one post.

Due to the thematic grouping of sub-forums and the chronological listings of threads, each webpage within a discussion forum follows a specific structure. A sample structure is shown in Figure 2, which illustrates multiple sub-forums displayed one after the other. Each sub-forum has a name, followed by a description of the topic(s) of discussion, the number of replies, and views. Figure 3 is similar to Figure 2, but

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02



Figure 2 - Forum content



Figure 3 - Thread content

```
<HTML>
    <DIV>
     <A href="index.html">Home</A>
    </DIV>
</HTML>
```

a) Ideal scenario

```
<HTML>
    <DIV>
     <A href="index.html">Home</A>
</HTML>
```

b) Real-Life result

Figure 4 - Ideal vs real-life HTML

depicts the structure of posts, rather than sub-forums. Posts exhibit the same implicit structure, as each post repetitively lists the *Author*, *body*, and *date* (for example). This structure is consistent from post-to-post and across threads, but usually varies from forum to forum. ODFC downloads each webpage (see Chapter III.A.1) and then performs extensive clean-up (Chapter III.A.2). Users then input rules, so that the software extracts relevant data (Chapter III.A.3), which can then be formatted through the use of 'patterns' (Chapter III.A.4).

1) Data Capture

The data-capture process starts with ODFC being given a list of sub-forums to retrieve (or a list of all sub-forums to download when the goal is to capture the entire forum). The process is as follows: 1) ODFC retrieves the first page of each sub-forum; 2) based on user-specified rules (see Chapter III.A.3), it detects the number of pages in the sub-forum, and calculates the URLs for all the pages; 3) for each URL (i.e. each page of the sub-forum), ODFC retrieves the content of the webpage; 4) rules are applied to the webpage to determine relevant thread information—the URL of each thread and the number of webpages it is spread through; 5) ODFC downloads all thread-pages sequentially and rules are applied to it to retrieve targeted data in each post (*author*, *date*, *body* of the post, etc.) which are then stored in a database. Once the entire forum is downloaded into the database, the data can be extracted in various ways, such as querying keywords and specific authors, or even constructing a social network.

2) Cleaning up the webpage

Before data can be extracted from a webpage, it has to be identified. To find relevant information, ODFC uses *rules* to retrieve various components from within the webpage. A *rule* is a {*path*, *pattern*} combination, where the *path* follows the *XPaths* standard for querying trees and *pattern* uses the standard *XQuery*. These techniques require the object being queried to be in a tree-like structure; however, HTML is not required to follow this requirement, as many website authors do not properly close tags. For example, in Figure 4, (b) has an unclosed <DIV> tag, while (a) shows the correct version. While web-browsers know how to interpret (b), it is not proper. Unclosed tags confuse the ODFC data retrieval process. Thus, retrieved webpages must first be scanned for any unclosed tags, which are then closed through the addition of HTML closing statements (such as the extra </DIV> in Figure 4b). The resulting HTML then follows a standard tree-structure, which can be translated into a XML file for further querying using our rules. This allows us to specifically pinpoint any detail in the webpage and store it in the database. Due to differences in structure, rules are customized for each forum.

3) Finding information within a webpage using paths

By applying a *rule* to a cleaned-up HTML webpage, ODFC is able to retrieve the exact information indicated in the rules (such as the *Author* of a thread). One of the components of a *rule* is the *path*. A *path* is defined as a sequence of branches from the root, using the format of the XPaths standard[1], and results in zero or more nodes in a tree. For example, to determine the author of the second post in Figure 5, it must first be recognized that there are three posts (attributed to *Author 1*, *Authors 2* and *Author 3*, respectively). To identify the various authors, a series of branches (*table*, *tr*, and finally *td*) must be traversed, resulting in a path "/html/body/table/tr/td". In Figure 4, this process results in three distinct authors (one for each post). To focus on the second author, the *rule* must be modified to traverse down specifically only on the second <*tr*> tag in <*table*>. The *rule* thus becomes "/html/body/table/tr[2]/td", which results in ODFC retrieving "*Author 2*".

4) Interpreting data through patterns

The application of a *rule* usually results in the proper data collection, but not necessarily in the proper format. That is, *rules* can be used to pinpoint where a data-element is, but cannot point to a piece of data within the larger data-element. For this reason,

[1] http://www.w3.org/TR/xpath/

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02



Figure 5 - HTML content as a tree.

Hacking Tools	# of References
botnet	1254
inject	1016
malware	671
virus	774
Trojan	501
-	-

Hacker Jargon	# of References
exploit	866
breach	341
backdoor	95
DDoS	918
SQL	1257
vulnerability	524

Critical Infrastructure	# of References
airline	312
bank	15 200
government	446
hospital	197
mobile	1302
-	-

Table 1 – Nouns extracted from hacker forum data used as keywords for sentiment analysis.

patterns are applied to eliminate unwanted text, perform counts, or perform mathematical operations on the results of the *path*. *Patterns* use the standard language of *XQuery*. Referring to Figure 5, to determine how many pages there are in this thread, the first step is to recognize this information is shown in the text ("Page 1 of 3") following traversing down <hr> then the <b> tag, resulting in *rule* "//hr/b". The end result is "Page 1 of 3", which contains the desired answer ("3") but is not in the specific format needed: an individual numerical value. To get to the "3", a *pattern* is created specific to that *path*. For example, the *pattern* "Page 1 of (?<RESULT1>)" would result in <RESULT1> = "3" being returned. This phrasing allows us to query multiple data-elements using a single *pattern*.

B. Forum selection

The hacker forum analyzed for this study was downloaded by ODFC in its entirety. This forum was selected for three reasons. First, it had large numbers of existing members, number of posts, and discussion threads. These characteristics increased the likelihood this forum contained relevant and analyzable data. Second, the topics of its sub-forums were relevant to malicious hacking, which broadly corresponds with the focus of this study. Finally, the discussion between forum members is freely or publicly accessible on the Internet, which means these individuals do not have any reasonable expectation of privacy. Adherence to this criterion satisfies ethical concerns associated with this study; however, to address issues concerning privacy and the potential uncovering of personally identifiable information, the domain of this hacker forum will not be identified, nor will the forum member's user names.

C. Keyword Extraction

A parts of speech (POS) tagger was used to analyze the body of the posts and identify linguistic features within the data. POS tagging is a computational technique that dissects text into nouns, verbs, adjectives, as well as other components and applies labels to them. POS tagging has been used across various disciplines to analyze the use of language. One area of focus involves using POS tagging for domain specific searches [11]. For example, Ferraro et al. found that natural language processing (NLP), with a dictionary specific to the clinical domain, was better able to identify specific text features in their samples [12]. While POS tagging is an effective tool in its own right, it is also useful in conjunction with other analysis tools. For example, Dalrymple found using POS tagging as a parsing method increased the efficiency of other established linguistic analysis methods [13]. In other words, it is a method for understanding text without the manual, labour-intensive efforts traditionally associated with large, text-based data. It has also been used for a myriad of other purposes, including supplementing other data analysis techniques, such as sentiment analysis [10].

To automate the analysis of the posts, OpenNLP was integrated into ODFC. OpenNLP is a POS tagger which dissects and categorizes supplied text, splitting sentences into their base components and then labelling each of those components based on their part of speech (nouns, adjectives, adverbs) [9]. The POS tagger then creates an enumerated list, detailing the amount of times a particular word occurs within the text. This list can be used to understand the relative frequencies of words within the text. For this study, we focused only on nouns. The POS tagger extracted a combined list of 11,432 nouns; however, not all were unique. Due to the design of the NLP software, issues regarding case sensitivity, the overlap between singular nouns and their pluralized form, and abbreviated or slang terminology particular to the hacker subculture all contributed to this greatly inflated number.

The list of nouns were manually revised and 16 applicable nouns were identified that had a relatively high frequency of in-text references and were considered to be potentially associated with malicious hacking and critical infrastructure (see Table 1). These keywords were used to query the database to arrive at a sample of posts, which were then cross-referenced and analyzed using sentiment analysis. The query function uses the root of the keywords, so that "virus" would retrieve posts containing both its singular and plural. Keywords were partitioned into three categories: 1) those that refer to hacking tools; 2) words specific to the hacker jargon; and 3) words potentially associated with critical infrastructure. Influenced by previous research, these keywords were selected based on their association with the hacker subculture and potential reference to critical infrastructure.

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

D. Sentiment Analysis

Sentiment analysis is a computational technique that attempts to analyze freeform text, by assigning *scores* to it after it is fed through a set of parameters or classifications. Sentiment analysis has been used primarily as a tool for analyzing user reviews of products and media [14]. Due to its ability to assign *positive* or *negative* scores to comments or reviews, companies can understand the general sentiment of online discussions; although sentiment analysis is also a useful tool for examining social networks and has been used in a variety of online contexts [15]. Within a security context, sentiment analysis has been used to examine the online activities of criminal subcultures, including extremist and hacker groups. For example, Li and Chen used deep learning-based sentiment to identify the key players involved in the sale of malware and stolen data [11].

For the current study, each of the critical infrastructure keywords were independently cross-referenced with each of the hacking tools and hacker jargon keywords; sentiment scores were computed for posts containing both types of keywords (henceforth, keyword pairing(s)). For example, each post that contained the keywords "bank" and "malware" would be assigned a sentiment score. In any instance that a post contained "bank", but not "malware" a score of 0 was assigned and were excluded from the analysis. It is important to note that keywords from the hacking tools and hacker jargon lists may have been used in the same post, so it is possible that posts may have occurred multiple times within the results.

Scores were calculated with the sentiment analysis tool SentiStrength—a dictionary based sentiment analysis program that can apply positive and negative values to text [16]. SentiStrength examines inputted text and then references its internal dictionaries to apply values or scores to it. More specifically, SentiStrength analyzes the words located on either side of keywords within a specified range. A custom 'Sentiment Wrapper' was used to expedite this process through simple automation of its command line inputs. Using this Sentiment Wrapper, the context surrounding each post for each keyword was analyzed. Specific words are given a numeric sentiment value based on the intensity of the sentiment. For example, "dumb" has a less negative sentiment value than "hate" [16]. Additionally, factors such as punctuation, capitalization, repetition, and specific emoticons can all affect sentiment values. In the context of this study, the output or score is a single value for each post and keyword combination. Each keyword was aggregated, both averaged and summed, across all posts to determine the sentiment around that keyword for the entire forum.

Sentiment scores are not limited or restricted to a specific range of numbers. That is, scores are entirely dependent on the negative or positive terms found in the analyzed text. Despite this seemingly infinite scale, extreme scores (both positive and negative) are unlikely. Due to the length of the average post, there are simply not enough words with associated sentiment values to garner extreme scores. While it is possible, the likelihood of a post with a sentiment score greater than 100 or less than -100 is low.

E. Contextual Analysis

The posts extracted by the POS tagger and analyzed with SentiStrength were also manually analyzed. A contextual analysis was used to determine the effectiveness of combining these methods to identify potential threats within this web forum, as well as the 'severity' or extent of these threats.

IV. Results

Table 1 displays the keywords used to query the data collected by ODFC, as well as the number of times they are referenced within the database. There were 5,134, 3,083, and 17,457 posts referencing the keywords in the "hacking tools", "hacker jargon", and "critical infrastructure" categories, respectively, for a gross total of 25,647 posts. Within and across categories, the POS tagger identified that certain keywords were more frequently used than others. For instance, "botnet" and "inject" were the most frequently referenced hacking tools with regards to the data collected by ODFC. Similarly, "SQL", "DDoS", and "exploit" were the most frequently used keywords within the hacker jargon. Finally, the number of posts referencing financial institutions were overwhelmingly overrepresented within the data, as illustrated by the 15,200 posts referencing the keyword "bank", compared to the other keywords relating to critical infrastructure. These figures suggest that certain keywords can potentially be used to distinguish between the level of interest afforded to particular infrastructure systems, as well as the most widely mentioned tools and jargon associated with attacks.

Table 2 displays a subset of the summary of sentiment scores for each pairing of the cross-referenced keywords. In this way, Table 2 provides a visual depiction of the potential threats (hacking tools and hacker jargon) expressed in this forum—specific to the critical infrastructure sector. Overall, the summary of sentiment scores indicates that posts mentioning hacker tools or hacker jargon along with "bank" and "government" grossed the largest scores, which were also negative. Rows without scores indicate that those keyword combinations did not occur within the data. For example, "hospital" and "backdoor" were not mentioned together in any post made to the forum. Furthermore, while the keywords "airline" and "hospital" were seldom cross-referenced, this is likely representative of the larger data, given the relatively few number of times they were independently referenced (Table 1).

In total, there were 688 posts made to this hacker forum that cross-referenced critical infrastructure keywords with hacking tool and hacker jargon keywords. This figure represents roughly 2.68 percent of the 25,647 gross number of references made to all 16 keywords. Again, some keywords were referenced more than others. From the hacking tools list, "botnet" was cross-referenced in 97 posts, while "Trojan", "malware", "inject", and "virus" were cross-referenced in 91, 86, 59, and 39 posts, respectively. From those words specific to the hacker jargon, "breach" was referenced along with critical infrastructure in 88 posts, whereas "exploit" and "vulnerability" were both cross-referenced with these keywords in 79 posts, and "DDoS", "SQL", and "backdoor" were cross-referenced in 40, 26, and 12 posts, respectively. With regards to critical infrastructure, "bank" was cross-referenced in 459 posts, "government" in 126 posts, "mobile" in 93 posts, and "airline" and "hospital" were

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

Keywords	airline		bank		government		hospital		mobile	
	Avg.	Sum	Avg.	Sum	Avg.	Sum	Avg.	Sum	Avg.	Sum
backdoor	-	-	-1.4	-7	-5	-15	-	-	-2	-8
botnet	-	-	-2.09	-140	-4.33	-13	-	-	-2.4	0
breach	-	-	-6.76	**-399**	**-7.7**	-154	-	-	**-8.57**	-60
DDoS	-	-	**-10.68**	**-299**	-4.86	-34	-	-	-5.75	-23
exploit	-4	-4	-2.15	-112	-4.25	-51	-4.5	-9	-1.78	-16
inject	-5	-5	-2.47	-89	-3.11	-28	-	-	-2.4	-12
malware	-	-	-4.71	**-245**	-3.9	-78	-3	-3	0	-63
SQL	-5	-5	**-12.44**	-112	-2.4	-24	-	-	-3.33	-20
Trojan	-	-	-3.15	**-189**	-0.75	-6	-	-	-2.69	-43
virus	-3	6	-3.6	-90	2.67	8	2	2	-1	-6
vulnerability	-3	6	-6.73	**-222**	**-7.71**	-185	-	-	-5.62	-73
* **Bold** indicates the largest average (white) and summed (gray) sentiment scores within the sample.										

Table 2 – Summary of average and summed sentiment scores for the cross-referenced keywords.

cross-referenced in 7 and 3 posts, respectively. More specifically, the three most frequent combinations of keywords all included "bank", which were cross-referenced with "botnet" in 80 posts, "Trojan" in 64 posts, and "breach" in 60 posts. Collectively, these posts represent 204 (29.65%) of the 688 total cross-referenced posts. These figures indicate the relative infrequency in which these keyword combinations were used in posts made to this hacker forum.

The sentiment scores for the keyword pairings (displayed in Table 2) generally indicate disparities between the average and summed scores. For instance, only the pairing of "DDoS" and "bank" was amongst the largest scores for both average and summed sentiment. The other largest average sentiment scores amongst the pairings were "SQL" and "bank", "breach" and "government", "vulnerability" and "government", and "breach" and "mobile". In contrast, the summed sentiment scores indicate the other largest sentiment scores were posts that cross-referenced "bank" with "breach", "malware", "Trojan", and "vulnerability". While the summed scores are attributed to the total number of posts that referenced these keywords, they may also suggest the likelihood of potential threats or risk to particular critical infrastructures.

A. Contextual Analysis

A content analysis of selected keyword pairings listed in Table 2 was used to gauge the accuracy of the analysis tools at identifying potential threats to critical infrastructure. The focus was on analyzing the context in which certain keyword pairings were used. Particularly, individual posts that had high sentiment scores were analyzed. For example, "SQL" was cross-referenced with similar frequency with both "bank" and "government"—these pairings were referenced in 9 and 10 posts, respectively—and used in similar contexts:

- "The emails are sourced exclusively from SQL databases only. Very high quality databases. More than 95 percent of them are sourced from government databases".
- "The hacker group Team GhostShell returns with a massive data leak targeting the banking, aerospace, legal, education, government, military, and airline sectors. The data dump contains database details, names, email addresses, mailing lists, addresses, and passwords. The dump also contains the SQL Injection vulnerable links."
- "Here is a website [hosting] sensitive information on banks, identified databases with bulk financial information, as well as simple mail transfer protocol (STMP) services."

Posts typically contained links to websites hosting information on banking and government institutions, information about recent data breaches from notorious hackers, advertised the sale of data stolen from computer databases, and websites that contained SQL injections and other hacking tools. Posts indicated that these websites, "... have some of the best tools to download. SQL Exploiter Pro v2.15 accesses databases and exploitable websites. It makes hacking websites easy. Pro hackers will love it for how organized it is. So simple, even novice hackers can use it."; however, "vulnerability" was used in varying contexts when cross-referenced with the keywords "bank" and "government". "Vulnerability" and "government" was used when referring to:

- "[Our group] doesn't only offer security for clients, but also hacking services—from owning your [corporate] competitors [or] breaking into [*.gov*] military-grade networks, financial institutions, or just sites you don't like."
- "A key actor in the [hacking] game is to always be ahead of the government. They are ahead of us when it comes to the key factor of technology, so to save yourself, here are some recommendations for guaranteed safety."
- "This may help the [hacker group Anonymous] at the end of the month, although I don't know how many of their members are on this forum. I'm going to list *.gov* websites that are vulnerable and explain how they are vulnerable to cross site scripting (XSS)."

In contrast, when used in conjunction with "bank", "vulnerability" was used to discuss the susceptibility of financial institutions in the operation of bank cards and how to fraudulently exploit these vulnerabilities. To a lesser degree custom written programs, such as web injections, were referenced as other methods that could be used to take advantage of these vulnerabilities. More commonly, this pairing was used to refer to the susceptibility of financial institutions against botnet services. As one post described, "It is possible to build large botnets, which [launch] brute force attacks in hours, or even minutes...Practically all banks are vulnerable, some more and some less."

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

As previously mentioned, "botnet" was almost always used in association with "bank", as indicated by the 80 (82.47%) references made to the pairing. The pairing of "botnet" and "bank" was used in the following contexts:

- "Current botnets available: Zeus (the latest public version with 18 private United States (US) bank [SQL] injections); Spyeye (the latest public version with private European Union and US bank [SQL] injections); and Citadel. I will contact the sellers on your behalf and set it up for you."
- "I have experience with botnet and spamming, looking for work. If you have any custom target, I can work on it, be it bank accounts [or anything else]."
- "Hello, I am one of actual speakers of Anonymous Russia. Today, Friday, February 3, 2012 at 12:00 GMT [we are] starting a lockout of the US economy (banks). This movement is supported by botnet [attacks]. This lockout is designed to punish banks for their involvement in the global crisis, but also to show Americans that no one can violate freedom on the Internet and those who try will be punished."

This suggests a predominant use of malware-infected networks to attack institutions in the financial sector. In this context, "botnet" was often used to advertise the sale of various types of malware, all of which are well known to the hacker community, including "Zeus", "Spyeye", and "Citadel". "Botnets" would generally be advertised with "special offers" of SQL web injections that could be used with the corresponding malware. Forum members that operated botnets would also advertise their services to others looking to attack banking websites and steal financial information. While these examples are specific to financial motives within the hacker community, the sentiment analysis tool also identified potential threats motivated by political ideologies. For example, a forum member claiming to represent the hacktivist group Anonymous encouraged other forum members to use or join botnets to support a campaign targeting US banks.

V. Discussion

Taken together, the largest, and most negative, average and summed sentiment scores were in reference to the keyword "bank". This is consistent with the argument that financial motives are pervasive within the contemporary hacker culture [7]; however, overall there is discrepancy between sentiment scores for "bank" that cross-reference the hacking tools and hacker jargon keywords. For example, "breach" had the largest negative summed score, but did not rank within the largest average scores. Similarly, "SQL" had the second largest negative average score, but was not among the largest summed scores. Thus, while the average scores indicate the most negative content, the summed scores indicate the most frequent pairings. This suggests that average scores may better indicate the most serious threats, while the summed scores indicate the most likely targets and sources of threat. For instance, the pairing of "bank" and "DDoS" had both the largest average and second largest summed sentiment scores. Financial institutions are popular targets for hackers and DDoS attacks are estimated to be the most frequent method of cyber-attack, accounting for 23 to 42 percent of all attacks against critical infrastructures [18, 19]. Thus, these findings invite speculation that sentiment may be correlated with the actual threat of attacks.

In comparison to other keyword pairings, those including "government" also had large sentiment scores. For instance, the pairing of "government" and "vulnerability" was amongst the largest average sentiment scores and also had a large summed sentiment score. These sentiments scores correspond with the increase in attacks targeting this sector. For example, reports indicate that the exploitation of vulnerabilities are attributed to 50 percent of all attacks against critical infrastructure systems, including those in the government sector [21]. More generally, "Trojan", "malware", and "virus" were cross-referenced in a gross total of 216 posts or 33.40% of the 688 posts that were analyzed; however, this figure is likely to be slightly inflated, due to the possibility that posts reference both "Trojan" and "virus". Nonetheless, this finding is consistent with statistics, which indicate that malware is responsible for approximately a third of cyber-attacks targeting critical infrastructure [20].

There were four main limitations associated with this study. First, sampling procedures ignored dialogue. In other words, the keywords used to query the data returned only those posts that contained the specified keyword pairings in the body of the text. Thus, the exchange of dialogue between forum members was missed; however, this limitation is inherent to our research design. While these methods missed the evolution of dialogue, they were efficient in identifying and extracting posts used for automated and manual analysis. Second, due to SentiStrength's intentionally broad base configuration, many words specific to the hacker jargon and subculture were likely ignored. For example, SentiStrength's dictionary is not built with negative connotations towards words like *infect*, *leak*, and *crack*, which limits its ability to understand the particular context in which these words are used. Future research on virtual hacker communities should take this into account by amending the SentiStrength dictionary with additional terms specific to hacker language; however, while user coded data may prove to be a valuable modification, an underlying issue remains: users may not be able to effectively label or assign a sentiment value to this subculture's lexicon. The third limitation surrounded the size and nature of the hacker forum. In regards to the nature of the forum, discussions about financial institutions were over-represented within the data. That is, posts identifying critical infrastructure were largely in reference to "banks" and other financial intermediaries, which provide the opportunity for monetary gain. This meant that "airlines" and "hospitals", which contain a wealth of personal information, but less financial information, were not as frequently discussed. This could be remedied by expanding the number of forums studied, as others may focus more on other interests. Finally, our use of NLP was limited to English, so relevant discussion in others languages was missed.

Future research using keyword-based queries and detection methods will use a stemming algorithm to extract keywords and prioritize critical infrastructures based on perceived threats, risks, and the relative importance of the services, assets, and security that they contribute to the functioning of society. Furthermore, while this hacker forum was large, it provided opportunity to determine the effectiveness of automated tools to identify potential threats within a large data source. That said,

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

future studies could consider smaller hacker forums for analysis. This research could compare whether or not smaller communities are more closed and tight-knit, which may potentially breed a more explicit criminal hacker subculture.

VI. Conclusions

Hacking is an ever growing issue in the modern world. With the rise of cloud computing and the tendency to store data outside of a local server, the threat to computer networks and data is ever present. This is especially true for critical infrastructures. Surveys show that these systems have become targeted in attacks and do not generally have up to date or sufficient digital protections. In some cases, hackers post about their attacks before or after they occur. This necessitates the development of analytic tools that can identify the language and targets of hackers, without manually monitoring the large volume of posts made to hacker forums. This study sought to produce one of these tools through the use of a customized web-crawler, POS tagging, and sentiment analysis tool. Specific keywords were pinpointed that resulted in some posts having larger sentiment values than others. Through this it was possible to determine which posts contained information regarding different facets of hacking, advertising of hacking services, requests for aid in hacking endeavors, and conversations surrounding major hacks perpetuated by well-known hacker groups.

Acknowledgements

The authors of this paper would like to thank Public Safety Canada, who, through their Cyber Security Cooperation Program, funded the research presented in this paper.

References

1) V. Benjamin and H. Chen. Securing cyberspace: Identifying key actors in hacker communities. Presented at Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on. 2012.

2) T. M. Chen. Stuxnet, the real start of cyber warfare? [editor's note]. Network, IEEE 24(6), pp. 2-3. 2010.

3) D. Décary-Hétu, B. Dupont and F. Fortin. "Policing the hackers by hacking them: Studying online deviants in IRC chat rooms," in Networks and Network Analysis for Defence and Security Anonymous 2014.

4) T. J. Holt. Subcultural evolution? examining the influence of on-and off-line experiences on deviant subcultures. Deviant Behav. 28(2), pp. 171-198. 2007.

5) T. J. Holt and M. Kilger. Know your enemy: The social dynamics of hacking. The Honeynet Project pp. 1-17. 2012.

6) T. J. Holt, D. Strumsky, O. Smirnova and M. Kilger. Examining the social networks of malware writers and hackers. International Journal of Cyber Criminology 6(1), pp. 891-903. 2012.

7) M. Kilger, O. Arkin and J. Stutzman. Profiling. The Honeynet Project (2 Nd Ed.), Know Your Enemy.Addison Wesley Professional 2004.

8) Shaw, G. (2014). Eric Filol on withdrawing from CanSecWest Vancouver: Talk potential blueprint for terrorists. [Online]. Available: http://blogs.vancouversun.com/2014/03/13/eric-filiol-on-withdrawing-from-cansecwest-vancouver-talk-potential-blueprint-for-terrorists/

9) Chambers, L., Tromp, E., Pechenizkiy, M. and Gaber, M. "Mobile sentiment analysis," in: Proceedings of the 16th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, (2012), San Sebastian.

10) Li, W., & Chen, H. "Identifying Top Sellers in Underground Economy Using Deep Learning-based Sentiment Analysis". Presented at Intelligence and Security Informatics (ISI), 2014 IEEE International Conference on. 2014.

11) Rayson, P., "From key words to key semantic domains," in International Journal of Corpus Linguistics, Vol. 13(4), (2008), p. 519-549

12) Ferraro, J., Daumé, H., DuVall, S., Chapman, W., Harkema, H., Haug, P., "Improving performance of natural language processing part-of-speech tagging on clinical narratives through domain adaptation," in American Medical Information Association, Vol. 20, (2013), pp. 931-939

13) Dalrymple, M., "How much can part-of-speech tagging help parsing?" in Natural language Engineering, Vol. 12(4), (2006), pp. 373-389

14) Dave, K., Lawrence, S., Pennock, D., "Mining the Peanut Gallery: Opinion Extraction and semantic classification of product reviews," in In Proceedings of the 12th international conference on World Wide Web (WWW '03). ACM, New York, NY, USA, 519-528.

15) Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), p. 82-88.

16) Thelwall, M., & Buckley, K. (2013). Topic-based sentiment analysis for the Social Web: The role of mood and issue-related words. Journal of the American Society for Information Science and Technology, 64(8), 1608–1617.

17) Brocklehurst, K. (2015, Feb 1). Cyberterrorists Attack on Critical Infrastructure Could be Imminent. [Online]. Available: http://www.tripwire.com/state-of-security/security-data-protection/security-controls/cyberterrorists-attack-on-critical-infrastructure-could-be-imminent/

18) Zorz, Z. (2014, Dec 31). 2015 predictions: Cyber attacks aimed at critical infrastructure, Aacks as a Service. [Online]. Available: http://www.net-security.org/secworld.php?id=17712

19) Lorhmann, D. (2015, Apr 17). Hacking Critical Infrastructure is Accelerating. [Online]. Avialable: https://www.techwire.net/hacking-critical-infrastructure-is-accelerating/

20) Paganini, P. (2014, Nov 24). Foreign Hackers Constantly Target US Critical Infrastructure. [Online]. Available: http://resources.infosecinstitute.com/foreign-hackers-constantly-target-us-critical-infrastructure/

21) Borilin, S. (2014, Sep 24). So malware attacks against critical infrastructure are inevitable. What's next?. [Online]. Available: https://business.kaspersky.com/so-malware-attacks-against-critical-infrastructure-are-inevitable-whats-next-2/2647

22) Dingman, S., Silcoff, S., & Greenspan, R. (2014, Nov 24). Hacked: The escalating arms race against cybercrime. [Online]. Available: http://www.theglobeandmail.com/report-on-business/hacked-the-escalating-arms-race-against cybercrime/article21305464/?page=all

23) Graham, A. "Canada's Critical Infrastructure: When is Safe Enough Sage Enough?" The Macdonald-Laurier Institute. 2014.

24) Miller, B., & Rowe, D. "A survey of SCADA and critical infrastructure incidents". In Proceedings of the 1st Annual Conf. on Research in Inf. Tech., ACM. 2012, pp. 51-56.

25) Auditor General of Canada. "2012 Fall Report of the Auditor General of Canada: Protecting Canadian Critical Infrastructure Against Cyber Threats". Office of the Auditor General of Canada No. FA1-2012. Ottawa, ON. 2012.

26) Public Safety Canada. "Action Plan for Critical Infrastructure (2014-2017)". Public Safety Canada No. PS4-66/2014E. Ottawa, ON. 2014.

27) Rege-Patwardhan, A. "Cybercrimes against critical infrastructures: a study of online criminal organization and techniques". Criminal Justice Studies, vol. 22, no. 3, pp. 261-271, 2009.

28) Lynn, W. J. "Remarks at the 28th Annual International Workshop on Global Security". June 16, 2011.

29) Holz, T., Engelberth, M., & Freiling, F. "Learning more about the underground economy: A case-study of keyloggers and dropzones" in Spring Berlin Heidelberg. 2009, pp. 1-18.

30) Frank, R., Macdonald, M., & Monk, B. "Identifying Potential Canadian Targets in Online Hacker Forums". Ottawa: Public Safety Canada (ongoing). 2015.

31) United States Department of Homeland Security. (2015). "What is Critical Infrastructure?". [Online]. Available: http://www.dhs.gov/what-critical-infrastructure

32) United Nations. (2011). "Cybersecurity: A global issue demanding a global approach". Available: http://www.un.org/en/development/desa/news/ecosoc/cybersecurity-demands-global-approach.html

Authorized licensed use limited to: Penn State University. Downloaded on June 17,2025 at 17:13:17 UTC from IEEE Xplore. Restrictions apply.

Appendix MACDONALD02

https://dl.acm.org/doi/10.1145/3145574.3145590 June 23, 2025

ACM DIGITAL LIBRARY | Association for Computing Machinery

Browse | About | Sign in | Register

Journals | Magazines | Proceedings | Books | SIGs | Conferences | People

Search ACM Digital Library | Advanced Search

Conference | Proceedings | Upcoming Events | Authors | Affiliations | Award Winners

Home > Browse Publications > css > ACM Other conferences > CSS 2017 > Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums

RESEARCH-ARTICLE

Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums

Authors: Mohammed Almukaynizi, Alexander Grimm, Eric Nunes, Jana Shakarian, Paulo Shakarian Authors Info & Claims

CSS 2017: Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas
Article No.: 12, Pages 1 - 7 • https://doi.org/10.1145/3145574.3145590

Published: 19 October 2017 Publication History Check for updates

23 545 Get Access

Abstract

We present an approach that combines social network analysis with machine learning techniques to predict future cyber threats through darkweb/deepweb discussions with hacking-related content. Our approach harnesses features derived from hacker social networks and from online sources of cybersecurity advisories. We address the problem of predicting the exploitability of software vulnerabilities to show that features computed from hacker social networks are important indicators of future cybersecurity incidents. We conduct a suite of experiments on real-world hacker and exploit data and demonstrate that social network data improves recall by about 19%, F1 score by about 6% while maintaining precision. We believe this is because social network structures related to certain exploit authors is indicative of their ability to write exploits that are subsequently employed in an attack.

References

[1] Luca Allodi and Fabio Massacci. 2014. Comparing vulnerability severity and exploits using case-control studies. ACM Transactions on Information and System Security (TISSEC) 17, 1 (2014), 1.
Digital Library Google Scholar

[2] Victor Benjamin, Weifeng Li, Thomas Holt, and Hsinchun Chen. 2015. Exploring threats and vulnerabilities in hacker web: Forums, IRC and carding shops. In Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on. IEEE, 85--90.
Crossref Google Scholar

[3] Alex Beutel, Leman Akoglu, and Christos Faloutsos. 2015. Fraud detection through graph-based user behavior modeling. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 1696--1697.
Digital Library Google Scholar

[4] LeoBreiman. 2001. Random Forests. Machine Learning 45, 1 (2001), 5--32.
Digital Library Google Scholar

[5] Benjamin L. Bullough, Anna K. Yanchenko, Christopher L. Smith, and Joseph R. Zipkin. 2017. Predicting exploitation of disclosed software vulnerabilities using open-source data. In Proceedings of the 2017 ACM International Workshop on Security and Privacy Analytics. ACM.
Digital Library Google Scholar

[6] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 15--15.
Digital Library Google Scholar

Appendix ALMUKAYNIZI01

[7] Michel Edkrantz and Alan Said. 2015. Predicting Cyber Vulnerability Exploits with Machine Learning. In SCAI. 48--57.
Google Scholar

[8] Thomas J Holt, Deborah Strumsky, Olga Smirnova, and Max Kilger. 2012. Examining the social networks of malware writers and hackers. International Journal of Cyber Criminology 6, 1 (2012), 891.
Google Scholar

[9] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. 2011. An Analysis of Underground Forums. In Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11). ACM, New York, NY, USA, 71--80.
Digital Library Google Scholar

[10] Kartik Nayak, Daniel Marino, Petros Efstathopoulos, and Tudor Dumitraş. 2014. Some vulnerabilities are different than others. In International Workshop on Recent Advances in Intrusion Detection. Springer, 426--446.
Crossref Google Scholar

[11] Mariam Nouh and Jason RC Nurse. 2015. Identifying Key-Players in Online Activist Groups on the Facebook Social Network. In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on. IEEE, 969--978.
Digital Library Google Scholar

[12] Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin, Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, and Paulo Shakarian. 2016. Darknet and deepnet mining for proactive cybersecurity threat intelligence. In Intelligence and Security Informatics (ISI), 2016 IEEE Conference on. IEEE, 7--12.
Digital Library Google Scholar

[13] Elizabeth Phillips, Jason RC Nurse, Michael Goldsmith, and Sadie Creese. 2015. Extracting social structure from darkweb forums. (2015).
Google Scholar

[14] John Robertson, Ahmad Diab, Ericsson Marin, Eric Nunes, Vivin Paliath, Jana Shakarian, and Paulo Shakarian. 2017. Darkweb Cyber Threat Intelligence Mining. Cambridge University Press.
Digital Library Google Scholar

[15] Carl Sabottke, Octavian Suciu, and Tudor Dumitras. 2015. Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits. In USENIX Security, Vol. 15.
Digital Library Google Scholar

[16] Sagar Samtani, Ryan Chinn, and Hsinchun Chen. 2015. Exploring hacker assets in underground forums. In Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on. IEEE, 31--36.
Crossref Google Scholar

[17] Jana Shakarian, Andrew T Gunn, and Paulo Shakarian. 2016. Exploring malicious hacker forums. In Cyber Deception. Springer, 261--284.
Google Scholar

[18] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. Social media mining: an introduction. Cambridge University Press.
Digital Library Google Scholar

Show fewer references

Cited By

View all

Kashihara K, Pal K, Baral C and Trevino R. (2024). Prompt-Based Learning for Thread Structure Prediction in Cybersecurity Forums. Intelligent Systems and Applications. 10.1007/978-3-031-47715-7_51. (766-781). Online publication date: 30-Jan-2024.

https://doi.org/10.1007/978-3-031-47715-7_51

Hughes J, Pastrana S, Hutchings A, Afroz S, Samtani S, Li W and Marin E. (2024). The Art of Cybercrime Community Research. ACM Computing Surveys. 10.1145/3639362. Online publication date: 10-Jan-2024.

https://dl.acm.org/doi/10.1145/3639362

Kühn P, Wittorf K and Reuter C. (2024). Navigating the Shadows: Manual and Semi-Automated Evaluation of the Dark Web for Cyber Threat Intelligence. IEEE Access. 10.1109/ACCESS.2024.3448247. **12**. (118903-118922). Online publication date: 2024.

https://doi.org/10.1109/ACCESS.2024.3448247

Rekha B, Rajeshwari S, Siddanna S and Kavya G. (2023). The Prediction of Attacks and Challenges in Cyber Security Using Machine Learning Techniques. Proceedings of the 2nd International Conference on Cognitive and Intelligent Computing. 10.1007/978-981-99-2742-5_45. . 433-445. Online publication date: 27-Sep-2023.

https://doi.org/10.1007/978-981-99-2742-5_45

Di Tizio G, Siu G, Hutchings A and Massacci F. (2023). A Graph-Based Stratified Sampling Methodology for the Analysis of (Underground) Forums. IEEE Transactions on Information Forensics and Security.

10.1109/TIFS.2023.3304424. **18**. 5473-5483. Online publication date: 2023.

https://doi.org/10.1109/TIFS.2023.3304424

Du Y, Huang C, Liang G, Fu Z, Li D and Ding Y. (2022). ExpSeeker: extract public exploit code information from social media. Applied Intelligence. 10.1007/s10489-022-04178-9. **53**:12. 15772-15786. Online publication date: 28-Nov-2022.

https://doi.org/10.1007/s10489-022-04178-9

Show More Cited By

Recommendations

Combining Cyber Security Intelligence to Refine Automotive Cyber Threats

Modern vehicles increasingly rely on electronics, software, and communication technologies (cyber space) to perform their driving task. Over-The-Air (OTA) connectivity further extends the cyber space by creating remote access entry…

Read More

Darkweb Cyber Threat Intelligence Mining

Read More

On Cyber Threats to Smart Digital Environments

ICSDE'18: Proceedings of the 2nd International Conference on Smart Digital Environment

Cyber threats and attacks have significantly increased in complexity and quantity throughout this past year. In this paper, the top fifteen cyber threats and trends are articulated in detail to provide awareness throughout the community and…

Read More

Comments

DL Comment Policy

Comments should be relevant to the contents of this article, (sign in required).

Got it

0 Comments

• Share

Best Newest Oldest

Nothing in this discussion yet.

Privacy Do Not Sell My Data

Download PDF

View Table of Contents

The ACM Digital Library is published by the Association for Computing Machinery. Copyright © 2024 ACM.

Appendix ALMUKAYNIZI01

Terms of Usage | Privacy Policy | Code of Ethics



Predicting Cyber Threats through Hacker Social Networks in Darkweb and Deepweb Forums

Mohammed Almukaynizi, Alexander Grimm, Eric Nunes,
Jana Shakarian, Paulo Shakarian
Arizona State University
Tempe, AZ 85281, USA
Email: {malmukay, argrimm, enunes1, jshak, shak} @asu.edu

ABSTRACT

We present an approach that combines social network analysis with machine learning techniques to predict future cyber threats through darkweb/deepweb discussions with hacking-related content. Our approach harnesses features derived from hacker social networks and from online sources of cybersecurity advisories. We address the problem of predicting the exploitability of software vulnerabilities to show that features computed from hacker social networks are important indicators of future cybersecurity incidents. We conduct a suite of experiments on real-world hacker and exploit data and demonstrate that social network data improves recall by about 19%, F1 score by about 6% while maintaining precision. We believe this is because social network structures related to certain exploit authors is indicative of their ability to write exploits that are subsequently employed in an attack.

1 INTRODUCTION

The existing methods for assessing the likelihood of software vulnerability exploitation are found to have many limitations, preventing them from being useful tools for prioritizing vulnerability remediation. They are known to assign high severity scores to many vulnerabilities that will never be exploited [1]. Recent studies show that less than 3% of the vulnerabilities reported are ever exploited in the wild [1, 15]. Therefore, cybersecurity researchers have used alternative methods leveraging machine learning techniques on features computed from social media feeds with explicit vulnerability mentions [5, 15]. In both studies, social network data has not been analyzed nor has it been evaluated as a predictor. Additionally, while other studies suggest that darkweb/deepweb (D2web) sites are among the best sources for gathering cyber threat intelligence [1, 9, 17], the value of such intelligence in predicting the likelihood of vulnerability exploitation is not quantified.

In this paper, we introduce an approach that leverages machine learning techniques on features derived from the social network of users participating in D2web forums, as well as features derived from the National Vulnerability Database (NVD)[1]. We demonstrate the viability of user connectivity metadata in predicting cyber threats. Our machine learning model achieves an F1 measure of 0.67 and generalizes well with the biased ground truth. Additionally, we propose an approach for generating a social network graph from the discussions on D2web forums, and we provide analysis on the social network structure. We specifically show that the users who discussed vulnerabilities are about 4 times more active in posting than other users.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CSS '17, CSSSA's Annual Conference on Computational Social Science, October 19–22, 2017, Santa Fe, NM, USA
© 2017 ACM. 978-1-4503-5269-7/17/10…$15.00
DOI: 10.1145/3145574.3145590

2 BACKGROUND

2.1 Vulnerability Lifecycle

In this subsection, we provide a general description of the typical vulnerability lifecycle. We note that vulnerabilities may have variations on their lifecycles depending on different factors.

Vulnerability Reporting. Researchers are encouraged to report discovered security flaws to any CVE Numbering Authorities (CNAs)[2]. Once a flaw is reported, the CNA reserves a new Common Vulnerabilities and Exposures (CVE) number if two constraints are satisfied: first, the flaw can be exploited in a way that violates the security policies; second, no CVE number has been previously assigned to this flaw. Once it is established that the flaw poses cyber threats, the software vendor is notified and allowed a period of time to release patches before any information about the vulnerability is publicly disclosed.

Vulnerability Disclosure. Once a legitimate vulnerability is validated and the time period provided to the vendor is over, the National Institute of Standards and Technology (NIST) releases information about the vulnerability to the cybersecurity community. This disclosure is published through the National Vulnerability Database (NVD) - which is known to be the comprehensive reference vulnerability database with over 92,000 disclosed vulnerabilities as of July 2017.

Vulnerability Exploitation. Ideally, after a vulnerability is disclosed by NIST, organizations immediately begin patching their systems to avoid potential security risks. However, with the ever-increasing number of disclosed vulnerabilities, many organizations

[1] https://nvd.nist.gov.
[2] For a complete list of CNAs, see https://cve.mitre.org/cve/cna.html.

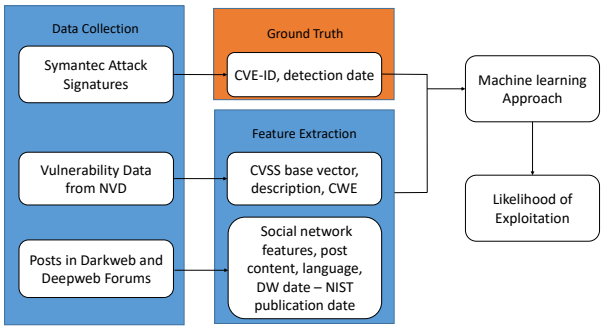


Figure 1: An overview of the predictive model.

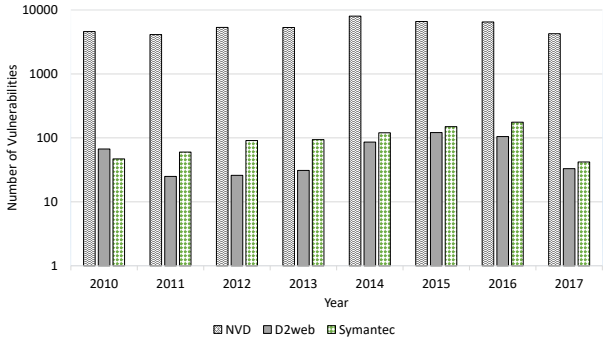


Figure 2: Vulnerabilities reported per year from NVD (per the disclosure year), D2web (per the disclosure year), and Symantec (per the exploitation year).

fall behind on vulnerability mitigation and become exposed to a wide range of newly deployed exploits that exist in the wild. Recent studies show that only a small fraction of vulnerabilities are found to be exploited in the wild (1-3% [1, 7, 10, 15]). Therefore, organizations look for methods that help to better assess the exploitability likelihood, when compared to the existing severity scoring systems.

3 APPROACH

We view the problem of predicting the vulnerability likelihood of exploitation as a binary classification problem, where the positive class is *exploited*, and the negative class is *not exploited*. Figure 1 gives an overview of our proposed approach.

3.1 Data Collection

Our machine learning models use features derived from two sources: vulnerability data feeds collected from NVD, and a D2web database of posts with cybersecurity-related content collected and filtered from 151 D2web forums [12]. The class labels are determined based on a ground truth set of attack signatures of exploits detected in the wild and reported by Symantec[3]. For the purpose of this paper, we focus our study on data collected from all the mentioned sources between January 2010- March 2017. Figure 2 shows the number of vulnerabilities published on NVD per year, the number of vulnerabilities mentioned on D2web, and the number of vulnerabilities reported by Symantec. Next, we provide details on the data collection process for the three sources we use.

NVD. Every vulnerability in NVD is assigned a unique CVE number. To collect vulnerability data, we use the JSON data feeds provided by NVD and extract information about the vulnerabilities we study (extracted information is discussed in Section 3.2). Further, a web scraper is developed to retrieve from the NVD's vulnerability webpages the data elements that are missing in the JSON files (e.g., the disclosure date).

[3]https://www.symantec.com

D2web Forums. We use a database of posts collected from 151 darknet and deepnet forums. The data collection system is described in [12]. In summary, the system extracts the hacking-related content from D2web sites using a learning algorithm with high accuracy. The database we use contains over 2,290,000 posts under 223,074 distinct forum topics and from 151 distinct forums. From these posts, only 3,082 have explicitly mentioned 624 distinct vulnerabilities by referencing their CVE-ID. Of those, 502 are within the time period we focus on, and they are found on 46 different forums. We only use these forums to build the social network of users, as described in Section 4.

Symantec Attack Signatures. We label vulnerabilities as *exploited* in the wild if the CVE-ID is mentioned in the description of the attack signatures reported by Symantec's anti-virus[4] or Intrusion Detection Systems' attack signatures[5]. The fraction of exploited vulnerabilities is found to be very small as shown in Figure 2, and it varies from year to another. The maximum value for the fraction of exploited vulnerabilities is 2.7% (for vulnerabilities published in 2016), and minimum value is less than 1% (for vulnerabilities published in 2017) - previous work has reported fractions comparable to our findings [1, 15].

3.2 Features Description

Table 1 summarizes the features we extract from both data sources discussed in Section 3.1. Here we provide discussions on each of these sets of features.

CVSS Base Score. CVSS is a vulnerability severity scoring framework designed to measure the exploitability and the impact of the vulnerabilities. We use versions v2.0 CVSS base score. There are two components for this set of features: (1) the base score (numeric): a given severity score ranges from 0 to 10, and (2) the CVSS vector: a vector of the metrics that determine the base score (categorical). The measures in the vector are *Access Vector, Access Complexity,*

[4]A complete list is found here https://www.symantec.com/security_response/landing/azlisting.jsp. The detection date is labeled with "Discovered".
[5]https://www.symantec.com/security_response/attacksignatures.

Appendix ALMUKAYNIZI01

Table 1: Summary of features

Source	Feature Set	Type
NVD	CVSS base score	Numeric and Categorical
	Description	TF-IDF uni-grams
	CWE	Categorical
D2web	Social Network Features	Numeric
	Post content	TF-IDF uni-grams
Combined	D2web date - NVD date	Numeric

Authentication, Confidentiality Impact, Integrity Impact, and *Availability Impact*. Each one of these measures can take one value from a predetermined set of possible values. For example, *Access Vector* indicates how the vulnerability is exploited. It can take one of these three possible values: *Local (L), Adjacent Network (A) and Network (N)*[6].

NVD Description. NIST provides a textual description of the vulnerability when it is released. The description summarizes the system/software in which the flaw exists and gives information on how it can be exploited. Each vulnerability description undergoes a preprocessing pipeline including stemming (reducing the words to their root forms) and stop words removal (e.g., *and, or, then*). The preprocessed documents are then turned into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF) computed for the 250 most frequent words in the collection of documents. Informally, TF-IDF assigns higher values to the words that appeared in a document with higher frequency while rarely did these words appear in other documents in the corpus.

CWE. It is a community-effort project comprising enumerating common software security weaknesses (categorical). These are categories of flaws that can be unintentionally made during the software development and can exist in the software architecture, design, or code.

D2web Social Network Features. This set of features contains measures computed from the social connections of users posting hacking-related content. The basic social network features (e.g., in-degree, out-degree) indicates how active a user is in the social graph. More advanced features measure the centrality of users in the social graph. Highly central users are more important; thus the vulnerability mentions should take more consideration. We compute the features for the set of users who explicitly mentioned one or more CVE-IDs in their posts.

Post Content. We found evidence for many vulnerability mentions with content ranging from exploit offers to content irrelevant of the mentioned vulnerability. This set of features is extracted the same way as NVD description features except that for non-English posts, we automatically translate the content to English using Google Translate API[7], then the TF-IDF are computed over the translated words.

[6]See https://www.first.org/cvss/v2/guide for complete documentation.

3.3 Classifier Training and Prediction

We use a supervised machine learning approach to train classifiers on the presented features. The output of the classifiers is a confidence score. A threshold can be set on the confidence score to determine the best decision boundary. The experimental settings and results are described in Section 6.

4 HACKER SOCIAL NETWORK

In Section 3.2, we describe the features we use, while here, we elaborate on the social network creation approach. We then provide our observations on the created graphs in Section 5. In this work, we adopt the same assumption made in much of the previous work on D2web data, where they consider the same usernames (case insensitive) across different D2web sites to belong to the same person(s) [14]. This allows for generating one network comprising a large number of D2web sites as opposed to a social network for each site [13].

4.1 Users' Social Graph

Formally, the users' social graph $G = (V, E)$ is a weighted, directed graph with no self-loops (i.e., every edge has a weight; every edge points away from one node to another node; there exists at most one edge between any pair of nodes). V is the set of vertices (D2web users) and E is the set of edges.

The graph is creating using posts between January 2010 to June 2016 (747,351 posts under 109,413 forum topics), and for every topic t_x, posts under t_x are grouped together in a list l_x ordered by the date and time of posts. Then, an edge is created (with weight = 1) from user v_i to v_j and labeled with the date of v_i's posting date only if: (1) $v_i \neq v_j$, (2) both v_i and v_j have posts in l_x, and v_i has posted after v_j, (3) the number of posts between v_i's post and v_j's post in l_x is less than thr (it is set to be 10 in all experiments in this paper), and (4) there is no existing edge originating from v_i to v_j and labeled with the same date. Once the edges are created, they are added to a multi-directed graph with parallel edges of weights = 1. The multi-graph is then transformed to a directed graph G by summing the weights of the parallel edges pointing at the same direction.

Degree distributions, for both incoming and outgoing edges, of G are found to resemble the power-law distribution as depicted in Figure 3. This means that there exist very few users with a very large number of connections, and many users with few connections - this observation is known to be common for social media sites [18][8].

[7]https://cloud.google.com/translate.

[8]In the D2web database, there are few users, each has exactly one post. Some of these posts are found in topics with a number of postings greater than thr; hence the little up-tick in Figure 3.

Appendix ALMUKAYNIZI01

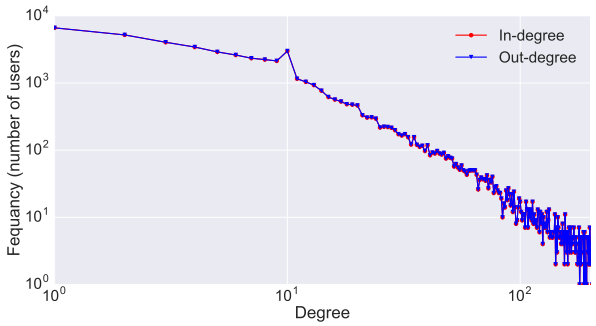


Figure 3: In-degree and out-degree distribution of users in G.

4.2 Social Network Measures

After creating the social network, we compute measures derived from the network structure. In this paper, we consider three categories of social network measures:

Network Structure Measures: the measures under this category are: (1) *In-degree*: the number of edges pointing to the user, (2) *Out-degree*: the number of edges originated from the user, (3) *Sum of In-degree weights*: the sum of the weights for all edges pointing to the user, (4) *Sum of out-degree weights*: the sum of the weights for all edges pointing away from the user. These measures describe the type of activities in which the user engages. For example, higher in-degree than out-degree may indicate the user tendency towards creating new topics or posting under topics short time after they are created.

Centrality Measures: three measures are computed: (1) *In-degree centrality*: it measures the popularity of a user v_i by normalizing v_i's in-degree by the maximum possible in-degree, (2) *Out-degree centrality*: measures how actively a user v_i replies to others by normalizing v_i's out-degree measure by the maximum possible out-degree, (3) *Betweenness centrality*: for a user v_i, Betweeness centrality measures the importance of v_i by computing the fraction of shortest paths between all pairs of users that pass through v_i.

Importance Measures: the number of connections user v_i has with other users, by itself, may not be indicative of importance; rather, v_i is important if his/her posts make other important users reply. Hence, influence metrics incorporate the centrality of users with outgoing edges to v_i into v_i's centrality (i.e., if an important user v_j replies to v_i, then the importance of v_i increases). Two measures are computed under this category: (1) *Eigenvector centrality*: measures the importance of v_i by assigning a centrality proportional to the sum of in-neighbors' centralities. Eigenvector centrality of v_i is the ith value of the eigenvector C_e corresponding to the largest eigenvalue of the network adjacency matrix A^t, and (2) *Pagerank centrality*: measures the centrality of v_i by incorporating fractions of the centralities of in-neighbors, such that each of v_i's in-neighbors passes the value of his/her centrality divided by the number of outgoing edges[9].

[9]In all the experiments, the damping factor is set to 0.85.

5 SOCIAL NETWORK ANALYSIS

In this section, we report our observations on the computed measures. Table 2 shows statistics for the D2web social graph G created according to our description, as well as statistics for (1) the subset of users who have discussed vulnerabilities in D2web (s.t., $vulUsers \subset V$), (2) the subgraph of G induced by the $vulUsers$, $G_{vulns}(vulUsers, vulEdges)$, (s.t., $vulEdges \subset E$), and (3), a subgraph induced by $vulUsers$ as well as all their in-neighbors and out-neighbors $G_{vulNei}(vulNeis, vulNeiEdges)$, (s.t., $vulNeis \subset V$ and $vulNeiEdges \subset E$). We create these three graphs to understand the differences in the user connectivity for the subset of users mentioned vulnerabilities (s.t., $vulUsers$, less than 1% of the total population) as opposed to the total user population.

These individuals, or *vulUsers*, are spread across many D2web forums (46 forums out of 151). Further, they are generally more active than other users in the same forums. As depicted in Table 2, the average in-degree (64.90) and out-degree (51.45) for the subset of *vulUsers* are orders of magnitude higher than the same measures for all users in the graph (13.74)– about 5 times higher in-degree and 4 times higher out-degree. This shows that the average hacker in *vulUsers* is exposed to a larger population of hackers than a normal hacker. Expectedly, we observe that a hacker from *vulUsers* is more likely to engage in discussions with another hacker from the same group than he/she does with others. For example, the average in-degree and out-degree (4.08) for the subgraph G_{vulns} indicates that, on average, a user has connections with about 4 other *vulUsers* (about 1% of the population of *vulUsers*); whereas he/she would reply to about 51 users and make about 64 other users reply to his/her post, less than 0.1% of the total population. For these reasons, *vulUsers* generally, exhibit significantly higher centrality and importance measures as compared to normal users in G. However, we observe that the distribution of network measures vary largely within this group of users. We also observe that about 30% of the $vulUsers$ have no communication history with other users within the same group. Figure 4 shows a visualization for the subgraph G_{vulns}, which confirms the two observations. Finally, about 25% of $vulUsers$ joined the D2web community less than three days before their first vulnerability mention.

6 EXPERIMENTAL SETUP

We perform experiments on the set of vulnerabilities mentioned on D2web forums in the time period from January 2010 to March 2017. We exclude the vulnerabilities that were mentioned by users with no communication history. We also exclude the vulnerabilities that had been detected in the wild by Symantec before they were mentioned in any of the D2web posts, since those vulnerabilities could be retrieved by querying the database without prediction. Our resultant dataset contains 157 distinct vulnerabilities, 24 of which have the class label *exploited*.

Different machine learning classifiers are compared[10], but we only report the performance achieved by the best performing classifier, which is Random Forest (RF) [4]. RF is an ensemble of decision

[10]Classifiers used include: Support Vector Machine (SVM), Logistic Regression, and Naive Bayes.

Appendix ALMUKAYNIZI01

Table 2: Statistics for the graph G with all users with at least one edge (an in-edge or an out-edge), the subset of users that have discussed vulnerabilities *vulUsers*, subgraph ($G_{vulns}(vulUsers, vulEdges)$), and subgraph ($G_{vulNei}(vulNeis, vulNeiEdges)$).

Property	D2web users	Users mentioning vulnerabilities		
		as a subset of nodes	as a subgraph	as a subgraph with all 1-hop neighborers
Nodes	53, 178	365	365	10469
Edges	730, 740	undefined	1, 492	202, 070
The average of:				
In-degree	13.74	64.90	4.08	19.30
Out-degree	13.74	51.45	4.08	19.30
Sum of in-degree weights	35.80	250	48.07	75.73
Sum of out-degree weights	35.80	215	48.07	75.73
In-degree centrality	$2.59e^{-4}$	$1.22e^{-3}$	$1.12e^{-2}$	$1.84e^{-3}$
Out-degree centrality	$2.59e^{-4}$	$9.68e^{-4}$	$1.12e^{-2}$	$1.84e^{-3}$
Betweeness centrality	$7.2e^{-5}$	$1.29e^{-3}$	$1.89e^{-4}$	$2.35e^{-4}$
Eigenvector Centrality	$2.18e^{-4}$	$4.68e^{-4}$	$5.95e^{-3}$	$8.21e^{-4}$
Pagerank	$1.90e^{-5}$	$1.27e^{-4}$	$2.74e^{-3}$	$9.6e^{-5}$

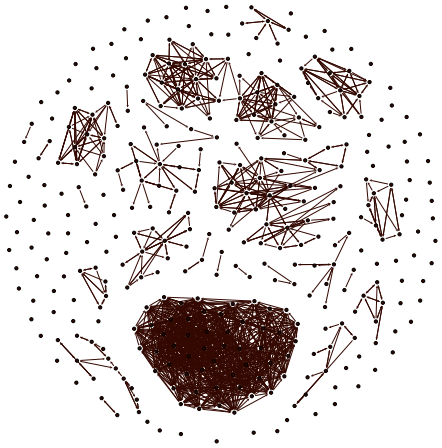


Figure 4: The subgraph of G induced by the set of users who have mentioned vulnerabilities in their postings.

trees generated from the training data. Each tree produces a prediction independent of other trees, with the algorithm taking the majority vote at the end. In each of the experiments, we report an averaged performance of five runs.

6.1 Performance Evaluation

We evaluate the performance of the predictors using precision, recall, and the harmonic mean of precision and recall, which is F1 measure. Precision is the fraction of vulnerabilities that were actually exploited from all vulnerabilities predicted as being exploited, and recall is the fraction of correctly predicted exploited vulnerabilities from the total number of exploited vulnerabilities. Table 3 defines how each metric is computed.

Table 3: Evaluation metrics. TP - true positives, FP - false positives, FN - false negatives, TN - true negatives.

Metric	Formula
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1	$2 * \frac{precision*recall}{precision+recall}$

We also compute Receiver Operating Characteristics (ROC) for each run and report Area Under Curve (AUC) of the classifier. ROC graphically illustrates the classification performance by plotting the true positive rate (TPR) against the false positive rate (FPR) at different points of the decision boundary.

6.2 Results

Experiments under Real-World Conditions. In this set of experiments, we sort the vulnerabilities by their D2web date, then we train our classifiers on the vulnerabilities mentioned before June 2016 (125 vulnerabilities), and test on the vulnerabilities from June 2016 to March 2017 (32 vulnerabilities, only 3 are *exploited*). The classification performance achieved by our RF model has an average precision of 0.57, recall of 0.93, and F1 of 0.67. The same classifier is able to achieve on average AUC of 0.95[11]. The lower score of precision is attributed to the fact that Symantec's data is biased towards reporting exploits targeting vulnerabilities that exist in software products from certain software vendors such as Microsoft, Adobe [15]. Since our model is found to predict vulnerabilities as being exploited from other vendors as well, we believe that some

[11]The results of 5 runs show relatively high variance due to the small number of samples on which the models are tested.

Appendix ALMUKAYNIZI01

false positives were actually exploited in the wild but never detected by Symantec.

Ablation Test and Cross-Validation. Since the number of vulnerabilities in our testing dataset in the previous experiment is relatively small, we further apply stratified 5-fold cross-validation on the whole dataset. In this experiment, the samples are intermixed; hence these conditions do not reflect the conditions of real-world streaming prediction (i.e., predicting the likelihood of exploitation at the time of the vulnerability mention). The average F1 achieved is 0.72, with a precision of 0.61, a recall of 0.89, and an AUC of 0.88[12].

To measure the impact of individual feature sets on the overall classification performance, we apply two tests: (1) an ablation test (depicted in Figure 5) where the change in precision, recall, F1, and AUC is recorded when each set of features is removed from the prediction model, and (2) a test on individual feature sets (depicted in Figure 6) where the classification performance is reported for models trained on only one set features at a time. In the ablation test, when the set of social network features is removed from the model, some decrease in performance was recorded as depicted in Figure 5. In the individual feature tests, the social network measures resulted in an improvement in performance that is significantly higher than the improvement recorded from the inclusion of other feature sets. We note that the simple classifier - which labels all vulnerabilities as being *exploited*, results in a precision of 0.16, a recall of 1, at an F1 of 0.27 and an AUC of 0.5.

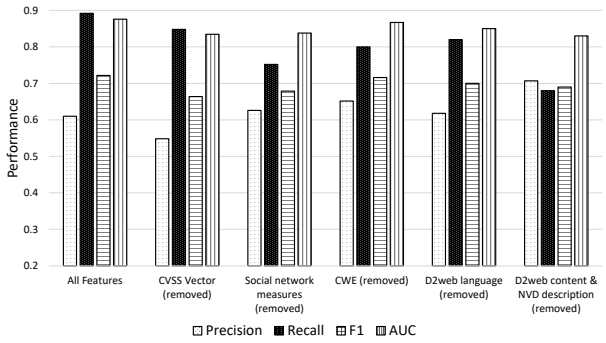


Figure 5: Classification performance achieved by applying ablation test with 5-fold cross-validation.

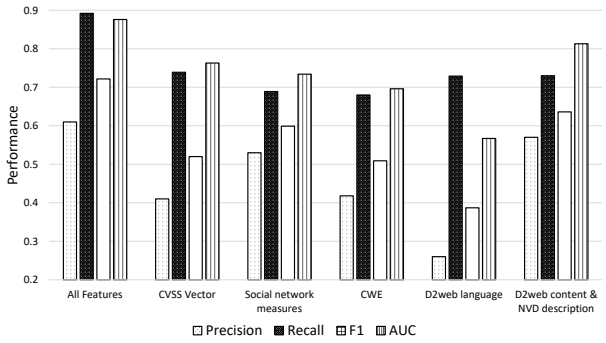


Figure 6: Classification performance achieved by individual feature sets.

7 DISCUSSION

In the ablation test, the largest drop in F1 occurred when CVSS vector set of features were removed, followed by the removal of social network measures with comparable drop in F1. It is important to note that CVSS vector is designed to assess the vulnerability exploitability and impact, and the assignment of values to the different components of the vector is manually done by domain experts[13]. Hence, being able to achieve performance comparable to the expensive CVSS vector when social network features are used is very promising. However, we do note that even the experiments here where we only use the CVSS vector differ substantially from previous work. As we have selected a-priori on vulnerabilities that have appeared on D2web, we observe a superior performance of CVSS score in this paper when compared to previous work.

Additionally, when the individual feature sets were examined, the best performing features were the TF-IDFs computed from both the content of D2web postings and the vulnerability description retrieved from NVD. The second best performing features were the social network measures, scoring F1 that is significantly higher than F1 scores achieved by the other individual feature sets - excluding the textual features. The D2web textual content provides rich information about the context in which the vulnerability is discussed. Furthermore, the software vendor (e.g., Microsoft, Adobe) can be easily derived from the NVD description; leading the model to potentially over-fit the biased ground truth. In all experiments, social network measures demonstrated their viability as predictors of potential cyber threats.

[12]Standard deviation for the evaluation metrics across the five runs is between 0.02 (F1 and AUC) and 0.06 (recall).
[13]See https://www.first.org/cvss/v2/guide.

8 RELATED WORK

8.1 Vulnerability Exploitation Prediction

A recent study found that CVSS base score metrics are poor indicators of exploitation [1]. Many vulnerabilities are assigned high scores, resulting in very high false positive rates. Sabottle et al. [15] proposed an SVM classifier that leverages data feeds from Twitter with explicit mentions to legitimate CVE identifiers to predict whether a vulnerability will have proof-of-concepts available in one experimental setup, and predict whether a vulnerability will be detected in the wild in other experimental setup. Results in their work are reported on time-intermixed samples (i.e., samples in testing set may have appeared before samples in training set), and including samples where the exploitation date is before any of the tweets are posted. Both practices have been discussed in [5]

Appendix ALMUKAYNIZI01

to measure their impact on real-world proactive exploitation prediction settings. The work in [5] replicates the experiments done in [15] (with data feeds spanning different time period) to predict the existence of proof-of-concepts from EDB. They found that the experimental methodology highly influence the results. In both papers, the dynamics of user connectivity are not studied.

8.2 Social Network Analysis

An extensive amount of work has focused on usage of measures computed from a social network of actors to identify malicious actions [3, 6, 11]. For example, Cao et al. [6] proposed a method called *SybilRank* that relies on social network measures to identify fake accounts (Sybils). Across multi-disciplines, hacker communities in underground hacking forums have been widely studied to understand the dissemination of information among hackers, the motives for hacking, and the reputation and skill level of hackers to detect threats [2, 8, 16]. However, the dynamics of connections within hacker communities have not been quantified as predictors for vulnerability exploitation. We find the measures computed from social connections to be promising predictors of future cyber attacks.

9 CONCLUSION

Our work contributes toward the understanding of user connectivities in D2web forums, and extracting measures that serve as predictors of cyber attacks. We address the problem of predicting the likelihood of vulnerability exploitation to demonstrate the value of these measures. Our experimental results also suggest that these measures produce models that generalize well when the ground truth is biased.

ACKNOWLEDGMENT

Some of the authors were supported by the Office of Naval Research (ONR) contract N00014-15-1-2742, the Office of Naval Research (ONR) Neptune program and the ASU Global Security Initiative (GSI). Paulo Shakarian and Jana Shakarian are supported by the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL) contract number FA8750-16-C-0112. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government.

REFERENCES

[1] Luca Allodi and Fabio Massacci. 2014. Comparing vulnerability severity and exploits using case-control studies. *ACM Transactions on Information and System Security (TISSEC)* 17, 1 (2014), 1.
[2] Victor Benjamin, Weifeng Li, Thomas Holt, and Hsinchun Chen. 2015. Exploring threats and vulnerabilities in hacker web: Forums, IRC and carding shops. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*. IEEE, 85–90.
[3] Alex Beutel, Leman Akoglu, and Christos Faloutsos. 2015. Fraud detection through graph-based user behavior modeling. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1696–1697.
[4] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32.
[5] Benjamin L. Bullough, Anna K. Yanchenko, Christopher L. Smith, and Joseph R. Zipkin. 2017. Predicting exploitation of disclosed software vulnerabilities using open-source data.. In *Proceedings of the 2017 ACM International Workshop on Security and Privacy Analytics*. ACM.
[6] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 15–15.
[7] Michel Edkrantz and Alan Said. 2015. Predicting Cyber Vulnerability Exploits with Machine Learning.. In *SCAI*. 48–57.
[8] Thomas J Holt, Deborah Strumsky, Olga Smirnova, and Max Kilger. 2012. Examining the social networks of malware writers and hackers. *International Journal of Cyber Criminology* 6, 1 (2012), 891.
[9] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. 2011. An Analysis of Underground Forums. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*. ACM, New York, NY, USA, 71–80. DOI:https://doi.org/10.1145/2068816.2068824
[10] Kartik Nayak, Daniel Marino, Petros Efstathopoulos, and Tudor Dumitraş. 2014. Some vulnerabilities are different than others. In *International Workshop on Recent Advances in Intrusion Detection*. Springer, 426–446.
[11] Mariam Nouh and Jason RC Nurse. 2015. Identifying Key-Players in Online Activist Groups on the Facebook Social Network. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 969–978.
[12] Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin, Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, and Paulo Shakarian. 2016. Darknet and deepnet mining for proactive cybersecurity threat intelligence. In *Intelligence and Security Informatics (ISI), 2016 IEEE Conference on*. IEEE, 7–12.
[13] Elizabeth Phillips, Jason RC Nurse, Michael Goldsmith, and Sadie Creese. 2015. Extracting social structure from darkweb forums. (2015).
[14] John Robertson, Ahmad Diab, Ericsson Marin, Eric Nunes, Vivin Paliath, Jana Shakarian, and Paulo Shakarian. 2017. *Darkweb Cyber Threat Intelligence Mining*. Cambridge University Press.
[15] Carl Sabottke, Octavian Suciu, and Tudor Dumitras. 2015. Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits.. In *USENIX Security*, Vol. 15.
[16] Sagar Samtani, Ryan Chinn, and Hsinchun Chen. 2015. Exploring hacker assets in underground forums. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*. IEEE, 31–36.
[17] Jana Shakarian, Andrew T Gunn, and Paulo Shakarian. 2016. Exploring malicious hacker forums. In *Cyber Deception*. Springer, 261–284.
[18] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. *Social media mining: an introduction*. Cambridge University Press.

Appendix ALMUKAYNIZI01

We keep the record straight by preserving government websites, news articles, historical data, and more. If you find our archive useful, please CHIP IN

INTERNET ARCHIVE WEB TEXTS VIDEO AUDIO SOFTWARE IMAGES SIGN UP | LOG IN UPLOAD Search

ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE

DONATE INTERNET ARCHIVE WayBackMachine Explore more than 946 billion web pages saved over time

https://researchrepository.wvu.edu/etd/11253/

Calendar · Collections · Changes · Summary · Site Map · URLs

Saved **15 times** between October 16, 2022 and May 15, 2024.



Wed, 26 Oct 2022 03:04:22 GMT (why: OAI-PMH-CRAWL-2022-10, customcrawlservices, ia_pub_crawls)

JAN

						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

FEB

		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28					

MAR

		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

APR

					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

MAY

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

JUN

			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

JUL

					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

AUG

	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

SEP

			1	2	3	
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

OCT

						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

NOV

		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

DEC

				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Note

This calendar view maps the number of times **https://researchrepository.wvu.edu/etd/11253/** was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the FAQ. Green indicates redirects (3xx).

FAQ | Contact Us | Terms of Service (Dec 31, 2014)

The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.

Your use of the Wayback Machine is subject to the Internet Archive's Terms of Use.

https://web.archive.org/web/20221026030422/https://researchrepository.wvu.edu/etd/11253/ June 23, 2025

INTERNET ARCHIVE WayBack Machine https://researchrepository.wvu.edu/etd/11253/ Go SEP OCT APR
15 captures 16 Oct 2022 - 15 May 2024 26 2022 2024 About this capture

Browse

Collections
Disciplines
Authors

Resources

WVU Libraries
WVU Research Office
WVU Research Commons
Open Access @ WVU
Digital Publishing Institute

Search

Enter search terms:
Search
in this series
Advanced Search
Notify me via email or RSS

Author Corner

Author FAQ
Submit Research

Home > Graduate Theses, Dissertations, and Problem Reports > 11253

< Previous Next >

GRADUATE THESES, DISSERTATIONS, AND PROBLEM REPORTS

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

Vivian Fiona Guetler, *West Virginia University* Follow

Semester
Spring

Date of Graduation
2022

Document Type
Dissertation

Degree Type
PhD

College
Eberly College of Arts and Sciences

Department
Sociology and Anthropology

Committee Chair
Jesse Wozniak

Committee Co-Chair
Katie Corcoran

Committee Member
Katie Corcoran

Committee Member
Abhik Roy

Committee Member
Joshua Woods

Abstract
This three-article dissertation focuses on cyber-related topics on terrorist groups, specifically Jihadists' use of technology, the application of natural language processing, and social networks in analyzing text data derived from terrorists' Dark Web forums. The first article explores cybercrime and cyberterrorism. As technology progresses, it facilitates new forms of behavior, including tech-related crimes known as cybercrime and cyberterrorism. In this article, I provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology by reviewing existing literature focusing on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers.

The second article is a methodological study examining the application of natural language processing computational techniques, specifically latent Dirichlet allocation (LDA) topic models and topic network analysis of text data. I demonstrate the potential of topic models by inductively analyzing large-scale textual data of Jihadist groups and supporters from three Dark Web forums to uncover underlying topics. The Dark Web forums are dedicated to Islam and the Islamic world discussions. Some members of these forums sympathize with and support terrorist organizations. Results indicate that topic modeling can be applied to analyze text data automatically; the most prevalent topic in all forums was religion. Forum members also discussed terrorism and terrorist attacks, supporting the Mujahideen fighters. A few of the discussions were related to relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards. LDA topic modeling is significant for finding topics from larger corpora such as the Dark Web forums. Implications for counterterrorism include the use of topic modeling in real-time classification and removal of online terrorist content and the monitoring of [illegible] terrorist groups use [illegible] to justify their goals and

Download

INCLUDED IN
Computer Sciences Commons, Criminology Commons, Data Science Commons, Political Science Commons, Statistical Methodology Commons, Technology and Innovation Commons

SHARE

recruit in such forums for supporters.

The third article builds on the second article, exploring the network structures of terrorist groups on the Dark Web forums. The two Dark Web forums' interaction networks were created, and network properties were measured using social network analysis. A member is considered connected and interacting with other forum members when they post in the same threads forming an interaction network. Results reveal that the network structure is decentralized, sparse, and divided based on topics (religion, terrorism, current events, and relationships) and the members' interests in participating in the threads. As participation in forums is an active process, users tend to select platforms most compatible with their views, forming a subgroup or community. However, some members are essential and influential in the information and resources flow within the networks. The key members frequently posted about religion, terrorism, and relationships in multiple threads. Identifying key members is significant for counterterrorism, as mapping network structures and key users are essential for removing and destabilizing terrorist networks. Taken together, this dissertation applies a computational social science approach to the analysis of cyberterrorism and the use of Dark Web forums by jihadists.

Recommended Citation

Guetler, Vivian Fiona, "Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11253.
https://researchrepository.wvu.edu/etd/11253

DOI

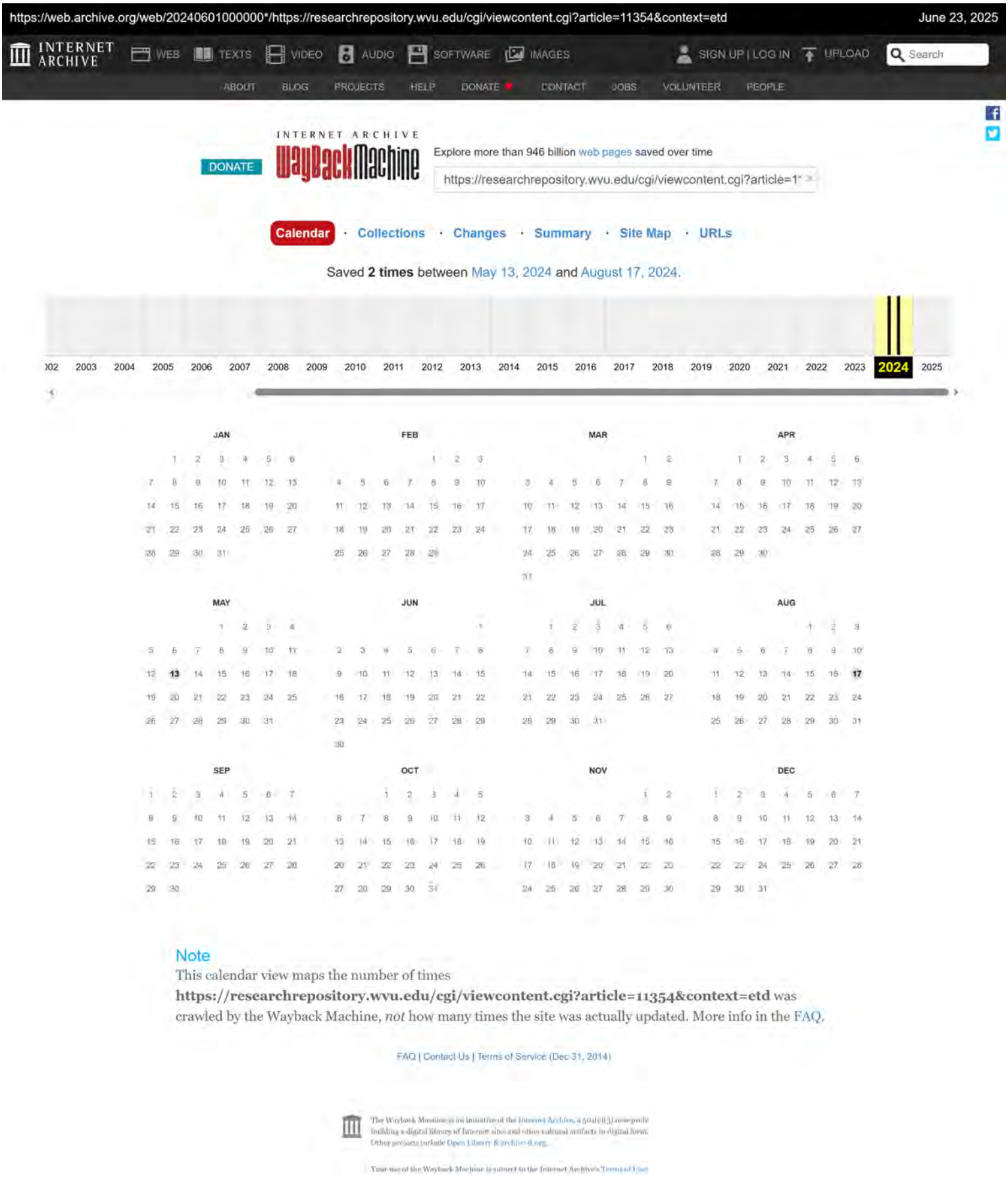
https://doi.org/10.33915/etd.11253

DIGITALCOMMONS* powered by bepress

Home | About | FAQ | My Account | Accessibility Statement

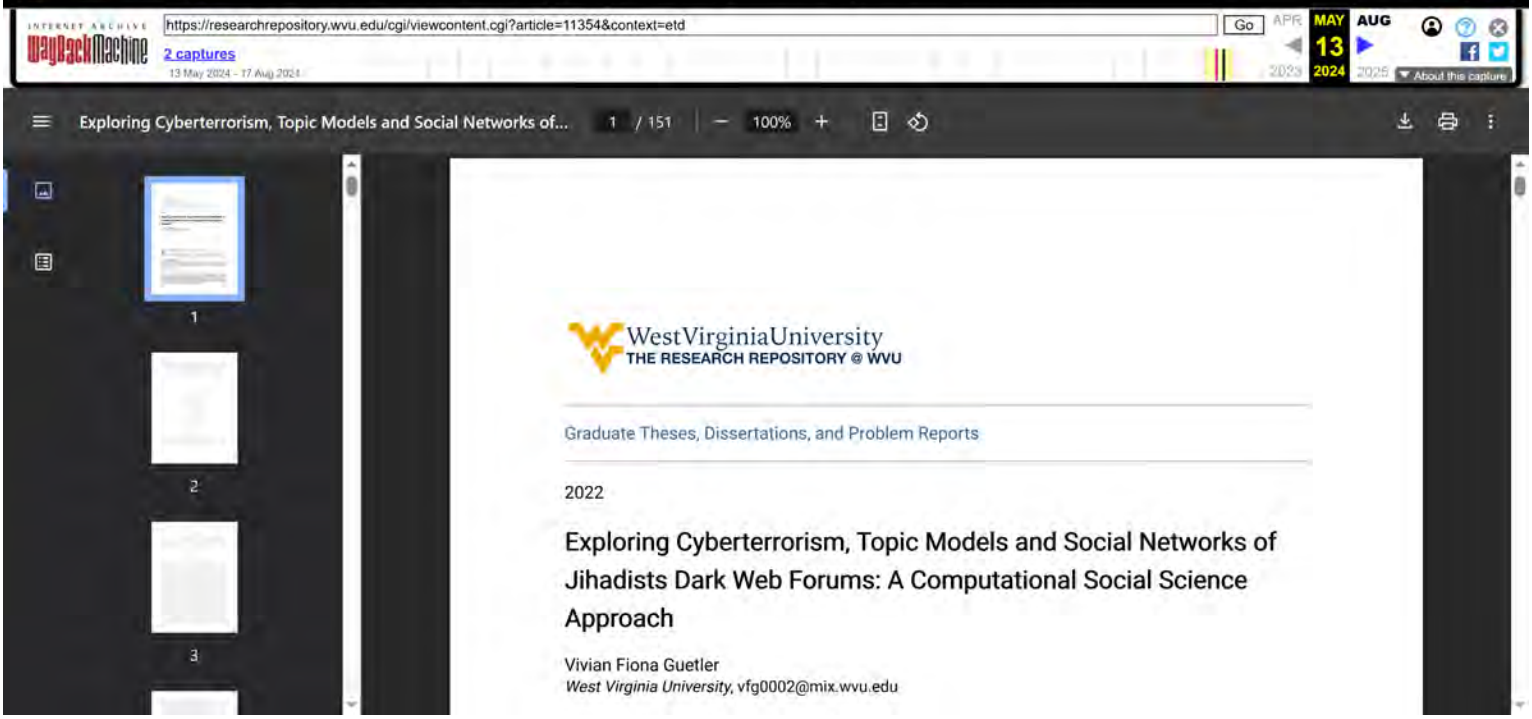
Privacy Copyright

https://web.archive.org/web/20240601000000*/https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=11354&context=etd June 23, 2025



Appendix GUETLER01

https://web.archive.org/web/20240513213612/https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=11354&context=etd June 23, 2025



WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2022

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

Vivian Fiona Guetler
West Virginia University, vfg0002@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

 Part of the Computer Sciences Commons, Criminology Commons, Data Science Commons, Political Science Commons, Statistical Methodology Commons, and the Technology and Innovation Commons

Recommended Citation

Guetler, Vivian Fiona, "Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11253.
https://researchrepository.wvu.edu/etd/11253

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Appendix GUETLER01

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

by

Vivian Fiona Guetler

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Sociology

Jesse Wozniak, Ph.D., Chair
Katie Corcoran, Ph.D.
Abhik Roy, Ph.D.
Joshua Woods, Ph.D

Department of Sociology & Anthropology

Morgantown, West Virginia
2022

Keywords: cyberterrorism, terrorism, Internet, Dark Web, machine learning, natural language processing, topics models, sentiment analysis, social networks, community detection

Copyright © 2022 Vivian Guetler

Appendix GUETLER01

Abstract

Exploring Cyberterrorism, Topic Models and Social Networks of Jihadists Dark Web Forums: A Computational Social Science Approach

by Vivian Fiona Guetler

This three-article dissertation focuses on cyber-related topics on terrorist groups, specifically Jihadists' use of technology, the application of natural language processing, and social networks in the analysis of text data derived from terrorists' Dark Web forums. The first article explores cybercrime and cyberterrorism. As technology progresses, it facilitates new forms of behavior, including tech-related crimes known as cybercrime and cyberterrorism. In this article, I provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology by reviewing existing literature focusing on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers.

The second article is a methodological study examining the application of natural language processing computational techniques, specifically latent Dirichlet allocation (LDA) topic models and topic network analysis of text data. I demonstrate the potential of topic models by inductively analyzing large-scale textual data of Jihadist groups and supporters from three Dark Web forums to uncover underlying topics. The Dark Web forums are dedicated to Islam and the Islamic world discussions. Some members of these forums sympathize with and support terrorist organizations. Results indicate that topic modeling can be applied to analyze text data automatically; the most prevalent topic in all forums was religion. Forum members also discussed terrorism and terrorist attacks, supporting the Mujahideen fighters. A few of the discussions were related to relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards. LDA topic modeling is significant for finding topics from larger corpora such as the Dark Web forums. Implications for counterterrorism include the use of topic modeling in real-time classification and removal of online terrorist content and the monitoring of religious forums, as terrorist groups use religion to justify their goals and recruit in such forums for supporters.

The third article builds on the second article, exploring the network structures of terrorist groups on the Dark Web forums. The two Dark Web forums' interaction networks were created, and network properties were measured using social network analysis. A member is considered connected and interacting with other forum members when they post in the same threads forming an interaction network. Results reveal that the network structure is

decentralized, sparse, and divided based on topics (religion, terrorism, current events, and relationships) and the members' interests in participating in the threads. As participation in forums is an active process, users tend to select platforms most compatible with their views, forming a subgroup or community. However, some members are essential and influential in the information and resources flow within the networks. The key members frequently posted about religion, terrorism, and relationships in multiple threads. Identifying key members is significant for counterterrorism, as mapping network structures and key users are essential for removing and destabilizing terrorist networks. Taken together, this dissertation applies a computational social science approach to the analysis of cyberterrorism and the use of Dark Web forums by jihadists.

This dissertation is dedicated to my dear mama, Rose Atieno, and my entire family: my grandparents, aunts and uncles, siblings, cousins and nephews, nieces, and extended family.

"It takes a village to raise and educate a child"

– Author Unknown

"I am because we are, and since we are, therefore I am"

– J.S Mbiti

Appendix GUETLER01

Acknowledgments

I would like to thank my dissertation advisor, Dr. Jesse Wozniak, for the support, encouragement, guidance, advice, and patience. You have always supported my ideas, and I am grateful for that. I inspire to be the same advisor to my future students.

I am grateful for my dissertation committee members Dr. Abhik Roy, Dr. Katie Corcoran, and Dr. Joshua Woods. Thank you for the advice, feedback, help with data analysis, and the countless recommendation letters you wrote on my behalf.

I am thankful for the entire WVU sociology and anthropology department faculty and staff, cohorts, and friends. Without your support, encouragement, and feedback, I would not have been where I am today.

I am thankful to the WVU Research Computing for the access of supercomputing systems (Thorny Flat) and their help in navigating the system.

I am thankful to my family for their prayers, support, and encouragement.

Lastly, I am thankful for Almighty God for making this possible.

Asanteni Sana, Thank You!

Appendix GUETLER01

Contents

Abstract **ii**

Acknowledgments **v**

List of Figures **ix**

List of Tables **xi**

1 Introduction **1**

1.1 Background 1

1.2 Research Objectives 8

1.3 Research Organization 8

2 Technology, Terrorism, and Crime: Exploring the Intersections of Cybercrime and Cyberterrorism **10**

2.1 Introduction 10

2.2 Understanding terrorism, cybercrime, and cyberterrorism: Definitions 13

2.3 Cybercrime and Cybercriminals 18

2.4 The cybercrime-cyberterrorism nexus: Relationship between cybercrime and cyberterrorism 21

2.5 Conclusions: Future Directions for Cybercrime and Cyberterrorism Research 26

Appendix GUETLER01

3 Text as Data: Topic Modeling of Dark Web Forums **31**

3.1 Introduction . . . 31

3.2 Machine Learning and Natural Language Processing . . . 33

3.2.1 Topic Modeling . . . 35

3.2.2 Sentiment Analysis . . . 37

3.3 Applications . . . 38

3.3.1 Detecting Terror on the Dark Web: A Case Study . . . 39

3.3.2 Dark Web Forum Data . . . 43

3.4 Text Analysis . . . 44

3.4.1 Data Pre-processing . . . 45

3.4.2 LDA Models . . . 46

3.5 Results . . . 47

3.6 Discussion and Conclusion . . . 61

4 Detecting Virtual Communities and Networks of Jihadists Dark Web Forums: A Social Network Analysis **67**

4.1 Introduction . . . 67

4.2 Dark Web's Technological Structure . . . 69

4.3 Related Work . . . 71

4.3.1 Virtual Communities and Terrorists Online Content . . . 71

4.3.2 Terror on the Dark Web . . . 74

4.3.3 Covert Social Networks . . . 76

4.4 Research Design . . . 80

4.4.1 Data . . . 81

4.4.2 Analysis . . . 81

4.5 Results . . . 85

4.6 Discussion and Conclusion . . . 92

Appendix GUETLER01

5 Conclusions **97**

5.1 Summary 97

5.2 Limitations and Future Research 102

References **106**

Appendix **137**

Appendix GUETLER01

List of Figures

3.1 Text Analysis Workflow 45

3.2 Ansar1, Gawaher and Islamic Network Top Frequent Terms 48

3.3 Ansar1 Dark Web Forum Topic Network. 80 Nodes represent topics, 3160 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted & undirected network. 49

3.4 Gawaher Dark Web Forum Topic Network. Topic network has 67 Nodes representing topics, 2211 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network. 52

3.5 Islamic Network Dark Web Forum Topic Network. Topic network with 72 Nodes representing topics, 2556 Edges representing the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network. 55

3.6 Word Sentiments of Dark Web Forums 58

3.7 Polarity Scores of Dark Web Forums 59

3.8 Dark Web Forums Top 5 Words and Emotions 60

Appendix GUETLER01

4.1 Islamic Web-Community (Myiwc) undirected network with eleven communities. 638 nodes, 4438 edges, node size represents the node degree, colors represent communities. 88
4.2 Ansar1 undirected network. 353 nodes, 8765 edges with 4 communities. Node size represents the node degree, colors represent communities. 90
4.3 Frequency of the term terror posted by key members in the network 91

Appendix GUETLER01

List of Tables

3.1 Dark Web Forums Data 44

4.1 Dark Web Forums Data 81

4.2 Interaction Network Statistics 86

5.1 Top 10 Nodes Centrality Scores – Myiwc Interaction Network 137

5.2 Top 10 Nodes Centrality Scores – Ansar1 Interaction Network 138

Appendix GUETLER01

Acronyms

ML	Machine Learning
NLP	Natural Language Processing
LDA	Latent Dirichlet Allocation
SA	Sentiment Analysis
SNA	Social Network Analysis
ANTMN	Analysis of Topic Model Networks

Appendix GUETLER01

Chapter 1

Introduction

1.1 Background

The digital revolution and development of new technologies, including the Internet, Web 2.0, artificial intelligence, and social media networks, have rapidly transformed all aspects of social, political, economic, communications, and cultural life. As of January 2022, more than 50% of the world's population is online (Forum, 2020; Kemp, 2019). Criminals and terrorists have embraced the development of the digital landscape to further their goals. Terrorists are attacking computer systems and networks, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt, 2012; Holt et al., 2017a). While not attacking cyberspace, terrorists predominantly use the Internet to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, and secure communication (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Gill et al., 2017; Sageman, 2011). Hence, facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence. The availability of large-scale information generated by terrorists and their supporters online can be collected and analyzed. However, the traditional man-

Appendix GUETLER01

ual methods for collecting and detecting terrorist content can be challenging, costly, and time-consuming (Brynielsson et al., 2013; Scrivens et al., 2021; Scrivens et al., 2018). This dissertation argues that novel computational methods such as machine learning and natural processing language techniques can advance criminological research on terrorism online content, cybercrime, and cyberterrorism.

Research into terrorists and technology use has gained pace in recent years, driven by the growth of domestic and international terrorist groups. S. Furnell et al. (2015) define cybercrime as cyber-dependent and cyber-enabled. Cyber-dependent crimes are offenses that can only be conducted using a computer, computer networks, or technologies such as spreading viruses, malware, hacking, and distributed denial of service (DDoS) attacks directed against computers or networks. In contrast, cyber-enabled crimes are traditional crimes that use computers or computer networks to widen their reach or scale. Cyber-enabled crimes include fraud, theft, and sexual offending. Hacking is the starting point from which cyber-dependent crimes begin by deliberately accessing an unauthorized computer system, networks, and data (S. Furnell, 2003; S. Furnell et al., 2015; Maimon & Louderback, 2019; Schell & Dodge, 2002). Online offenses include hacking, website defacement, Distributed Denial of Service (DDoS), malware and ransomware, piracy, cyberbullying, and online fraud and scams. It is crucial to note cybercrime has become more professional, automated, complex, and stealthy. The technology used to commit cybercrime is commercialized, and experienced hackers can be hired to commit the crime (Grabosky, 2016; Holt, 2012; Wall, 2001b). Cybercrime is predicted to inflict damages totaling $ 6 trillion by 2025 (Morgan, 2021).

Terrorists are known to engage in cybercrime offenses, referred to as cyberterrorism, which is the convergence of cyberspace and terrorism (K.-s. Choi et al., 2018; Conway, 2014; Denning, 2000, 2001). The term cyberterrorism has sparked an intense debate over whether terrorist groups are capable of attacking cyberspace and critical infrastructures to further their goals (Conway, 2003, 2014; Hardy & Williams, 2014; Jarvis & Macdonald, 2014;

Weimann, 2004, 2005). Although terrorist groups have yet to cause massive loss of life or economic chaos by hacking into critical infrastructure systems, they have engaged in cyber-attacks against websites and other non-essential infrastructure through hacking (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt, 2012). Additionally, the ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities (Carlin & Graff, 2018; S. M. Furnell & Warren, 1999). Cybercriminals and cyberterrorists are attacking computer systems and networks, critical infrastructures, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy.

Most of the studies on terrorists' use and/or misuse of the Internet have identified different ways in which terrorists use the internet. The fundamental use is to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, information dissemination, and secure communication (Conway, 2017; Hoffman, 2006b; Weimann, 2006). The groups use both the indexed and accessible Surface Web and the non-indexed and hidden Dark Web to further their goals. The radical right, ISIS, and other Jihadist groups use multiple platforms; for instance, they would use Twitter for audience development and YouTube or Instagram to share videos with the hope of recruiting (Weimann, 2016b). Interactive platforms such as chat rooms, message boards, forums, and social media platforms are beneficial for the interactions amongst like-minded individuals. Hence bringing together individuals, fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst participants (Bloom & Daymon, 2018; Gaudette et al., 2020; Piazza & Guler, 2019). Hence, active online discussions about ideologies and opinions with friends or strangers have fostered like-minded individuals' belongingness and virtual communities of like-minded individuals (Bowman-Grieve, 2009; Scrivens et al., 2018). Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community.

Terrorists use the internet predominantly to spread propaganda and psychological warfare. Islamist groups, for instance, Al-Qaeda and ISIS, have dedicated media councils overseeing their communication strategy (Hoffman, 2006b; Ligon et al., 2017). The role of the media jihad (Erez et al., 2011) is to produce, distribute, and collect jihadist content for TV, Radio, websites, and social media platforms. ISIS is the most prolific group on the internet, the most sophisticated in media and technology usage (Bisgin et al., 2019; Ligon et al., 2017). To reach their global audience, the groups publish and share information about attacks, ideologies, strategies, and justifications in Arabic, English, and other local languages. They publish online magazines such as Al-Qaeda's Inspire and ISIS' Dabiq and Rumiyah, in Arabic, English, French, Russian and Turkish (Bisgin et al., 2019; Welch, 2018; Wozniak et al., 2020) to encourage their supporters to commit terrorist attacks, provide instructions on how to make a bomb, assemble weapons, and become a homegrown terrorist. In addition, they create films known as 'Mujahideen films,' which have been posted on the internet or sold as DVDs. The films depict groups attacking U.S military forces on patrol, transmitting the last words of kidnapped Iraqis and foreigners about to be executed, and appealing for financial contributions.

The internet has become a better platform for terrorist groups to spread their ideologies through meta-narratives. A meta-narrative is a "theory that tries to give a totalizing, comprehensive account to various historical events, experiences, social, and cultural phenomena based upon the appeal to universal truth or universal values" (Ali & Bwana, 2015, p. 22). The metanarratives of Islamist terrorist groups such as Al-Shabaab, al-Qaeda, and ISIS are based on the misinterpretation of Islam's religious traditions and historical events for ideological purposes. Al Qaeda and ISIS base their political ideology on Islam, "on the religious tradition of Islam, appropriating and transforming key elements from the Quran and the Hadith, from Mohammed's life story and from the early history of Islam for its own ideological purposes" (Schmid, 2014, p. 4). Their messaging is strategic; for instance, Al-Shabaab's metanarratives for its targeted audience in Kenya include: there is a war against Islam in

Kenya; Muslims in Kenya have a duty to wage 'holy war' against non-Muslims, and the government of Kenya; and the only way to address and achieve victory against the 'war on Islam' is through armed struggle, self-sacrifice and/or active support to their version of 'Jihadist' cause (Ali & Bwana, 2015). They use the meta-narratives of victimization to appeal to the recruits, inform, educate and solicit support (Anzalone, 2016; Hoffman, 2006a).

Jihadist's ideologies ascribed by Al Qaeda, ISIS, and their affiliates follow three themes; the West is hostile to Islam and the Muslim world; the caliphate, a political entity will replace the corrupt rules under the Western influence; the only way to address the West is in the language of violence, by waging jihad (holy war) against the non-believers, the West (Hoffman, 2006b; Pelletier et al., 2016; Schmid, 2014; Weimann, 2011). The ideology is meant to inspire believers to join the global jihad to create an Islamic State that enforces strict Islamic law. As Weimann (2006) notes, a terrorist campaign has four target audiences for its messages, namely the supporters, the population being served, the enemy, and international public opinion. Messages vary depending on the audience; the message amplifies pride, success, commitment, and vision to the supporters, potential recruits, and community. To the enemy, the message is meant to cause fear and threat with the hope of pressuring the government to accept the terrorists' demands. To the rest of the world, the message is to justify their cause explain why and what they want to achieve, i.e., freedom, equal rights, a free state, etc. Therefore, the messaging seeks to express outrage, call for jihad, violence against the West, and to gain support for their Islamic cause. The groups have gained sympathizers and insurgents through their propaganda tactics and narratives.

There has been a growing effort in preventing and countering violent extremism (P/CVE) (Greenberg, 2016; Romaniuk, 2015; Saltman et al., 2021; Tuck & Silverman, 2016). As a policy and practice, countering violent extremism (CVE) has rapidly emerged as the most significant development for counterterrorism efforts. The common strategies include disruption, diversion, and counter-messaging (Greenberg, 2016). Officials and practitioners are responding to online extremism by implementing counter-narrative measures as the 'soft'

solution to the problem of terrorism. The measures include using social media, newspapers, radio, and texting messages as tools for countering online extremism. The counternarratives provide discussions of Islam that correct the misinterpretation of the Quran and religion, refuting the "idealization of life inside the Caliphate. Instead of a perfect, protected, peaceful life, viewers see images of rape, death, and general suffering for members of the Caliphate" (Greenberg, 2016, p. 172). Other CVE programs include community engagement and outreach through town halls, roundtables, capacity building of youth, and community development (Gartenstein-Ross & Barr, 2016; Hussain & Saltman, 2016; Romaniuk, 2015).

Diversion and disruption occur through technical interventions by removing and disrupting terrorist platforms and online content. For instance, social media platforms used by terrorist groups and their supporters such as Facebook and Twitter, and forums such as Gab and Stormfront are being monitored and suspended (Behr et al., 2013; J. M. Berger, 2016; Caló & Hartley, 2019; Zelin & Fellow, 2013). The monitoring and disruption from Surface Web have led to the migration to encrypted platforms such as Telegram and the Dark Web (Malik, 2018; Weimann, 2016a). Most Internet users only access the surface web, also known as the Internet, easily accessible through browsers such as Google and Yahoo. The Dark Web, a subsection of the Deep Web, is a hidden part of the Internet that can only be accessed via special browsers such as TOR (Malik, 2018; Mirea et al., 2019; Monk et al., 2018; Weimann, 2016b). The Dark Web or Dark Net can only be accessed through encrypted Internet browsers; it offers anonymity, making it attractive for individuals interested in privacy and avoiding censorship and criminals, including terrorist groups evading law enforcement. Terrorists use the Dark Web for secure communication, discussion, and networking with their supporters through forums, as a repository of their propaganda materials, and financing using cryptocurrencies (Malik, 2018; Weimann, n.d.).

While previous research has highlighted the problem of terrorist use of the Internet, specifically social media platforms and websites, there has been little research exploring cyberterrorism and the use of Dark Web forums by terrorist groups and their supporters.

Additionally, although criminologists have applied classical research methods in the analysis of crime and terrorism and contributed to the development of novel research methods, crime theories, findings, and policies, few have yet to embrace computational approaches to the study of crime and crime-related problems, specifically terrorism and extremism studies (Bisgin et al., 2019; Greene & Lucas, 2020; Hernandez-Suarez et al., 2018; Scrivens et al., 2018). Social scientists can benefit from the new interdisciplinary field of computational social science, which is the development and application of computational methods, such as machine learning and natural language processing, to analyze complex, large-scale behavioral data (Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020). The digitization and availability of large-scale information generated by internet users (Hofman et al., 2021; Kemp, 2022) have rendered traditional manual methods for collecting and analyzing terrorist content challenging, time-consuming, and costly (Brynielsson et al., 2013; Scrivens et al., 2018). As such, social scientists working with text data can benefit from several statistical and computational techniques such as topic modeling algorithms latent Dirichlet allocation (LDA) that automatically uncovers topics and the hidden thematic structure of texts (Blei et al., 2003; DiMaggio et al., 2013; Maier et al., 2018; Walter & Ophir, 2019). Additionally, apply social network analysis (Borgatti et al., 2018; Scott & Carrington, 2014) to expose and map covert networks and structures (Koschade, 2006; Mullins, 2013; Perliger & Pedahzur, 2011).

The availability of digital data and the advancement of computational processing power and techniques has not only reduced the human effort in analyzing text data but allowed social science researchers to analyze large-scale text data, ask new questions, observe human interactions and behavior, thus advancing criminological research on terrorism, cyberterrorism, extremist online content, and counterterrorism. This dissertation aims to address this gap by exploring cyberterrorism and applying computational techniques and social networks to analyze Dark Web forums of international jihadists and their supporters.

1.2 Research Objectives

This three-article dissertation address three problems from an interdisciplinary, computational social science perspective, the first is an exploration of existing research on cybercrime and cyberterrorism. The second is methodological; providing insight into how researchers can analyze the large-scale text data generated by terrorists and their supporters. The article introduces and demonstrates topic modeling, specifically, Latent Dirichlet Allocation (LDA), in highlighting topics and themes from jihadists Dark Web forums. The third article explores Jihadists' network structure and behavior in Dark Web forums using social network analysis techniques.

The purpose of this three-article dissertation composed of three distinct but connected studies is threefold a) to explore the issue of cybercrime and cyberterrorism from a criminological perspective, b) highlight Jihadist topics and discussion on the Dark Web using innovative computational techniques, specifically topic modeling, c) illuminate and map Jihadist Dark Web forums' social structure and behavior using social network analysis. Additionally, the dissertation aims to contribute to the field of criminology innovative computational social science approaches for detecting terrorists' online contents and technology use and adds to the existing literature knowledge on cyberterrorism and online behavioral patterns of Jihadists and their supporters.

1.3 Research Organization

The dissertation is organized as follows, the first article, chapter two, explores the literature on terrorists' use of technology, cybercrime, and cyberterrorism. The nature of cybercrime and cyberterrorism are technological, criminal, and social, requiring analysis from a disciplinary and interdisciplinary lens. In this chapter, I provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology. The article focuses on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals

Appendix GUETLER01

commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers. The article ends with a discussion of cybercrime and cyberterrorism's future directions.

The second article, chapter three, is a methodological study introducing the application of topic modeling in analyzing large-scale text data. I make a case for applying natural language processing computational techniques, specifically LDA topic models and topic network analysis of text data. I begin by discussing machine learning, natural language processing, its challenges, and the various topic models algorithms. I then demonstrate the potential of topic models by inductively analyzing large-scale textual data of terrorist groups and supporters from three Dark Web forums to uncover underlying topics. Topic models' inductive approach and measurements are a powerful tool for the exploratory and descriptive text analysis for social scientists and data. I conclude the chapter by discussing the benefits and limitations of topic models in uncovering topics.

Chapter four, the third article, is an exploratory study of the social networks of Dark Web forums used by international Jihadist groups and their supporters. Using social network analysis, I construct undirected networks to analyze the network's structure, the interactions and participation patterns between members within the forums and identify the important members and virtual communities. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? What are the posting activities of key members in the forums? The paper begins by describing the technological structure of the Internet, specifically the Dark Web. Next, the related literature in terrorists' use of the Internet, virtual communities and terrorists' online content, terror on the Dark Web, and social networks of terrorist groups. The last sections explore the research design, methods, data, and results. Finally, the article concludes with a discussion about Dark Web forums and their network structures. The dissertation concludes with a summary of the major findings, limitations, and future research on cyberterrorism and use of the Internet.

Appendix GUETLER01

Chapter 2

Technology, Terrorism, and Crime: Exploring the Intersections of Cybercrime and Cyberterrorism

2.1 Introduction

The digital revolution and the development of new technologies, including the Internet, Web 2.0, artificial intelligence, Internet of Things, the metaverse, and social media networks, have rapidly transformed all aspects of social, political, economy, communications, and cultural life. More than 50% of the world's population is online, with about one million additional people joining the Internet each day (Forum, 2020; Kemp, 2019). As a result, cybercrimes and cyberterrorism are the unintended consequences of the digital revolution. Cyberspace has become a powerful platform for criminals. Cybercriminals and cyberterrorists are attacking computer systems and networks, critical infrastructures, stealing data and information, thus increasingly posing serious threats to national security, economy, and public policy. Progressing in number, scale, and consequences, cybercrimes are a challenge for intelligence and security agencies tasked with combating online crimes. According to the FBI Internet Crime

Appendix GUETLER01

Complaint Center (IC3, 2020), in 2020, they received 791,790 complaints from Americans with a reported loss exceeding $4.2 billion, representing an increase of nearly 69% in total Internet crime complaints between 2019 and 2020. The most common types of cybercrime reported by victims include phishing scams, non-payment/non-delivery, extortion, personal data breach, and identity theft. Some victims also complained of ransomware, denial of service, malware, terrorism, and hacktivism (IC3, 2020). Cybercrime is predicted to inflict damages totaling $6 trillion by 2025 (Morgan, 2021).

Research examining the advancement of technology and terrorist groups has increased dramatically over the past decade, the focus has been chiefly on the use of the Internet, including social media platforms and websites for communication, financing, radicalization, and recruitment (Aly et al., 2014; Bermingham et al., 2009; Bloom et al., 2019; Gill et al., 2017; Hoffman, 2006b; Macdonald et al., 2019; Sageman, 2011; Scrivens et al., 2021; Weimann, 2011). However, there is a limited but growing body of criminological scholarship examining cybercrimes (Burruss et al., 2012; Holt, 2013; Holt & Bossler, 2009; Holt, Cale, et al., 2021; Howell et al., 2019; Lee & Holt, 2020; Maimon & Louderback, 2019). Few researchers have examined terrorist groups attacking digital infrastructures, stealing information, sensitive data, and finances to further their ideologies (K.-s. Choi et al., 2018; Holt, Cale, et al., 2021; Holt et al., 2019; Holt & Steinmetz, 2020). Some scholars consider such cyberattacks by terrorist groups an act of cyberterrorism (T. M. Chen et al., 2014; S. M. Furnell & Warren, 1999; Holt et al., 2019; Holt, Navarro, et al., 2020). Others debate whether cyberterrorism is a real threat (Conway, 2003, 2014; Denning, 2000, 2001; Pollitt, 1998; Weimann, 2004). Nonetheless, cyberspace remains a valuable borderless environment for criminal and terror activities, where the "likelihood of detection and prosecution is estimated to be as low as 0.05% in the United States" (Forum, 2020, p. 63).

Prior to the 9/11 terror attacks, criminological and sociological studies on terrorism were nonexistent, overlooked, and neglected (Armborst, 2010; Black, 2004; Deflem, 2004; Silke, 2004). The field was dominated by political scientists, international relations, and terror-

ism studies scholars, conducting research on terrorism from a political and psychological perspective. Furthermore, "criminology has been remiss in its research into the phenomena of cybercrime and has been slow to recognize the importance of cyberspace in changing the nature and scope of offending and victimization" (Jaishankar, 2007, p. 1). Despite the empirical gap, much has since changed within criminology and the research on terrorism. Criminologists are now researching terrorism, and a limited number of criminological studies examining cybercrime and cyberterrorism are published in top-tier criminological journals (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Maimon & Louderback, 2019). In addition, criminology scholars are advancing a new and developing field of criminological study known as cyber criminology, defined as "the study of causation of crimes that occur in the cyberspace and its impact in the physical space" (Jaishankar, 2007, p. 1). Thus, a multidisciplinary field relying upon criminology, sociology, and computer sciences (Diamond & Bachmann, 2015). However, as Diamond and Bachmann (2015) note, the field of cyber criminology is "largely ignored or marginalized by mainstream criminology, and that many criminologists refrain from examining this important, future-oriented issue" (Diamond & Bachmann, 2015, p. 25). Furthermore, the shortage of cybercrime and cyberterrorism research has been attributed to data limitations, the lack of official statistics on cybercrime and cyberterrorism incidents (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Navarro, et al., 2020) and criminologists' unfamiliarity with the technology and tools for collecting online data (Diamond & Bachmann, 2015; Maimon & Louderback, 2019). Research on cybercrime and cyberterrorism is increasing despite these challenges, albeit still limited.

The nature of cybercrime and cyberterrorism are technological, criminal, social, and economical, requiring analysis from a disciplinary and interdisciplinary lens (Payne & Hadzhidimova, 2020). Drawing from previous research that highlighted the state of the field and gap in cybercrime (Maimon & Louderback, 2019; Payne & Hadzhidimova, 2020) and ideologically-motivated cyberattacks (Holt, Lee, et al., 2020; Holt et al., 2012; Holt, Turner, et al., 2021),

this study aims to provide an analysis of the problems of cybercrime and cyberterrorism within the field of criminology. Focusing on (a) the issues in defining terrorism, cybercrime, and cyberterrorism, (b) ways that cybercriminals commit a crime in cyberspace, and (c) ways that cyberterrorists attack critical infrastructure, including computer systems, data, websites, and servers. Finally, the paper ends with discussing cybercrime and cyberterrorism's future directions.

2.2 Understanding terrorism, cybercrime, and cyberterrorism: Definitions

To understand the relationship between cybercrime and cyberterrorism, it is critical to understand terrorism and extremism. Unfortunately, there is no single consensus on terrorism, cybercrime, and cyberterrorism (Crenshaw, 2011; Hoffman, 2006a; LaFree & Ackerman, 2009; Schmid, 2004; Silke, 2004). Most books and studies on terrorism begin with discussing the definition of terrorism and its varieties (Silke, 2004). Consequently, studying terrorism is complicated, as it poses conceptual and methodological challenges as opposed to most types of criminal violence (Armborst, 2010; LaFree & Ackerman, 2009; LaFree & Dugan, 2015; Schuurman, 2019). Initially, criminologists shied away from studying terrorism, citing a) terrorism is different from other forms of violence, terrorism is not 'common' violence, despite terrorists being arrested and tried on murder, assault, weapon charges, b) terrorism cannot be explained by theories of violence, c) terrorism is political violence, criminological theories are intended to explain non-political violence (Rosenfeld, 2002). Yet, "while it seems reasonable that criminological theory alone cannot sufficiently explain terrorist violence, it is not convincing to exclude it from criminological research for this reason alone" (Armborst, 2010, p. 417). Neither does seeking first to find a "general theory" of terrorism necessary (Laqueur, 2004). Terrorism is a criminal matter (Hamm, 2007; Hamm & Van de Voorde, 2005), and terrorist groups engage in other types of crime, including bank robbery, international drug

trade, drug trafficking, smuggling humans, and money laundering.

The FBI defines international terrorism as the "violent, criminal acts committed by individuals and/or groups who are inspired by, or associated with, designated foreign terrorist organization or nations" they also define domestic terrorism as "violent, criminal acts committed by individuals and/or groups to further ideological goals stemming from domestic influences, such as those of a political, religious, social, racial, or environmental nature" (FBI, 2021). In this study, I adopt the criminologist's operational definition of terrorism, which is "the threatened or actual use of illegal force directed against civilian targets by nonstate actors in order to attain a political goal through fear, coercion, or intimidation" (LaFree & Ackerman, 2009, p. 348). Hence, the varying definitions of terrorism are framed in the context of socio-political, religious, cultural identification and the use of violence, fear, and threats to accomplish the ideological goals (Borum, 2011; LaFree & Dugan, 2009; Silke, 2004).

In addition, there are varying definitions of the term extremism. The definitions are based in the context of violent extremism and are frequently used interchangeably with the term terrorism (J. M. Berger, 2018; J. Berger, 2017). However, extremism is not always violent and associated with non-state actors. J. Berger (2017) defines extremism in terms of in-group/out-group relations, perceptions, and beliefs. Thus, extremism is "a spectrum of beliefs in which an in-group's success is inseparable from negative acts against an out-group. Negative acts can include verbal attacks and diminishment, discriminatory behavior, or violence" and violent extremism as "the belief that an in-group's success is inseparable from violence against an out-group. A violent extremist ideology may subjectively characterize this violence as defensive, offensive, or pre-emptive" (J. Berger, 2017, p. 6). Ideologies are the central element of violent and non-violent extremism (McCauley & Moskalenko, 2008). Categorized as international terrorism, individuals and groups espousing Islamist ideologies practiced by jihadists seek a violent military strategy to accomplish their goal of establishing an Islamist society. Often categorized as domestic terrorism, the far-right

ideology is espoused by White supremacists, militia, and militant gun rights adhering to racial supremacy and anti-government beliefs. In contrast, the far-left groups advocate for workers and animal rights, environment, and race-based issues.

There is no agreement on what constitutes cybercrime. As such, the debate on defining cybercrime centers around whether cybercrime should be conceptualized as a new crime type or traditional crime conducted through a new medium, in this case, technology (Diamond & Bachmann, 2015; Holt, 2016; Holt et al., 2012). For example, Furnell and colleagues (2015) define cybercrime as cyber-dependent and cyber-enabled. Cyber-dependent crimes are offenses that can only be conducted using a computer, computer networks, or technologies such as spreading viruses, malware, hacking, and distributed denial of service (DDoS) attacks directed against computers or networks. In contrast, cyber-enabled crimes are traditional crimes that use computers or computer networks to widen their reach or scale. Cyber-enabled crimes include fraud, theft, and sexual offending (S. Furnell et al., 2015).

Some scholars define cybercrime based on four types of activities and behaviors (Holt & Bossler, 2014; Wall, 2001a). The four categories include cybertrespass or hacking, which is the unauthorized access of a networked system or computers. It is considered an intrusion when an individual accesses a computer system, network, or data without the owner's permission. Hackers are known to access the networks of individuals, businesses, and governments without authorization.

Cyber-deceptions or cyber-theft include identity theft, online fraud, and digital piracy. The use of the Internet to steal information, credit and debit cards, money from bank accounts belonging to individuals or businesses. Cyber-deception also includes crimes simplified by the internet but do not require it. For instance, the Nigerian or 419 messages and work-at-home schemes using spam emails have adapted the offline fraud schemes to the online environments. Additionally, digital piracy, which is the stealing of intellectual properties, has been simplified by the Internet and technology.

Cyber-pornography/obscenity crimes include the distribution of sexually explicit mate-

rials online, child sexual exploitation materials, and the virtual sex trade. Cyber-violence causes psychological harm to victims through cyberstalking, cyberbullying, harassment, and hate speech. Extremists and terrorist groups also engage in acts of cyberviolence (A. M. Bossler & Berenblum, 2019; Holt, 2016; Holt et al., 2012; Wall, 2001b). Based on these definitions and categorizations, the range of activities viewed as cybercrimes are challenging for any definition of cyberterrorism, and terrorist groups may also engage in the same acts as non-ideologically motived cybercriminals (Holt et al., 2012).

The term cyberterrorism has sparked an intense debate over whether terrorist groups are capable of attacking cyberspace and critical infrastructures to further their goals (Conway, 2003, 2014; Hardy & Williams, 2014; Jarvis & Macdonald, 2014; Weimann, 2004, 2005). Similar to the term terrorism, various definitions of cyberterrorism exist. Cyberterrorism referred to the convergence of cyberspace and terrorism and was coined by Barry Collin in the 1980s (K.-s. Choi et al., 2018; Conway, 2014; Denning, 2000, 2001). Subsequent definitions have focused on the previous definitions and challenges in defining terrorism. Denning (2000, 2001) defines cyberterrorism as the

> "Unlawful attacks and threats of attacks against computers, networks and the information stored therein when done to intimidate or coerce a government or its people in furtherance of political or social objectives. Further, to qualify as cyberterrorism, an attack should result in violence against persons or property or at least cause enough harm to generate fear. Attacks that lead to death or bodily injury, explosions, or severe economic loss would be examples. Serious attacks against critical infrastructures could be acts of cyberterrorism, depending on their impact. Attacks that disrupt nonessential services or that are mainly a costly nuisance would not" (Denning, 2000, p. 29).

It is implied that for cyberterrorism to occur, terrorist groups must attack computers or networks, for instance, attack a power grid or access air traffic control leading to death, injury, or economic loss. In addition, (Conway, 2014) argues contrary to what the media

reports or government authorities, per this definition, terrorist groups sending pornographic emails to minors, defacing web pages, posting offensive content on the Internet, stealing credit card information does not constitute cyberterrorism. Thus, an attack only qualifies as cyberterrorism if it fits the definition of terrorism. However, some scholars argue terrorist groups can commit cyberterrorism by targeting government or businesses' online critical infrastructures (S. M. Furnell & Warren, 1999; Holt et al., 2019; Holt, Lee, et al., 2020). Furthermore, law enforcement and government agencies acknowledge the capabilities and potential cyberattacks by terrorist groups (Holt, Lee, et al., 2020). In this case, cyberterrorism can be understood as "the use of computer network tools to shut down critical national infrastructures (such as energy, transportation, government operations) or to coerce or intimidate a government or civilian population" (Yannakogeorgos, 2014, p. 44). This definition does not include violence as the end goal; thus, cyberterrorism should be distinctly defined from other forms of terrorism. Therefore, "while there is no single agreed upon definition for cyberterror, it is clear that this term must encapsulate a greater range of behavior than physical terror due to the dichotomous nature of cyberspace as a vehicle for communications as well as a medium for attacks" (Holt et al., 2012, p. 341). An act of cyberterrorism should not necessarily result in physical harm or intended to generate fear if its intention is the interference of the political, social, or economy of a nation or lead to physical violence.

Weimann (2004, 2005) claims at present, terrorists are using the internet and benefit from it rather than attacking it. Yet, cyberterrorism is an attractive option for terrorists who value its potential to inflict massive damage, its psychological impact, and the media appeal. No one has yet been physically hurt or killed by cyberterrorism. However, this does not mean terrorist groups are not interested in cyberterrorism. Nevertheless, whether terrorist groups weigh the benefits and costs of a cyberattack or possess the technical skills to attack critical infrastructure instead of bombing or suicide should not be the focus of a debate. The fact that the groups have access to the Internet, can learn technical skills, or hire hackers deserves serious attention from scholars, governments, and national security

policymakers. The lack of consensus on the definitions or elements comprising cyberterrorism should not be the focus of disciplinary and interdisciplinary debate but rather on the threat and changing dynamics of cyberspace and technology and the motivations and capabilities of terrorist groups to engage in cyberterrorism.

2.3 Cybercrime and Cybercriminals

Criminological research examining cybercrimes has increased over the past decade and addresses a variety of online offenses, including hacking, website defacement, malware, piracy, and cybercrime offenders (Holt & Bossler, 2009; Holt, Turner, et al., 2021; Maimon & Louderback, 2019; Wall, 2001b), advance-fee fraud and romance scams (Dion, 2010; Holt & Graves, 2007; Kopp et al., 2016; Whittaker & Button, 2020), identity theft (Archer, 2012; J. Choi et al., 2021; Gies et al., 2021; Rudner, 2008) and cyberbullying (Kerstens & Veenstra, 2016; Marcum & Higgins, 2012; Su & Holt, 2010; Zhang et al., 2021). Some scholars have analyzed the application and testing of criminological theories in explaining cybercrime, such as the explanatory power of self-control theory and routine activities theory (Back et al., 2018; A. Bossler & Burruss, 2011; A. Bossler & Holt, 2009; Howell et al., 2019; Kigerl, 2011; Leukfeldt & Yar, 2016; Louderback & Antonaccio, 2020; Reyns et al., 2011), social learning theory (Holt, 2010; Li et al., 2016; Miller & Morris, 2016; Morris & Higgins, 2010; Skinner & Fream, 1997), neutralization theory (Higgins et al., 2008; Morris, 2011) and strain theory (Hay et al., 2010; Jang et al., 2014). Cohen and Felson (1979) routine activities theory has been "applied in most of the victim-based studies of cyber-dependent crime" (Maimon & Louderback, 2019, p. 201), finding moderate support for the theory. As a result, "while support for different theories exists, no apparent general theory of cybercrime explains all of the offenses. Instead, some of them appear to be better suited for certain types of crimes" (Payne & Hadzhidimova, 2020, p. 82). Indeed, finding the general theories of crimes inadequate in explaining cybercrime, cyber criminologist Jaishankar (Jaishankar, 2007) developed

the space transition theory to explain the causation of crimes within cyberspace. The theory claims people behave differently (conform or non-conform) when moving from one space to another, i.e., from physical space to cyberspace and vice versa.

Hacking is the starting point from which cyber-dependent crimes begin, whereby criminal hackers deliberately access a computer system, networks, and data unauthorized (S. Furnell, 2003; S. Furnell et al., 2015; Maimon & Louderback, 2019; Schell & Dodge, 2002). Once hackers gain access to a computer system, they may develop or spread viruses and malicious software, launch website defacement or distributed denial of service attacks, destroy or alter files, redesign or configure hardware and software systems (Marcum et al., 2014). In addition, criminal hackers are behind major ransomware attacks in businesses and governments, posing infrastructure and human risk.

Cybercriminals' predominant hacking tools and tactics include phishing or spamming, a social engineering attack where an attacker uses social skills to obtain and compromise information from an individual or organization's computer system. For instance, a user may click a malicious link or file in an email from a 'reputable' credit card company or bank, allowing hackers access to a personal account or financial information. A typical cyberattack is the use of malicious software or malware; this occurs when a virus, worm, or Trojan horse is spread to gain access into a computer system. Ransomware is malware that extracts and encrypts data making it inaccessible and unusable. The victim must first pay a ransom, often using cryptocurrency before the information or system is released and available. It is essential to note the malware software is automated. It infects the computer, takes the victim's money, and deposits it to the attacker's account without the attacker being present or near the victim's computer. Malware can be spread via email attachments, advertisements, fake software installations, phishing emails, text messages, infected USB drives, and apps.

Web defacements or vandalism occurs when an attacker changes the "front face" of the website with their messages, images, or text. Such defacement attacks are meant to cause the website to be offline and unavailable, damage the reputation or brand of the individual,

organization, or government. Distributed Denial of Service (DDoS) attacks intentionally overwhelms a website, usually a government or business website, by disrupting the regular traffic of the server, service, or network, thus becoming unavailable to users (Hayward & Maas, 2021; Holt, 2010; Holt & Steinmetz, 2020; Singer & Friedman, 2013).Cybercrime has become more professional, automated, complex, and stealthy (Wall, 2015). Furthermore, the technology used to commit cybercrime is commercialized, and experienced hackers can be hired to commit the crime. While some hacker tools may be free to download, the better ones are available for sale or rent (Grabosky, 2016). For instance, the Cybercrime-as-a-Service (CaaS) kits sold on the Dark Web enable cybercriminals to attack cyberspace without the technical expertise of the computer or systems. The commercialization of technology tools for cybercrime may be due to the low technical capabilities and skills of most hackers but also provides cybercriminals with access to vulnerable and compromised systems with the possibility of evading detection or arrest.

Hackers are motivated by curiosity, technical achievement, fun, addiction, power and prestige, peer associations, ideology, and, most importantly, monetary gain (Grabosky, 2016; Weulen Kranenbarg et al., 2018; Weulen Kranenbarg et al., 2019). Within the hacker subculture, norms such as privacy and secrecy, meritocratic value, knowledge distribution networks exist (Holt & Graves, 2007; T. Jordan & Taylor, 1998; Steinmetz et al., 2020). In addition, hacking is dominated by males; often a male world, young, technologically oriented, and unfriendly to women (Adam, 2005; Bachmann, 2010; Holt, Navarro, et al., 2020; Hutchings & Chua, 2016; T. Jordan & Taylor, 1998; Steinmetz et al., 2020; Turkle, 2005). Hacking behaviors begin during early adolescence (Holt, 2007; T. Jordan & Taylor, 1998; Lee & Holt, 2020), and many cybercriminals may work alone, but "a great deal of cybercrime is the work of organizations" (Grabosky, 2016, p. 37). Furthermore, cybercrime organizations have structures and roles including "disorganized groups of cyber-dependent offenders that mainly operate in online environments yet have no clear chain of command (i.e., a swarm structure) and organized groups with clear leadership and command structure (i.e., a hub

structure)" (Maimon & Louderback, 2019, p. 195). For instance, the hacktivist group Anonymous, which hacks for a political cause, is decentralized and coordinated (Singer & Friedman, 2013). Anonymous have launched distributed denial-of-service (DDoS) attacks against the government and prominent organizations' websites. These include attacking the Australian Parliament, the US Federal Bureau of Investigation, the US Department of Justice, PayPal, MasterCard, Visa (Hardy & Williams, 2014). Hackers can also be classified as white hat hackers, employed to test system security thus authorized, versus black hat hackers who access systems unauthorized. Since cybercrime is a global phenomenon and a transnational business, the offenses can originate globally, and offenders can easily cross borders without detection (Payne & Hadzhidimova, 2020).

2.4 The cybercrime-cyberterrorism nexus: Relationship between cybercrime and cyberterrorism

The ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities (Carlin & Graff, 2018; S. M. Furnell & Warren, 1999). Terrorists use the Internet to further their goals and advance and adapt to technological changes. As a result, cyberspace is indeed "the new great equalizer, the Internet of Things is the new battlefield, anonymizing tools are the new weapons for stealth, malware is the new air strike, vulnerabilities are the new focus of attack, and robust zero day arsenals are the new symbols of supremacy as the world order is technologically toppled and supplanted by the new" (Scott & Spaniel, 2016, p. 2). Although terrorist groups have yet to cause massive loss of life or economic chaos by hacking into critical infrastructure systems, they have engaged in cyber-attacks against websites and other non-essential infrastructure through hacking (Carlin & Graff, 2018; K.-s. Choi et al., 2018; Hardy & Williams, 2014; Holt et al., 2017a; Holt et al., 2012). Indeed, the first recorded cyberattack by terrorist groups was by the Tamil Tigers, who carried out a denial-of-service attack against

Sri Lankan embassies worldwide (S. M. Furnell & Warren, 1999). Recently, in 2020, the US Justice Department, the Department of Homeland Security, and the Department of Treasury dismantled three terrorist cyber-enabled finance campaigns. The Al-Qassam Brigade terror group ran an online cryptocurrency fundraising effort soliciting bitcoins from its supporters. Al-Qaeda and its affiliated groups also operated a bitcoin money-laundering network using Telegram channels and other social media platforms and a scheme by an ISIS facilitator to sell fake masks online for funding purposes (ICE, 2020).

However, less research has explored cybercrimes and similar methods conducted by terrorist groups and ideologically motivated cyber attackers (T. M. Chen et al., 2014; Holt et al., 2019; Jarvis & Macdonald, 2014; Yannakogeorgos, 2014). As such, the dearth of cyberterrorism research poses challenges in developing counterterrorism policy and prevention strategies related to ideologically motivated cyberattacks (Holt et al., 2017b). There is no difference in skills and techniques between cyberterrorists and cybercriminals. Both require and utilize the same techniques for attacking cyberspace. However, cyberterrorists are motivated by political or ideological issues (S. M. Furnell & Warren, 1999; Holt et al., 2012). Although cyberterrorists are also known to hire and fund available hackers to carry out attacks on their behalf (Carlin & Graff, 2018), some hackers may not believe in the terrorist's cause but only work for financial gain, signifying collaborations between criminals and terrorist groups. In addition, some of the terrorist supporters and group members possess less technologically sophisticated skills and are taught by the skilled hackers "operating ISIS's Cyber Help Desk" (Scott & Spaniel, 2016, p. 2). As a result, cyberattacks may require only a few skilled hackers to train the rest and fewer financial resources and workforce.

Furthermore, terrorist groups such as ISIS recruit technology-savvy hackers known as cyber-jihads, members of the Cyber Caliphate, a hacker division of ISIS (Scott & Spaniel, 2016). The group mostly attacked websites by replacing the content with the ISIS flag and phrases. The digital jihadis are believed to have begun with the British terrorist of Pakistan descent, Junaid Hussain (Carlin & Graff, 2018). With a group of seven friends,

Junaid formed a hacker group called TeaMpOisoN. The group became famous in 2011 after they defaced/vandalized websites with pro-Palestine messages. In addition, they attacked websites belonging to NASA, NATO, UN, and figures such as former Prime Minister Tony Blair. Hussain was a prominent recruiter within the Islamic State, and it is believed he is responsible for developing much of their cyber and social media strategies. He may also have been developing al-Qaeda's Dabiq and Kybernetiq publications on information technology, communication, and security. Another hacker, Ardit Ferizi, accessed U.S computer systems to obtain personal information on 1,300 military and federal employees and then provided this information to ISIS members (Holt et al., 2017a). More cyber jihadis groups have been formed, including the Cyber Caliphate Army and United Cyber Caliphate.

As mentioned elsewhere, known cyber-attacks by terrorist groups include web defacements and information release, killing lists of military personnel with the hope that lone wolf recruits will conduct attacks on the targets, and credit card fraud to finance their cause. The defacements are meant to scare Western businesses and organizations than recruit new followers (Scott & Spaniel, 2016). Distributed denial of service (DDoS) is another method terrorist groups may use to attack websites and servers. This attack involves flooding requests to the website and server, rendering it 'busy' and unusable to others, causing financial harm to the victim (Holt, Turner, et al., 2021). The economic costs for the victims include damage and destruction of data, deletion of hacked data and systems, loss of revenue from site downtime, productivity costs, and reputational damages (Holt, Lee, et al., 2020).

Additionally, supporters of Al-Qaeda also operate web forums to distribute hacker tools and coordinate attacks. For instance, the hacker Younis Tsoulis promoted hacking tools against various targets in support of global jihad. Using the handle Irhabi 007, or Terrorist 007, he published a manual entitled "The Encyclopedia of Hacking the Zionist and Crusader Websites," detailing attack methods and a list of vulnerable targets online (Holt et al., 2012, p. 344). As a result, Tsoulis and his compatriots became the first people in Britain to be convicted of conspiring to commit terrorism on the Internet (Carlin & Graff, 2018).

Indeed, as Verton and Brownlow (2003) claim, "the cyber-threat from terrorist organizations such as al-Qaeda does not hinge on the willingness of the core leadership or even the most dedicated and radicalized jihadists within the movement to adopt cyber-tactics as a main attack method. There are already a vast number of cyber-based alliances of hackers and internet activists (aka "hacktivists") cyber-jihad movement, or electronic holy war, against the West" (Verton & Brownlow, 2003, p. 101). Like the cyber jihadis, far-left extremists have also engaged in website defacement campaigns against industrial and governmental targets to shame their victims for harming animals and the environment. The far-right groups have also defaced websites serving Holocaust memorial sites (Holt, Lee, et al., 2020). Indeed, "such actions would correspond in part to physical terror and extremist violence, which target infrastructure and persons who symbolically represent their ideological beliefs" (Holt, Lee, et al., 2020, p. 4). As such, high-profile websites such as those belonging to governments, businesses, and organizations are likely targeted by terrorist groups and extremists.

Terrorist groups are developing their cyberattacks strategies. For instance, the "lower-level actors such as Boko Haram are upgrading their 419 scams with ransomware, RaaS, and MaaS" (Scott & Spaniel, 2016, p. 2). Primarily by using Ransomware as a Service (RaaS) and Malware as a Service (MaaS) which ransomware developers lease to people without much technical knowledge and skills enabling them to launch ransomware attacks just by paying for the service (Europol, 2021; Ledesma, 2021). Ransomware remains a threat as cybercriminals tend to increase pressure by threatening the publication of data if victims do not pay (Europol, 2021), affecting not only the organizations or governments but those whose data is compromised. The prepackaged exploitation of vulnerabilities such as malicious software or malware (Singer & Friedman, 2013) is significant for terrorist groups. Mainly since "the use of known vulnerabilities may be associated with semi-skilled attackers as they may be able to utilize existing, pre-written attack programs to compromise vulnerable systems. In fact, a number of attack tools available for download include exploitable code for known vulnerabilities so as to facilitate attacks by lower-skilled cyberattackers" (Holt,

Lee, et al., 2020, p. 5). The dark web is the marketplace for malicious software and hacker tools, where cybercriminals and terrorists buy and sell the tools and services needed for a cyberattack (Singer & Friedman, 2013; Weimann, 2016a).

The technical sophistication and capabilities of terrorist groups to engage in cyberterrorism is often the central contention in the cyberterrorism debate (Evan et al., 2017; Singer & Friedman, 2013). The extent to which terrorist groups may engage in cyberterrorism is based on their technical capabilities and sophistication. For instance, according to Scott and Spaniel (2016), Al-Qaeda relied on more innovative partner terrorist organizations such as the Tunisian Cyber Army for cyber-attacks. In 2015, the group developed its cyber-arm that performs electronic jihad operations. The group was active on Twitter and video distribution sites but did not possess the social media presence, recruiting capabilities, and technical sophistication such as ISIS. The less technically sophisticated Al-Shabab shares many strategies with Al-Qaeda. Al-Shabaab predominantly uses the Internet to disseminate propaganda and recruitment. However, Al-Shabaab's "disorganized internet strategy and lack of technical proficiency prevent it from posing a real threat in cyberspace" (Scott & Spaniel, 2016, p. 8). Since Boko Haram switched allegiance to ISIS, Boko Haram's social media presence and the distribution of propaganda have become more sophisticated. However, the group still raises funds through fee fraud or 419 scams. ISIS is the most sophisticated group. It has recruited a following of technology-savvy kiddies and wannabe hackers (Scott & Spaniel, 2016). ISIS has dedicated its offensive cyber capabilities to compromise social media accounts and websites belonging to individuals, businesses, and government organizations. They have also used malware, preconfigured tools, and insider threats (Scott & Spaniel, 2016). Thus, the groups are also innovating and adapting to technological changes.

On the other hand, Evan et al. (2017) developed three levels of capability scales, enabling, disruptive and destructive, to classify the cyber-capabilities of these groups. The groups use the Internet for publicity, propaganda, recruitment, communication, and disseminating manuals/magazines online within the enabling level. The disruptive level is online activities

disrupting the computer systems, e.g., cyber breaches, malware, financial theft and fraud, denial of service attacks, phishing, and web defacements. Within the destructive level, the groups engage in cyber-attacks that trigger physical damage or injury through sensor spoofing technology, digital control systems, disabling control, and safety systems. However, Evan et al. (2017) claim most terrorist organizations such as ISIS, Al-Qaeda, Al-Shabaab, Boko Haram possess the 'enabling' cyber capability as they have websites, social media accounts, and media. They have some 'Disruptive' abilities, as seen by the cyber breaches, web defacement, and financial fraud. However, none of the groups have shown that they possess the higher destructive capability yet. An organization would have to have supporters and group members possessing advanced hacking and engineering skills for higher cyber capability. So far, none of the terrorist groups have this capability. However, "the imminent acquisition of cyber capabilities by terrorist groups has been long expected but has so far failed to materialize, and there have been no known terrorist attacks using cyber means to trigger physical damage and destruction" (Evan et al., 2017, p. 4). Overall, the benefits of cyberterrorism include reduced risk of capture, inflicting financial damage without loss of life, hackers and tools are available for hire, and the groups may widely publicize the successful attack while failure would go unnoticed. As such, researchers should not perceive the threat of cyberattacks by extremists and terrorist groups as unrealistic or non-existent. Evidence suggests that terrorist groups have the skills and means to engage in cyberattacks and use the Internet to further their goals.

2.5 Conclusions: Future Directions for Cybercrime and Cyberterrorism Research

Due to the rapid technological advancement and emerging technologies, including artificial intelligence, and the unintended consequences of these technologies, it is challenging to predict cybercriminals and cyberterrorists' technical capabilities and cyberattacks. In-

deed, while there have been few incidents of cyberterrorism globally, mostly attacks against government, organizations, and businesses networks, we cannot ignore the threat of severe cyberattacks and critical infrastructures posed by terrorist groups. Countries face a growing risk of cybercrime and cyberterrorism as we rely more on the Internet and technologies for communication, business, health, economy, governing, and critical infrastructures. While terrorism scholars have been hesitant in acknowledging that terrorists engage in cyberterrorism, law enforcement and government agencies do recognize terrorists and extremist groups may target government and civilian managed online critical infrastructure (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021).

The lack of access to official data on cyberattack incidents is one of the reasons for the dearth of cyberterrorism and cybercrime research (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Lee, et al., 2020). A few official data sources provide data on computer crimes such as hacking. Industry or open-source government data "are rarely available due to underreporting on the part of victims, particularly within the private industry over fears of economic loss and a decrease in consumer confidence" (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021, p. 6). Businesses and organizations tend not to report cyberattacks to the police or the public. Furthermore, cyberattack incidents are often omitted from terrorism databases "due to their failure to qualify as a "violent" incident" (Freilich et al., 2014; Holt, Lee, et al., 2020; Holt, Turner, et al., 2021). As such, most scholars have resorted to examining a sample of college and school students, mainly surveying their hacking and digital piracy experiences, analyzing bulletin boards and blogs, and field experiments (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Burruss et al., 2012; Holt & Steinmetz, 2020). Perhaps the innovative data collection from honeypots, simulated computer systems created to be attacked, researchers then analyze the data on system intrusions (Maimon et al., 2013; Testa et al., 2017; Wilson et al., 2015) could solve the lack of data challenges. However, while honeypots collect data on behavior (Bossler A.M., 2017), this type of data collection has limitations, including the inability to collect

demographic data from the actors.

There are currently no existing statutes for cyberterrorism within US federal criminal code. Terrorists engaging in cyberattacks are often prosecuted under existing computer hacking criminal laws (Holt, Lee, et al., 2020; Holt, Turner, et al., 2021). Thus, to defeat cyberterrorism globally, creating a "legal framework for prosecution with a strong foundation in international law" (Stockton & Golabek-Goldman, 2014, p. 214) would be significant for pursuing and prosecuting cybercriminals. Perhaps, cybercrime and cyberterrorism researchers may consider a collaborative research analysis with legal scholars to analyze the existing laws and contribute to the implementation of relevant policies on cybercrime and cyberterrorism from the perspectives of criminology.

Crime is one of the unintended consequences of old and new technologies. The field is changing, criminological research on cybercrime is progressing, researchers and colleges are advancing and offering a new subfield of cyber criminology. Research clusters are also developing, for instance, digital criminology (Powell et al., 2018), computational and Big Data criminology (Chan & Bennett Moses, 2016; Smith et al., 2017), and artificial intelligence crime (AI-Crime) (Hayward & Maas, 2021; King et al., 2020). Research groups are formed in educational institutions such as the Evidence-Based Cybersecurity Research Group at Georgia State University, which researches human factors in cybersecurity. Additionally, a significant and growing number of cybercrime and cyberterrorist presentations at national and international conferences exist (A. M. Bossler & Berenblum, 2019). The American Society of Criminology (ASC) Division of Cybercrime was formed in 2019. Undergraduate, graduate, and certificate offerings in cybercrime are increasing. Nodeland et al. (2018) argue, "students pursuing careers in criminal justice must be prepared to respond to modern threats including, but not limited to, a range of cyber-related offenses (e.g. cyberbullying, cyberstalking, sexting and cybersex crimes) as well as the traditional hacker, identity theft, terrorists, and nations wishing harm on the United States" (Nodeland et al., 2018, p. 71). Yet, most universities are lagging in their educational and training opportunities in the field

of cybersecurity for their criminal justice students (Nodeland et al., 2018).

Social scientists researching cybercrime tend to focus on the offenders and victims while placing less emphasis on the technical aspects of cybercrime (Holt, 2016). As a result, less criminological research in cybersecurity focuses on securing cyber systems and preventing cybercrimes. Technical analysis has been left to computer scientists with less input from the social sciences. Thus, "it is vital to identify potential methods for proactive actor motivation detection by cybersecurity professionals based on actor behavior during and after the incident is complete to better defend against future attacks and minimize harms" (Holt, Turner, et al., 2021, p. 544). Indeed, the "behavioral side of cybersecurity needs more research and can improve faster if it is integrated with human factors, and benefit from sophisticated modeling and simulation techniques" (Maalem Lahcen et al., 2020, p. 2). As Hayward and Maas (2021) note, "if many computer scientists are understandably guilty of concentrating their efforts on the technology side of what we might call here the 'tech-crime nexus', then the same is true in reverse of criminologists. While our discipline continues to advance knowledge about crime and punishment, it does so largely oblivious of the many social challenges posed by technological disruption" (Hayward & Maas, 2021, p. 210). The study of cybercrime and cyberterrorism requires a multidisciplinary focus and collaboration. Criminologists must collaborate with computer scientists and industry to develop knowledge on cybercrime, cyberterrorism, and cybersecurity to better anticipate cybercriminal activities and prevent, respond to, and mitigate cyberattacks.

With the advancement of artificial intelligence technologies and algorithms, new crimes have been developed, and new types of policing, punishment, and legal decision-making (Brayne, 2020; Brayne & Christin, 2021; Hayward & Maas, 2021). Therefore, criminology research should also focus on the policing and security practices that use technologies and algorithms, such as predictive policing and surveillance technologies, and how these tools are used to investigate criminals of all types, including cyberterrorists. The U.S criminal justice system uses predictive technologies. For instance, police departments use predictive

software programs to predict where and when future crimes will occur and target potential victims and offenders (Brayne, 2020; Brayne & Christin, 2021; Ferguson, 2017). Indeed, as Ferguson (2017) argues, "new technologies threaten to impact all aspects of policing, and studying the resulting distortions provides a framework to evaluate all future surveillance technologies" (Ferguson, 2017, p. 4). Hence, criminological research should also focus on the potential harms of predictive and surveillance technologies used by law enforcement.

Finally, social scientists, including criminologists, can also benefit from the adoption and application of a computational social science (Y. Chen et al., 2021; Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020; Molina & Garip, 2019; Salganik, 2019) approach for the analysis of crimes in cyberspace. The technological revolution has led to the digitization of social, economic, political, and cultural activities of billions of Internet users, thus generating "vast repositories of digital data as a byproduct" (Hofman et al., 2021, p. 181). While existing research methodologies, theories, and data have proven utility in examining crime, new innovative interdisciplinary and borrowed techniques from computer science and data science are necessary to analyze big data, including text and image data. Future research applying computational and network techniques to explore the online content and messaging of terrorist groups and their supporters on the Dark Web is imperative for counterterrorism. As terrorists have migrated to encrypted platforms such as the Dark Web, which is also a cyberthreat, it has become challenging for law enforcement and security agencies to monitor, detect, and analyze their use of encrypted platforms. The groups use the Dark Web marketplaces to hire hackers and purchase hacking tools to attack cyberspace. The Dark Web forums also provide secure communication and discussion platforms for the groups and their supporters. The following studies examine the topics and networks of Jihadists Dark Web forums from a computational social science approach.

Chapter 3

Text as Data: Topic Modeling of Dark Web Forums

3.1 Introduction

As of January 2022, over half of the global population is online, and Internet users spend more than 40 percent of their day online. As such, billions of Internet users' social, economic, political, and cultural activities have been digitized, generating vast amounts of data (Hofman et al., 2021; Kemp, 2022) that can also be used for descriptive and causal inferences. Traditional research methods and techniques may not analyze the influx of big data. The text and image big data generated from search and social media platforms are often noisier and unstructured, not fitting tidily into the row and column structure of traditional social science data such as surveys and experiments (Hofman et al., 2021). To analyze big data, social scientists can benefit from the new field of computational social science, which is the development and application of computational methods, such as topic modeling, to analyze complex, large-scale behavioral data (Hofman et al., 2021; Keuschnigg et al., 2018; Lazer et al., 2020).

Appendix GUETLER01

Natural language processing (NLP), a subfield of computer science, uses computational techniques to learn, understand and produce human language. The computational analysis of text data has grown over the past 20 years mainly due to increased computing power, the availability of large amounts of text data, and the development of successful machine learning (ML) methods (Hirschberg & Manning, 2015). NLP combines concepts from linguistics, statistics, machine learning, and deep learning enabling computers to understand spoken and written human language. Some NLP applications on text data include sentiment analysis that measures subjective content such as attitudes and emotions; and named entity recognition that identifies valuable words or phrases such as names and locations. As one of the NLP approaches, topic modeling is a computational technique that uncovers topics and the hidden thematic structure of a collection of texts. As a result, social scientists working with text data can benefit from several statistical topic modeling algorithms, such as the latent Dirichlet allocation (LDA) (Blei et al., 2003; DiMaggio et al., 2013; Maier et al., 2018; Walter & Ophir, 2019). Although there are several topic modeling algorithms, latent Dirichlet allocation (LDA) (Blei et al., 2003) is one of the widely used statistical algorithms for analyzing large-scale text data to find the topics and thematic structure. Most importantly, LDA's inductive approach and quantitative measures make it suitable for the exploratory and descriptive analyses of text data (Maier et al., 2018). Additionally, LDA models are useful for automatically detecting and removing terrorist content from online platforms.

The traditional approach of manual hand-coding text data is limited in terms of resources, the analysis is time-consuming, and concerns arise about biases of researchers or challenges in achieving acceptable intercoder reliability scores, especially when there is a large corpus to analyze (DiMaggio et al., 2013; Roy & Rambo-Hernandez, 2021). It is important to note, however, manual coding of text can be applied to analyze smaller samples of text data and effectively discover themes. Furthermore, the qualitative approach to text data using grounded theory (Glaser Strauss, 1967) and the quantitative approach from natural language processing (Blei et al., 2003) involve similar processes and produce similar results

(Baumer et al., 2017). However, the availability of digital data and the advancement of computational processing power and techniques have reduced the human effort in analyzing text data and allowed social science researchers to analyze large-scale text data, ask new questions, and observe human interactions and behavior.

In this methodological study, I make a case for applying natural language processing computational techniques, specifically LDA topic models and topic network analysis of text data. I demonstrate the potential of topic models by inductively analyzing large-scale textual data of terrorist groups and supporters from three Dark Web forums to uncover underlying topics and sentiments. I argue that topic models' inductive approach and measurements are a powerful tool for the exploratory and descriptive text analysis for social scientists and data. The organization of this article is as follows. I begin by discussing machine learning, natural language processing, its challenges, and the various topic models algorithms. In the next section, to highlight the utility of topic models, I provide an example of topic models by analyzing Dark Web forums of jihadist terrorist groups. I conclude by discussing the benefits and limitations of topic models in uncovering topics.

3.2 Machine Learning and Natural Language Processing

Within the field of artificial intelligence, machine learning (ML) focuses on developing algorithms and computer systems that improve automatically through experience and the use of data (M. I. Jordan & Mitchell, 2015). In ML, the algorithm learns from training data and improves as it learns more from processing data. Machine learning is an interdisciplinary field comprising computer science and statistics. It is used to develop software for computer vision, speech recognition, natural language processing, other applications, and predictions and inferences. Machine learning approaches in text analysis can be categorized into supervised, unsupervised, and semi-supervised learning methods with inputs from both

Appendix GUETLER01

humans and machines. In supervised machine learning, the algorithm 'learns' the patterns from input (training) data to make predictions. For instance, a researcher can use available labeled text data to train the model, then the model learns it and classifies the rest of the unseen (test) data. In unsupervised learning, algorithms learn from unlabeled data to find hidden patterns. Here, the researcher's role is to interpret the model's results. Finally, the algorithms learn from both smaller labeled and larger unlabeled datasets in semi-supervised learning to find patterns. Here the researchers' label/annotate a smaller sample first to train a model (Lucas et al., 2015; Vajjala et al., 2020).

Another subfield of artificial intelligence is natural language processing (NLP), which provides the tools and techniques for the learning, understanding, production, and analysis of human language content (Hirschberg & Manning, 2015). As such, the goal of NLP is to make the human language accessible to computers so that the machines can understand and respond to human languages. Since the "statistical or corpus ("body of words")-based NLP was one of the first notable successes of the use of big data, long before the power of ML was more generally recognized or the term "big data" even introduced" (Hirschberg & Manning, 2015, p. 261), NLP is a suitable approach for the analysis and understanding of language. NLP combines concepts and techniques from linguistics, computer science, statistics, and machine learning models. Methodological applications of NLP include machine translation, speech-to-text dictation, chatbots, spell check, spam detection, spoken-dialog devices (e.g., Siri, Alexa), information extraction, text classification, and automatic summarization, semantic and sentiment analysis.

NLP computing is complex and challenging mainly because natural language is ambiguous, complicated, evolving, and diverse and requires context to convey the intended meaning, domain-specific, social, and world knowledge (Hirschberg & Manning, 2015; Manning & Schutze, 1999). Without context and world knowledge, it is difficult for humans or machines to discern whether an utterance is sarcastic. For instance, the sentence "Oh, it is snowing again" may be sarcastic if you understand the context; it has been snowing for a long time,

and the speaker has had enough of the cold. This task may be easy for a human being to understand based on the context. However, while the NLP models may have learned all the definitions and synonyms, understanding the sentence in context is challenging.

Language is diverse and evolving, with different vocabularies, accents, and regional dialects, including slang. While it may be less challenging for some humans to understand the various languages and slang/colloquial based on one's culture, machines may find it challenging to understand the evolving language, including informal expressions and slang, mainly because these words do not exist in the dictionary or have a formal definition. As a result, models trained to detect hate speech are likely to falsely identify a dialect, for instance, the African American Vernacular English, as hate speech (Bender, 2019; Hvitfeldt & Silge, 2021; Sap et al., 2019). Furthermore, NLP "still cannot perform common-sense reasoning or draw on world knowledge in a general and robust manner" (Bird et al., 2009, p. 33). The models must be constantly updated with the language changes to reflect the new changes. The update would require more data for the algorithm to learn. Despite these challenges, statistical NLP approaches solve some of these problems by "automatically learning lexical and structural preferences from corpora [...] the use of statistical models offers a good solution to the ambiguity problem: statistical models are robust, generalize well, and behave gracefully in the presence of errors and new data" (Manning & Schutze, 1999). The statistical NLP approaches include topic modeling, which I describe below.

3.2.1 Topic Modeling

Topic models are generative probabilistic models used for text mining and uncovering latent topics from documents. In generative probabilistic modeling, data is treated "as arising from a generative process that includes hidden variables. This process defines a joint probability distribution over both the observed and hidden random variables" (Blei, 2012, p. 79). Topic model algorithms automatically discover themes and how the themes are connected in a large and unstructured collection of documents (Blei, 2012). As such, "this method scales to

billion-word datasets and can arguably provide an analysis driven more by the documents than by human preconceptions" (Baumer et al., 2017, p. 2). In analyzing terrorists' online outlets, topic models are more beneficial in the automatic and quick detection of large-scale terroristic content than manual annotation. Topic modeling can also find patterns in genetic data, images, surveys, and social networks (Blei, 2012; L. Liu et al., 2016).

There are several topic models; in this study, I am interested in the unsupervised latent Dirichlet allocation (LDA) topic model, which does not require prior annotations or labeling of the data (Blei, 2012; Blei et al., 2003). LDA automatically finds topics from documents based on words co-occurrence and frequency distribution. The documents and words are observed, while the topic structure, which includes topics, per-document topic distributions, and the per-document per-word topic assignments, is hidden (Blei, 2012). The benefit of topic models is that the hidden topic structure resembles a document's writing process, making it easier to interpret and label the topics. The high probability words in each topic distribution can be easily interpreted as a theme, therefore a topic (Baumer et al., 2017). Additionally, topics are "inferred from a given collection without input from any prior knowledge. Since topics are hidden in the first place, no information about them is directly observable in the data" (Maier et al., 2018, p. 94). The assumptions of LDA models include a) a topic is a distribution over vocabulary, i.e., a set of words, and b) a document is a mixture over topics. The researcher must specify the number of topics to be modeled.

Extending LDA, other topic models incorporate meta-data into the topic models, such as author, title, location, links, and year (Blei, 2012). For example, both topics and authors are generated in the author-topic model (Rosen-Zvi et al., 2004). In addition, the topics are attached to authors, providing the analysis of which authors are likely to write similar work and topics. Significant for social science research, the structural topic modeling (STM) (Roberts et al., 2013) includes metadata to analyze the topics and topic structure. For instance, the author, political views, and ideologies contribute to the words used in any given topic and document. Both these topic models can be applied to analyze text data; the

choice of an algorithm depends on the research questions and objective of the analysis.

Other topic models have extended the function of LDA to combine topic and network analysis. For example, the relational topic model (RTM) (Chang & Blei, 2009) models the documents and assumes the links between the documents depend on the distance between their topic proportion. "Unlike traditional statistical models of networks, the relational topic model takes into account node attributes (here, the words of the documents) in modeling the links" (Blei, 2012, p. 83), taking into account links such as citation and attributes such as text (Chang & Blei, 2009). Words or concepts are modeled as nodes and relationships, such as co-occurrence within documents, as edges (e.g., Walter & Ophir, 2019). Similarly, Walter and Ophir (2019) have developed the analysis of topic model networks (ANTMN) to include topics models and network analysis. In ANTMN, the topics are nodes, and edges are the relationships between the nodes calculated by their co-occurrence over documents and clustered using community detection techniques into topics. Community detection algorithms identify communities in networks based on the structural position of nodes (Molina & Garip, 2019).

3.2.2 Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is another NLP method for analyzing text data and is significant for sociological research (B. Liu, 2012; Pozzi et al., 2016; Puschmann & Powell, 2018; Roy & Rambo-Hernandez, 2021). Sentiment analysis (SA) extracts and analyzes subjective information from text data, such as opinions, sentiments, evaluations, attitudes, and emotions towards topics, events, organizations, individuals, or things (B. Liu, 2012). Using the words written by a user, SA measures how a person might feel about a topic (positive, negative, neutral, emotions – anger, happy, frustrated etc.); hence textual data expressing opinions, emotions, and sentiments are imperative for SA rather than factual data (Puschmann & Powell, 2018). Opinions influence our behaviors and attitudes, and the analysis of Jihadists forums and their supporters highlights their

sentiments towards terrorism (Abbasi & Chen, 2007; Ahmad et al., 2019; Greene & Lucas, 2020).

The approaches to SA are mainly supervised and unsupervised machine learning. In supervised learning, the models learn from labeled data, while in unsupervised learning, the models are dictionary-based, trained using lexicons and sentiment scoring. Lexicons are dictionaries or lists of terms classified with a polarity score as negative or positive and used to analyze the sentiment of the text (Puschmann & Powell, 2018; Roy & Rambo-Hernandez, 2021; Schweinberger, 2022). There are several lexicons; the most common include AFINN (Nielsen, 2011), NRC word-emotion association (Mohammad & Turney, 2013), Bing lexicon (Hu & Liu, 2004) based on customer reviews, and VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014) which is specifically adapted to sentiments from social media data.

Although SA is conducive to measuring opinions from text data that affect our behaviors, the method has limitations. The lexicons used to train the data are domain-specific; while they can accurately detect sentiment in one context/data, for instance, marketing research, the lexicons may not be applicable in other contexts, such as terrorism. Nevertheless, both topic models and sentiment analysis can be applied to analyze and understand the topics and opinions from online platforms. However, the goal of this research was to examine and find the hidden topics discussed on Jihadists' Dark Web forums. Therefore, most of the analysis and discussion will focus on topic models and the results.

3.3 Applications

In the following section, I introduce three applications of LDA models using text data from Dark Web forums as case studies. First, I explore the background and then describe the data and the text analysis process.

3.3.1 Detecting Terror on the Dark Web: A Case Study

Although most Internet users are using it for legitimate reasons and benefiting from the technologies, terrorist groups and their supporters use the Internet to further their ideologies and goals posing new national security and counterterrorism challenges. Terrorism is a form of communicative violence (Aly et al., 2017), as such terrorist groups use the Internet, specifically social media platforms and messaging apps, to spread propaganda, recruitment, radicalization, financing, and communication, establish their appeal and popularity, and interaction with low risk of being captured (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Gill et al., 2017; Sageman, 2011). One way of examining the online content and narratives of terrorist groups is by applying real-time computational techniques to highlight the topics discussed.

Social media accounts and websites of terrorist groups and their supporters are removed or suspended from the Internet by social media companies and counterterrorism agencies. As a result, the groups have moved to the Dark Web, the part of the Internet that is hidden, not indexed, unregulated, and therefore not accessible via search engines. Within the Dark Web, "the decentralized and anonymous networks enable evading arrest and the closure of the terrorist platforms" (Weimann, 2016a, p. 41). The groups moved to the encrypted platforms to protect their information from being removed and further disseminate news and propaganda, finance, and recruitment. As such, individuals seeking terrorist content can do so on the Dark Web without being monitored or detected (Malik, 2018). Additionally, "terrorist forums on the Darknet not only encourage individual radicalization but also promote a "self-starter" type of terrorism. This is a strategy endorsed by IS both online and offline and motivates vulnerable individuals to commit violence in the organization's name in an attempt to "crowdsource" terrorism" (Malik, 2018, p. 20).

Before the growth of the Internet, most terrorist networks resulted from face-to-face interactions among friends and family (Sageman, 2011). The Internet has allowed like-minded individuals to interact and form virtual communities; however, passive absorption of in-

formation does not change someone's mind. Communication through forums or chat rooms contributes to interactions paving the way to changing minds and radicalization. As such, active online discussions about ideologies and opinions with friends or strangers and the sense of belonging within the virtual community (Bowman-Grieve, 2009; Scrivens et al., 2018) or the virtual Muslim community (Ummah) with jihadist ideologies (Erez et al., 2011) has the potential of changing someone's mind rather than just reading impersonal stories. Participants post and share their views and knowledge within the virtual communities and offer support and justification for using violence and validation. The virtual networks comprise unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals. The forums, not passive websites, are crucial in radicalization. Thus, creating opportunities for self-radicalization, connection, and interaction of like-minded people globally while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence.

Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community. As time goes on, the users in these platforms have roles including producing and aggregating content, translating and curating content across platforms, or passively consuming the information (Bloom et al., 2019; Shehabat & Mitew, 2018). As participants begin to feel comfortable and welcomed on these platforms, skeptics and lurkers will be encouraged to share their beliefs and become more vocal. For instance, ISIS uses an army of 'media mujahideen' to amplify their online messaging campaign, "these IS-supporting internet users meticulously share and repost official content across several social media platforms, as part of a coordinated effort to maximize the organization's impact and relevance, and to assist in the recruitment to the group" (Malik, 2018, p. 21). In addition, since participation in these forums is an active process, users tend to select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). Therefore, the analysis of topics is significant for understanding terrorists' communication patterns and narratives within online platforms.

The more time individuals spend online interacting with terrorists and their supporters, the more likely they may be radicalized into violent or radical social networks. A user can be defined as radicalized based on the content they post or share, i.e., anti-Western, pro-terrorist rhetoric, if they post very negative and inciting comments to the forum. For example, a long-time member of a discussion forum who posts harmful material over time suggests a dedication to an extremist movement. Or an author who participates for a short time but posts very radical views implies the user has become radicalized (Scrivens et al., 2021; Scrivens et al., 2018). However, researchers attempting to classify users and their online posts as a signifier of radicalization (Ahmad et al., 2019; J. R. Scanlon & Gerber, 2015; Scrivens et al., 2021; Scrivens et al., 2018) have found that classifying a text or user as extremist is challenging due to the different kinds of extremism and ideologies, various targets, different ways of expression and language. For example, a pro-terrorist material and discussion forum would include terms and topics such as a call for jihad, Islam, mujahideen (members of the Taliban), apostate, caliphate, Ummah (Muslim community with jihadists worldview, nation), crusaders, promote violence towards the West, Christians, Jews, call for martyrdom, suicide operations, infidels (enemies, or non-believers), and offensive or abusive language such as kuffar (non-believers, non-Muslims). Forum posts of jihadists can be classified into themes based on information dissemination, religious preaching, instruction or training, and social interactions (Erez et al., 2011; Rowe & Saif, 2016). As such, topic models are imperative in classifying topics and themes within forums and websites.

Criminological/sociological theories and scholarship emphasis social interactions (Goffman, 1959) and criminal behavior (Akers et al., 2014; Gottfredson & Hirschi, 1990) within a social context, providing frameworks for understanding online social ties and terrorism. In the presentation of the self, Goffman (1959) argues social interactions are performances where individuals manage how they want others/ the audience to perceive them. As such, one would expect individuals interacting in an online forum to conform, hide their true identities and beliefs to fit in. Indeed, individuals interacting in online platforms imagine

their audiences and act accordingly (Marwick & boyd danah, 2011; Németh & Koltai, 2021). Social control theorists argue we are all capable of committing a crime (Hirschi, 1969, p. 31). The critical question is not what causes crime but why individuals conform. Control perspectives maintain that deviation from conforming behaviors is likely to occur when social bonds to family, peers and conventional society are weak or nonexistent. The Internet's role in radicalization and violent extremism has become a growing concern. Online content and platforms are conducive to online radicalization, as users seeking justification for joining or intending to join terrorist groups are likely to be indoctrinated by these messages and radicalized others.

The social bond theory (Hirschi, 1969) remains one of the most cited theoretical frameworks in sociological and criminological research. The theory holds that individuals are less likely to commit a crime if they have stronger social bonds to family, peers, and legitimate institutions. The theory has four central elements: attachment, commitment, involvement, and belief. The first element is attachment to others and sensitivity to their opinion. Attachment contains a moral component, as an attached individual considers others' reactions, such as parents, peers, or teachers, before engaging in deviant acts. Commitment refers to conventional lines of action; it is the rational component of conformity. An individual invests time in getting an education, starting a business, or establishing a reputation. Thus, before engaging in deviant behavior, they consider the act's cost. Those who have invested time and energy in getting an education, building a business, or establishing a reputation are less likely to engage in crime for fear of losing their investment. Involvement in conventional activities prevents individuals from engaging in deviant behavior. A person busy doing conventional activities has less time to engage in crime. Finally, belief refers to the commitment and devotion to society's standard value system; individuals who believe in society's laws are likely not to break them. The four elements are interrelated; the more individuals are affectingly attached to conventional others, the more they are committed to conventional systems, the more involved in those conventional systems, and the more they

believe in conventional values and norms. As a result, they are likely to be more conforming and less delinquent.

Social learning theory (SLT) (Akers et al., 1979; Akers et al., 2014) presumes that individuals' differences in criminal behaviors stem from learning through association or interactions with others. The premise of SLT is that all social behavior is learned, including crime, and has four key elements: 1) differential association, which refers to direct social interaction with members of a primary group. 2) imitation, which occurs when individual copies the behavior of others, 3) definitions, which serve as behavioral guideposts for how we think about certain behaviors as good or bad, rewarding or punishing and 3) differential reinforcement, which exists in both social and nonsocial forms, are anticipatory or prospective (Akers et al., 2014; Akins & Winfree, 2016). Social learning theory utilizes general principles of social behavior to offer a social psychological explanation of criminal and deviant behavior, the acquisition, maintenance, and change in criminal or conforming behavior through a learning process (Akers et al., 2014). Therefore, SLT can explain why individuals support and sympathize with terrorist groups. Exposure to deviant peers and family (differential association) predicts an individual's likelihood of engaging in crime. Individuals associating with others online who define violent extremism as favorable are likely to adopt those beliefs and eventually engage in violent extremism. As such, social bond and social learning theories have great potential in explaining the role of online bonds and interactions in violent extremism. Although these theories were formulated to explain offline interactions and behavior, few studies have yet to apply the theories and topic models in explaining whether online interactions and posting behavior are related to violent extremism.

3.3.2 Dark Web Forum Data

Data used for analysis is accessed from the University of Arizona Artificial Intelligence Lab's Dark Web Forums (https://www.azsecure-data.org/dark-web-forums.html). The forums were collected as part of the lab's study of the international Jihadi social media and

movement (H. Chen, 2011; H. Chen et al., 2008). The open-source dataset has provided researchers with data that would typically not be collected without computational knowledge, tools, or access to the forum. The Dark Web Forum repository includes twenty-eight jihadist dark web forums in Arabic, English, Spanish, French, German, and Russian. Forums are organized into threads and posts; the metadata includes members and dates collected between 2004 and 2012.

I analyzed three forums; the first is Ansar1, an English-language forum with 29,492 posts and 11,244 threads made by 382 members between 12/8/2008 and 1/20/2010. The second forum is Gawaher, an English language Islamic forum dedicated to discussing the Islamic world and Islam issues. Some of the forum members sympathize with radical Islamic groups. This forum entails 372,499 posts and 53,235 threads made by 9,269 members between 10/24/2004 and 6/7/2012. Finally, the Islamic Network forum is dedicated to various topics of interest to Muslims, ranging from theology and world events. Some members sympathize with and support terrorist organizations. This forum has 91,874 posts and 13,995 threads created by 2,082 members between 6/9/2004 and 11/10/2010.

Forums	**Year**	**Posts**	**Threads**	**Users**
Ansar1	2008-2010	29,492	11,244	382
Gawaher	2004-2012	372,499	53,235	9,269
Islamic Network	2004-2010	91,874	13,995	2,082

Table 3.1: Dark Web Forums Data

3.4 Text Analysis

Text analysis is conducted using R and Python statistical analysis tools and LDA packages. R packages for LDA include topicmodels (Gruen & Hornik, 2011), LDAtuning (Nikita & Chaney, 2020), tm, quanteda (Benoit et al., 2021), tidyverse packages (Wickham, 2021), tidytext (Queiroz et al., 2021) and Python package langdetect (Danilak, n.d.). In addition,

Appendix GUETLER01

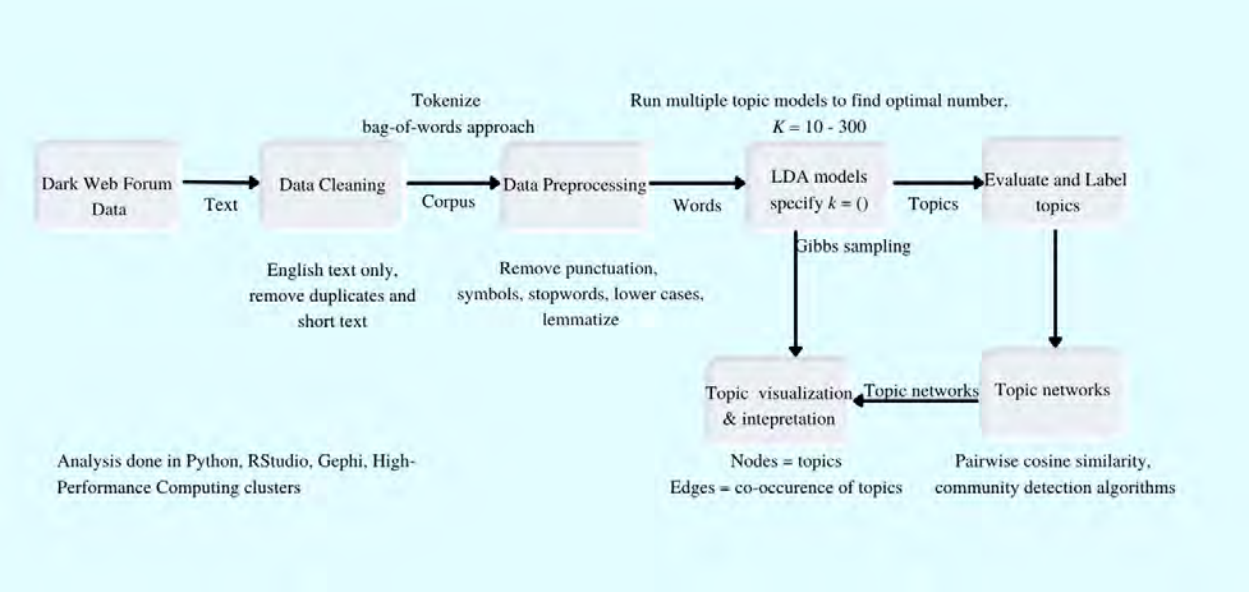


Figure 3.1: Text Analysis Workflow

there are multiple ways for preprocessing text data for analysis (Baumer et al., 2017; Roy & Rambo-Hernandez, 2021), and the order of the preprocessing matters (Maier et al., 2018). Below I outline the steps I took in my analysis.

3.4.1 Data Pre-processing

In LDA, the text collection to be analyzed is referred to as corpus (body of text), for instance, the Ansar Dark Web forum. An item within the corpus is a document, i.e., the message/post, and words within the document are called terms. To analyze text, one must first process it into a structured, machine-readable format. While text analysis tools include translation packages that English content by filtering and removing non-English text from the data by using a language-detection algorithm in Python, mainly because the data is predominantly in English. I then removed short texts and duplicates to analyze a sample of the raw data. The next step was to tokenize the text and break down the texts/sentences into words or tokens using the bag-of-words approach, which analyzes groups of words together, representing words in a phrase disregarding the order in which they appear (Lucas et al., 2015; Silge & Robinson, 2017), resulting in a matrix with the unique terms and words frequencies.

I removed punctuations, numbers, and symbols and converted capital letters to lowercase letters. Common stop-words such as "the," "and" "a" were also removed as they are not helpful for the detection of relevant topics. Additionally, words that appeared frequently and infrequently were filtered and removed. Similar to Walter and Ophir (2019) and Maier et al. (2018), words that occurred in more than 99% of the documents or less than .5% were removed. This process reduces the size of the corpus vocabulary, enhancing the algorithm's performance (Lucas et al., 2015; Maier et al., 2018). Stemming and lemmatization are NLP processes used to reduce inflection forms of a word to a common base or dictionary form, i.e., its stem or root, lemma. However, due to outputs of the stemming process often not making sense, it is difficult to interpret the words, thus impacting the topic quality (Maier et al., 2018; Walter & Ophir, 2019) by degrading topic stability (Schofield & Mimno, 2016). Therefore, I skipped the stemming process as "interpreting word stems correctly can be tough, or even impossible" (Maier et al., 2018, p. 101). Finally, I lemmatized some of the words to their lemma forms, i.e., 'study,' 'studies,' 'studying,' are collapsed to 'study.'

3.4.2 LDA Models

The next step is to train the LDA models using Gibbs sampling iteratively. This algorithm uses "thousands of iterations through the documents, considering the topic proportion of each word token in turn. Words that frequently occur together in documents are likely to be placed into the same topic so that each document contains relatively few topics" (Baumer et al., 2017, p. 7). The number of topics, K, is a parameter that must be defined for an LDA model. The number of topics can be anything from 2 to 100 based on the research questions, the overall size of the corpus, and evaluation metrics. The more topics, the more specific the topic outputs will be. On the other hand, too many topics may lead to a similar outcome that is not meaningful, or too few topics may lead to broad outputs that combine different terms that should have been separated (Maier et al., 2018). Walter and Ophir (2019) suggest specifying many topics rather than a few as duplicate topics can then be grouped using the

community detection algorithms.

To assess and evaluate the ideal number of topics for each Dark Web forum, a range of topic numbers from 10 - 300 were estimated and evaluated using four different model fit indicators from the LDAtuning package. The model fit metrics are based on the accuracy (Arun et al., 2010), density (Cao et al., 2009), latent concept modeling (Deveaud et al., 2014), and Markov chain Monte Carlo algorithm (Griffiths & Steyvers, 2004). Then I used the output from the indicators to select a final number of topics for each data set. The next step was to evaluate the relevant topic numbers and then rerun the model. Extract the theta and beta statistics from the output and plot, interpret, and label the topics. Similar to Walter and Ophir (2019), I analyzed the results of the top words in the topics, words that occur frequently and are exclusive to each topic, and the original documents representing each topic. I then interpreted and labeled the topics by examining the words of each topic and the documents. Not all topics in topic models are relevant or interpretable; some of the topics contained words that did not make sense and were, therefore, challenging to interpret. After labeling the topics, a topics network was created using pairwise cosine similarity and community detection algorithms (see Walter & Ophir, 2019). Nodes represent topics, and the edges are the co-occurrence of topics in the documents.

3.5 Results

The unit of analysis in topic models is the key terms or words present within the text data collection. For example, textual data visualizations (figure 3.2) show that the most frequent words in the Ansar1 forum include the terms 'said,' which occurred 29611 times, 'kill' occurred 18424 times, and 'attack' occurred 15,047 in the whole forum. For the Islamic Network forum, the terms 'quot' (quote) appeared 92,309, 'will' 40,196 times, and 'Allah' 39,474 times. The most frequent terms occurring within the Gawaher forum include 'link' 84,610 times and 'post' 60,468 times.



Figure 3.2: Ansar1, Gawaher and Islamic Network Top Frequent Terms

Ansar1 Dark Web Forum

The first topic model was conducted on the Ansar1 Dark Web Forum. The raw sample was N = 29,492 messages and 382 users. After removing duplicates and short text, the final sample was 12,744 messages used to analyze topic models and networks. The topic model for the Ansar Network consisted of 82 topics; 2 of the topics were not coherent enough and to be labeled and included in the network analysis (figure 3.3). Two communities, subgroups connected based on language co-occurrence than the rest of the group, were detected using the Walktrap clustering algorithm (Pons & Latapy, 2005). Walktrap clustering algorithm was selected because it captures the community structures of a network and computes efficiently and faster than other community detection algorithms (Pons & Latapy, 2005). The blue/on the left community focuses on religion and prayers, support for the mujahideen fighters, jihad, Al-Qaeda and Taliban, Muslims vs. the crusaders, the arrest of terrorists, and counterterrorism measures. The other community on the right/in red, topics ranged from mujahideen media, the death of U.S soldiers killed by the terrorists, death of terrorists and their leaders, Somalia government and Al-Shabaab, militants attacks in Afghanistan and Pakistan, suicide attacks and bombings, insurgency in Afghanistan.

A topic is outlined by its probability distribution over words, and the results also include the representative texts. Below I show the ten most probable words for a given topic and the full representative text of these topics. For instance, topic 23, which I labeled 'Mujahideen

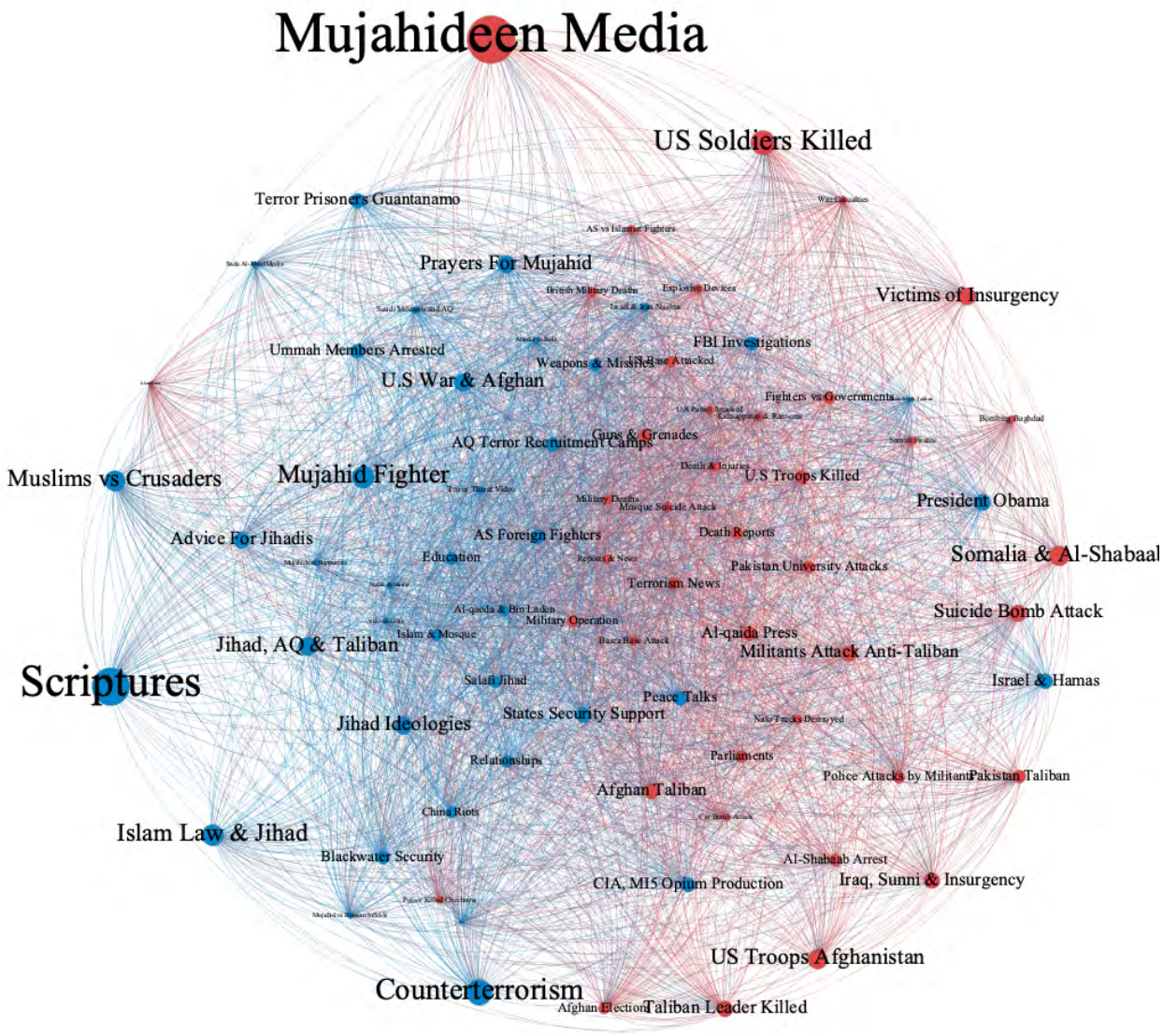


Figure 3.3: Ansar1 Dark Web Forum Topic Network. 80 Nodes represent topics, 3160 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted & undirected network.

Media' after reading the representative texts and words, has the following top 10 terms: Mujahideen, terrorist, kill, report, destroy, Islam, Afghanistan, Allah, Emir, puppet. The first complete text example was:

> "[1] Mujahideen of Islamic Emirate of Afghanistan military operations against the kafirs, munafiqs and the worshippers of Idols. This page is updated throughout the day as new operations are reported [...] the Mujahideen operations against the enemies of Islam terrorists in Afghanistan are reported to Theunjustmedia.com by the official Mujahideen of Islamic Emirate of Afghanistan spokesmen Qari Muhammad Yousuf and Zabeehullah Mujahid by e-mails." (Ansar1 Dark Web Forum).

Further reading of the rest representative texts suggests the forum members posted and discussed news from the Mujahideen media. Another example is topic 29, which I labeled 'Advice for Jihadis' has the following top 10 words: God, people, Muslim, religion, said, fight, one, victory, great, believe, faith, infidel. The representative text was:

> "[1] Al-Sahab is pleased to present you with a sound file titled "Advice for the People of Jihad" for Sheikh Mansur al-Shami. (Rajab, 1430) To our Mujahidin brothers greetings and prayers, It is mentioned in the Qur'an, and according to al-Sa'diy that if people don't fight for each other, land will be lost. The infidels will prevent the Muslims from worshipping God, spreading his religion, and they will establish their apostate idols everywhere. Because God is kind, he has allowed Muslims to go to Jihad to fight for the sake of their happiness. He has made this available to them for reasons they do know and reasons they don't know" (Ansar1 Dark Web Forum).

This topic incites and encourages supporters to join the fight for jihad, the Islamic religion, and their nation to prevent the West or non-Muslims from taking over their faith and country.

Gawaher Dark Web Forum

The following topic models were conducted on the Gawaher forum. First, I analyzed a sample (n = 52,588) based on the forum's most active year, 2007, for optimal analysis. Then, after removing duplicates and short text (less than 600), I remained with 19,236 texts for final analysis. Eighty-eight topics were analyzed and plotted; figure 3.4 illustrates the results of the 67 topics. Gawaher forum has two communities; the largest community in green topics are primarily about religion, Islamic scripture, the Bible vs. Quran, converting to Islam, and non-Muslims. When reading the full documents, there were some philosophical discussions about the existence of God and evolution, and relationships. Topics within the smaller community in red were about the U.S military and the U.S in Iraq, terrorists, Israel, Palestine, and the Hamas, the Taliban, and the imprisonment and torture of terrorists by the U.S. A few of the topics not included in the network mainly were posts from the forum admin reminding the members about posting rules and that posting links was forbidden until one had submitted 50 posts.

The most prevalent topic in the Gawaher topic network is topic 80, labeled "U.S Military and Admin." The top ten words include official, new, president, unit, report, American, Washington, military, said, intelligence. The first representative text was about the U.S admin:

> [1] "The Zionists are controlling the US regime and manipulating the USA and they are using the US regime to terrorize and invade Muslim countries and terrorize and massacre innocent and defenseless Muslims for the brutal and barbaric Zionist cause: Zionists in the Bush Administration: Paul Dundes Wolfowitz - Deputy Secretary, Department of Defense Richard Perle - Assistant Secretary of Defense for International Security Policy. Ari Fleischer - White House Press Secretary Josh Bolten - Deputy Chief of Staff ...." (Gawaher Dark Web Forum).

Appendix GUETLER01

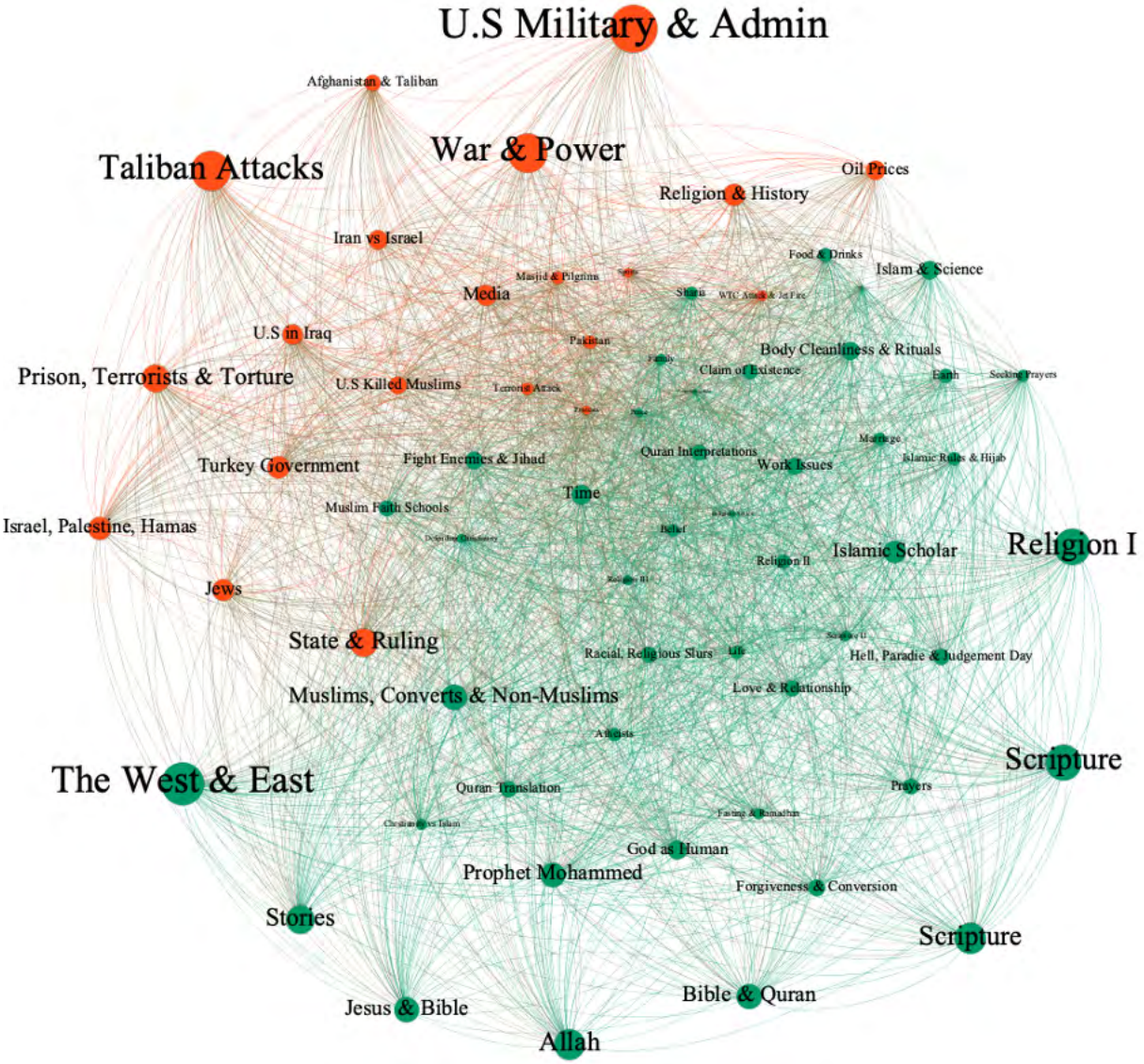


Figure 3.4: Gawaher Dark Web Forum Topic Network. Topic network has 67 Nodes representing topics, 2211 Edges represent the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network.

In topic 87, 'Taliban Attacks,' the top words include: said, kill, soldier, force, report, police, attack, bomb, military, Baghdad. The first representative text is a news report from the Associated Press:

> [1] "Taliban Torture, Kill 5 Afghan Police Associated Press — November 19, 2007 KANDAHAR, Afghanistan - Taliban militants tortured five abducted policemen in southern Afghanistan and then hung their mutilated bodies from trees in a warning to villagers against working with the government, officials said Nov. 18. The discovery of the bodies came as officials said that recent violence and clashes had left at least 63 other people dead across Afghanistan. The officers had been abducted two months ago from their checkpoint in southern Uruzgan province, said Juma Gul Himat, the provincial police chief. [...] Elsewhere, 23 Taliban militants were killed during a U.S.-led coalition operation on Thursday aimed at disrupting a weapons transfer in southern Afghanistan, the coalition said. Copyright 2007 Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed" (Gawaher Dark Web Forum).

Another interesting topic was Topic 9, "Forum Posting Rules" the probable top words include: forum, discussion, video, topic, can, site, read, please, may, website, thread, member. The first representative text:

> [1] "When starting a topic, please choose a title that describes it as best as you can. Non-descriptive titles confuse others, while good descriptive titles invite more members to come into your topic and share by reading or replying to you. Descriptive titles are beneficial and important in many ways: it makes searching and finding info much easier it invites more members to click your topic to read and reply choosing good descriptive titles allows members to experience 'informative clicking' titles help search engines index our topics and return more accurate results good titles bring more web searchers and so our daa'wa insha'Allah can

expand to more horizons Choosing a non-descriptive title for your topic is a violation of our forum rule 7. Kindly read our (you are not allowed to post links yet)"you can't post links until you reach 50 posts you are not allowed to post links yetgawaher(contact admin if its a beneficial link) to make sure you don't break them." (Gawaher Dark Web Forum)

This text exhibits how the forums are organized and how members use descriptive titles to attract volume and interactions.

Islamic Network Dark Web Forum

The Islamic Network forum had 91,874 messages. After removing duplicates and short text, the final sample was 22,787 messages resulting in 92 topic models. After labeling, 72 topics were applied to construct the topic network, which resulted in three topic clusters (figure 3.5). The 72 topics for the Islamic Network contributed to three major topic clusters. The topics in green were on the Guantanamo Bay and how prisoners are tortured and interrogated in that prison. Other topics were terrorism, US troops in Iraq, the death of soldiers from terrorist groups, the Taliban in Pakistan, and the relationship between Israel and Palestine. The topics in red/orange were a mix of electronics sold on the Dark Web, jihad, religion, the discussion on nationalism, people sharing stories, advertisements about the London talk by an Islamic scholar, investments and money, advertisements about selling fake documents, and IDs. Finally, the cluster in blue had topics on religion, Quran recitations and verses, Allah and God, Imam, and Islamic schools/scholars.

For instance, the most prevalent topic 14 is labeled 'Guantanamo,' top words include prison, Guantanamo, torture, detained, court, case, release, charge, terror, British. The representative text was:

[10] "The crimes at Abu Ghraib are part of a larger pattern of abuses against Muslim detainees around the world, Human Rights Watch said on the eve of the April 28 anniversary of the first pictures of U.S. soldiers brutalizing prisoners

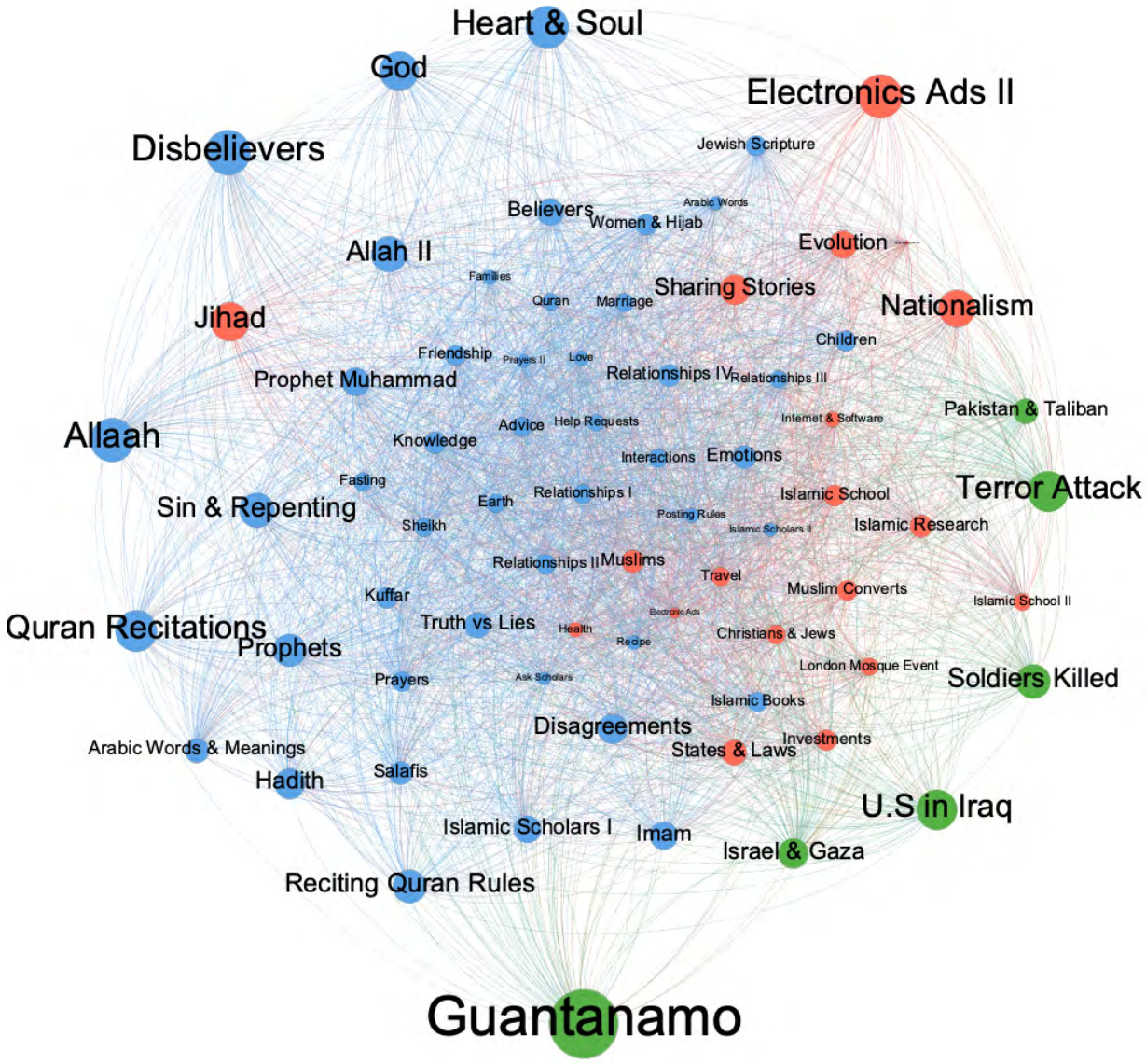


Figure 3.5: Islamic Network Dark Web Forum Topic Network. Topic network with 72 Nodes representing topics, 2556 Edges representing the co-occurrence of topics within the documents. The node size means the importance of each topic in the network. Color represents the community membership, weighted and undirected network.

at the Iraqi jail. Human Rights Watch released a summary (below) of evidence of U.S. abuse of detainees in Iraq, Afghanistan, and Guantanamo Bay, Cuba, as well as of the programs of secret CIA detention [. . .] Six Algerians held in Bosnia were transferred to U.S. officials in January 2002 (despite a Bosnian high court order to release them) and were sent to Guantanamo. SOURCE: Human Rights Watch" (Islamic Network Dark Web forum).

Another prevalent topic is topic 49, labeled 'Electronics Ads,' the most probable words for this topic include: soni, nokia, apple, htc, Samsung, Ericson, phone, notebook, pioneer, pc, tv, ipod, new, digit. The representative documents include:

[6] "We deal in the sales of consumer electronic. Main products include portable DVD player, TFT LCD monitor, Mobile phones and Our company have big quantity stock for mobile phone, they are 100% brand new set, wholesale price. . . . PRICE LIST'S BELLOW: NOKIA PHONES N95 8GB Latest/Black Edition.$350usd N73 Cost.....$120usd N75 Cost.....$150usd N76 Cost.....$170usd . . . ." (Islamic Network Dark Web forum).

Although not prevalent, topic 64, labeled 'Fake Documents Ad,' is interesting. The top words include passport, fake, false, sell, new, id, buy, identity, driver, sale, private. The representative text was:

[1] "Our team is a unique producer of quality fake documents. We offer only original high-quality fake passports, driver's licenses, stamps and other products for following countries: USA, Australia, Belgium, Brazil, Canada, Finland, France, Germany, Israel, Mexico, Netherlands, South Africa, Spain, United Kingdom. To get the additional information and place the order just visit our website: . . . (If in some technical reasons you are unable to visit our website we are always happy to answer your questions on email addresses mentioned below)" (Islamic Network Dark Web forum).

The topics were primarily about religion and a few on relationships, terror attacks, and electronics sales in three clusters. Most interesting is the illegal activities conducted on the Dark Web forums, including forums used by terrorist groups and their supporters.

In addition to the topic models, sentiments displayed by the words/topics were measured using the NRC lexicon (Mohammad & Turney, 2013). Sentiment analysis adds to topic models the sentiments of forum members based on the words and topics discussed. The NRC lexicon categorizes words into eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgusts) and polarity (negative and positive). Figure 3.6 shows the result of the forum sentiments, Ansar1 forum posts tend to display sentiments such as anger, fear, sadness, and trust compared to Gawaher and Islamic Network forums. The emotions highly displayed by Gawaher forum include disgust, joy, and trust, while Islamic Network sentiments are anticipation and trust. Overall, the sentiments of fear and trust was prevalent in the forums. Polarity scores (figure 3.7) indicate that the posts in the forums were mostly positive, words used in the forums were scored as positive for the Gawaher forum by (10.9%), Islamic network (10.3%) and Ansar1 (8.6%). Words associated with negative were fewer within the three forum Ansar1 by 8.3%, Gawaher (6.9%), and Islamic Network (6.5%). Overall, Gawaher is the most positive forum, has the highest average polarity ratio followed by Islamic Network while Ansar1 has the most negative sentiments.

Finally, results of sentiment analysis also display the top words contributing to the anger, anticipation, fear, and trust emotions (figure 3.8). For instance, in the Ansar1 forum, the top words associated with anger include attack, fighting, destroyed, enemy, killing. While the terms, words, attack, bad, death and evil were used differently in the Gawaher forum to portray anger. Words associated with fear are used differently in the forums, for instance, the words police, attack, government, military, and war conveyed different polarity within the forums. The words government, military, death, war appear in all three forums and are associated with the sentiment fear. The term, God, appears frequently within the Islamic Network and Gawaher forums and associated with the anticipation, fear, and trust emotions.

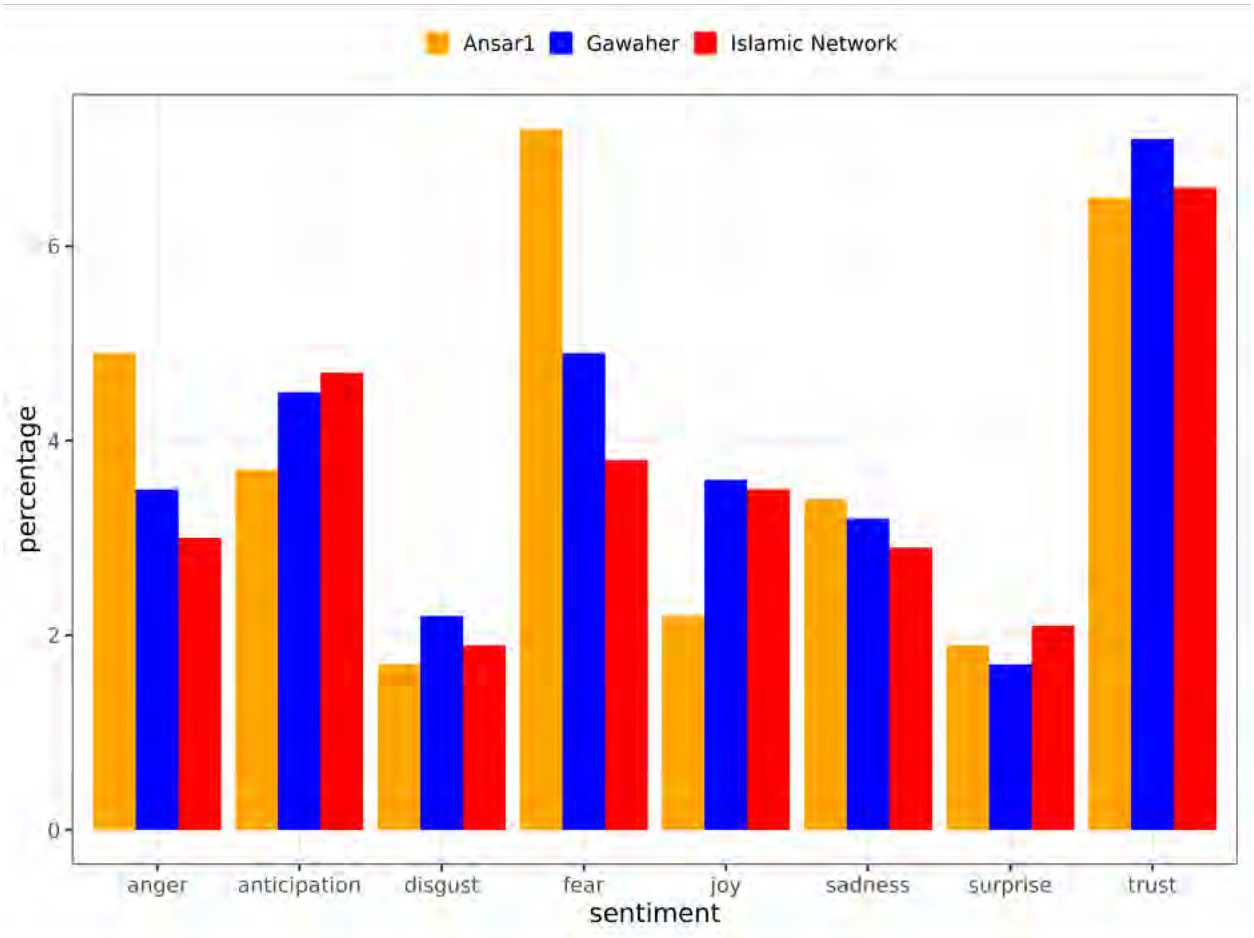


Figure 3.6: Word Sentiments of Dark Web Forums

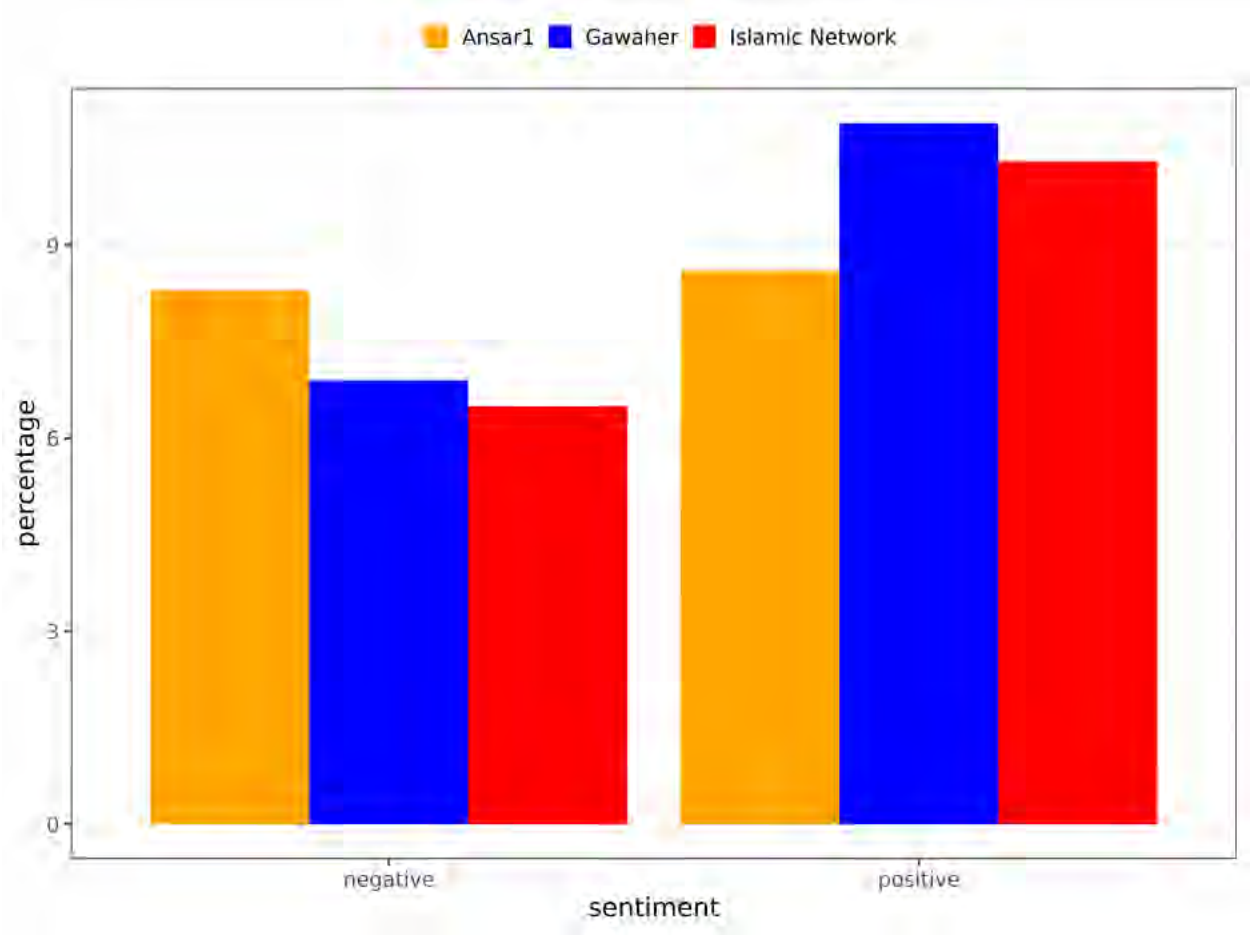


Figure 3.7: Polarity Scores of Dark Web Forums

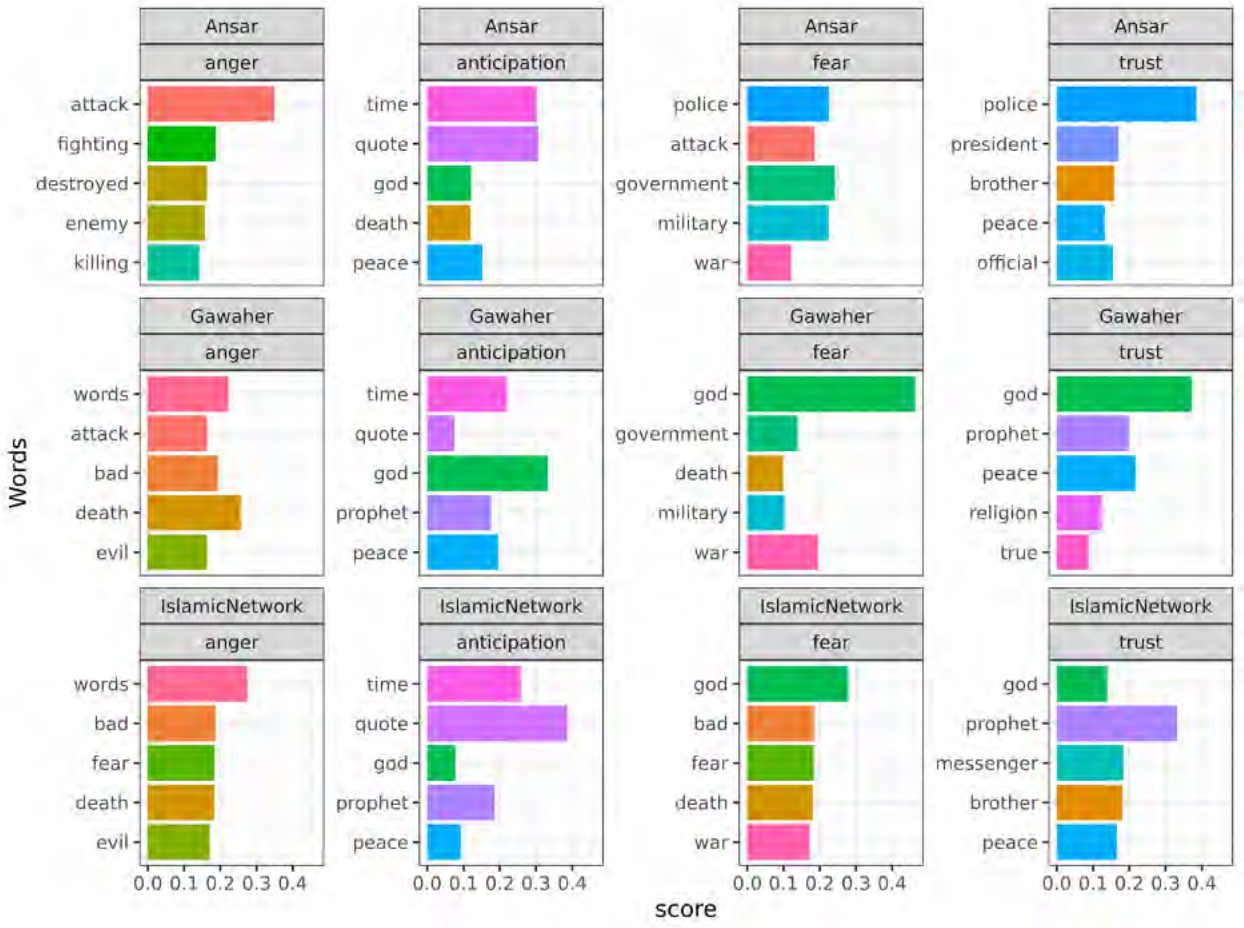


Figure 3.8: Dark Web Forums Top 5 Words and Emotions

By comparing the words and sentiments provides a further understanding of the topics, attitudes and opinions on religion and terrorism and how they differed between the forums.

3.6 Discussion and Conclusion

As Jihadists continue to use the Internet for communication and recruitment, topic models are valuable techniques that automatically detect and classify their online content. Topic model algorithms automatically discover themes and topics and how the themes are connected in a large and unstructured collection of documents based on the co-occurrence and frequency distributions of words (Blei, 2012). While the topic structure is hidden, the documents and words are observed (Blei, 2012; Blei et al., 2003). The high probability words in each topic distribution can be easily interpreted as a topic (Baumer et al., 2017). The unsupervised latent Dirichlet allocation (LDA) topic model does not require prior annotations or data labeling (Blei, 2012; Blei et al., 2003). Therefore, there is no need to manually label a data sample before identifying the topics. The researcher only has to specify the optimal topic numbers and label the output. LDA models assume that a topic is a distribution over vocabulary, i.e., a set of words, and a document is a mixture of topics. Topics can be inductively observed through topics modeling, network analysis, and community detection (Walter & Ophir, 2019). Community detection identifies groups of vertices/nodes, particularly topics, that are more densely connected than the rest of the network. Thus, topic models enable the analysis of text data that is too large to code by hand. Furthermore, topic models might highlight topics that a researcher using manual coding techniques may otherwise not have found (DiMaggio et al., 2013).

Terrorist groups and their supporters use the Internet to further their ideologies and goals posing new national security and counterterrorism challenges. To evade detection or disruption, the groups and their supporters have moved to the Dark Web (Malik, 2018; Weimann, 2016a). To show the utility of LDA topic models, I used three cases from Dark

Web forums of jihadist groups and their supporters, the Ansar1 (Ansar AlJihad) forum, the Gawaher forum, which is dedicated to discussions of Islam and the Islamic world, and the Islamic Network forum, which is also devoted to the theology and world events. Although the forum discussions pertaining to the Islamic world and Islam, some members of these forums sympathize with and support terrorist organizations. Therefore, participating in forum discussions is considered a form of jihad (Erez et al., 2011).

Using LDA, I identified coherent topics and clusters of the forums. As expected, the most prevalent topic in all forums was religion. Topics about religion included Islam, Christianity, and Judaism, discussion about Allah and God, believers and disbelievers, Quran recitations, religious rituals, and philosophical discussions about religion and science. These topics are expected, as the forums were dedicated to the Islamic faith and issues concerning Muslims. The topic of terrorism was prevalent in all forums, mainly within the Ansar1 forum. Here the discussion was about terrorists and attacks, support for the Mujahideen fighters, jihad, Al-Qaeda, Taliban, the U.S in Iraq, Afghanistan, and the arrests and torture of terrorists and foreign fighters. One could argue the discussion or posts about terror groups were created by the terrorist groups and their media centers and supporters, also known as 'media mujahideen' (Malik, 2018). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). A few of the discussions were on relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards. Topics on selling fake IDs and documents indicate the presence of criminals in terrorist networks and possibly a partnership between criminals and terrorists, also known as a crime-terror nexus. Not all these topics would have been easily identified via manual hand-coding, especially since a predefined set of themes would have been set as guidance for coding.

By clustering the topics into communities, one can highlight the frequency of language use from the corpus used by each community, in this case, the extent of discussions based on terrorism and/or Jihadism. For instance, in the Ansar1 forum, 51% of the corpus was

associated with the community on the left, in blue. In contrast, 49% of language used in the forum corpus is likely to be used by the community on the right, in red. Similarly, for the Gawaher forum, the topics in the green community were most likely to be associated with 69% of the language in the corpus and 31% for the topics in the red community. In the Islamic Network forum, the topics in blue are most likely to be associated with 65.3% of the words in the corpus and 26.4% for the topics in red, and 8.33% for the topics in the green community. In addition to identifying the topics, and the prevalence of language use, one can also classify extremist content from the Dark Web. Results suggest that from the 82 topics identified in the Ansar1 forums, there is an equal mixture of religious and terrorism content. Most of the Gawaher forum's 67 topics were about religion (61%) and terrorism content (31%). The Islamic Network forum's 72 topics resulted in three classifications. The prevalent topic is religion (65.3%), a mixture of religion, legal and illegal activities (26.4%), and terrorism (8.33%). The Ansar1 forum was mainly dedicated to discussing or sharing terrorist content as opposed to the other two forums, which were especially about Islam and Muslims. It should be noted that the Ansar1 forum is the English forum of the Arabic Ansar AlJihad Network, which is an invitation-only Jihadist forum popular with western Jihadists (J. Scanlon & Gerber, 2014). Additionally, as mentioned elsewhere, topics on religion can also be classified as Jihadist discourse, using religion to justify their ideology.

LDA topic modeling has several advantages over the manual annotation of text data, especially when analyzing larger corpora. The analysis of text data has predominantly been based on manual content analysis, where researchers create a set of themes based on research questions and objectives, create a coding sheet, and read the text and code by hand. As DiMaggio et al. (2013) note, manual coding is limited when the text to be analyzed is large, and "the more analytically interesting are the research questions, the harder it is to achieve acceptable levels of intercoder reliability; and c) the approach presumes that the researcher knows what is worth finding in the texts before having analyzed them" (DiMaggio et al., 2013, p. 577). However, a manual approach can help analyze smaller samples. When analyzing

big data generated by Internet users, automatic computational techniques are more effective, faster, and reliable. Topic models quickly discover hidden topics, thus providing real-time text analysis. The topics are inductively uncovered and require no previous knowledge of the content from the large-scale text data. Additionally, topic models offer insights into extensive unstructured text data instead of the traditional structured data such as surveys and interviews. The keywords discovered from the topic models can further be utilized to organize and search for content in other text data. For instance, the keywords derived from the three Dark Web forums can be used to analyze or search other Dark Web forums or websites for similar topics.

Compared to manual annotation, topic models require less time and money and reduce researcher biases. Topic modeling is data-driven and thus requires very minimal human supervision. The only work for the researcher is to input a topic number and interpret and label the output. Hence, topic models illustrate that textual data can also be analyzed quantitatively and mitigate researcher bias. Additionally, compared to other text classification algorithms of unlabeled data, such as k-means clustering analysis, where documents can be classified into only one cluster, topic models are mixture models where each document is assigned a probability of belonging to a topic; therefore, a document can belong in different topics.

However, some limitations abound when using topic modeling. First, depending on the corpus size to be analyzed, topic modeling can be computing-intensive. Depending on the computer's power, an analysis can run for two days or weeks. Secondly, language is ambiguous, and not all topics can be labeled and interpreted into a recognizable theme. For instance, despite considering the full text and context, some of the topics in the Gawaher and the Islamic Network forums were challenging to interpret and label, were not meaningful, and thus omitted from the analysis. In addition, unrelated topics may be attributed to the stylistic conventions of text data and the algorithm (Walter & Ophir, 2019). Finally, it is important to note while topic models are data-driven, they assist but do not replace human

interpretation (Baumer et al., 2017).

Since the focus of this example case study was to exemplify how topic modeling, specifically LDA, can be used to discover topics from the Dark Web forums, I only analyzed the messages posted by members and did not include the members or time, the year they posted or were active on the forums. Future directions may include the analysis of the members and their contributions to the topics. For instance, an analysis of the topic models would consist of the members and using structural topic modeling, author-topic models, or other algorithms to classify the posts and members as radical or non-radical, based on the posts, time spent on the forum, engagement with others, i.e., responding to all threads, etc. Another future study may analyze the social network structure, including the key members, and detect the communities to highlight how the members behave and interact in the Dark Web forums. It is important to note that this work applied topic models in classifying and detecting Jihadist online content; I did not seek to imply that the forum members are terrorists. Instead, I was interested in detecting topics posted on the forums. Lastly, while the application of sentiment analysis to measure polarity was not the scope of this study, sentiment analysis of the Dark Web forums highlights the opinions and attitudes of forum users toward the major topics of religion and terrorism. Overall, Ansar1 sentiments tend to be negative (i.e., in support of terrorism) as compared to Gawaher and Islamic Network where the sentiments were positive i.e., in favor or support of the Islamic religion. As such, analyzing the sentiments in addition to the topics provides an overall idea about the perceptions on religion and terrorism. A future study may analyze the sentiments using the key words identified by the topics, to create a lexicon that is terrorism specific.

From the topic classification of jihadist Dark Web forums, we can conclude that topic modeling and sentiment analysis can be applied automatically to analyze large-scale text from extremist content online. The detection of extremist content online using topic models is significant for analyzing users' beliefs and countering and/or disrupting their online narrative. Additionally, sociological study of social interactions and theories such as social learning and

social control theories can be applied to explain online radical content and interactions of like-minded individuals based on their topics. Social bond theory (Hirschi, 1969) holds that individuals are less likely to commit a crime if they have stronger social bonds to family, peers, and legitimate institutions. On the other hand, social learning theory (Akers et al., 1979; Akers et al., 2014) presumes that individuals' differences in criminal behaviors stem from learning through association or interactions with others. Therefore, as social behavior is learned, one would expect individuals participating in the forums and interacting with those who post about terrorism are more likely to change their beliefs and become radicalized into violent extremism. Additionally, individuals posting on the forums may seek to manage their impressions (Goffman, 1959) by selecting what they post in order to be perceived as either in support or against terrorism. Although these theories were formulated to explain offline interactions and behavior, a future study would apply the theories in explaining whether online interactions and posting behavior are related to violent extremism.

Chapter 4

Detecting Virtual Communities and Networks of Jihadists Dark Web Forums: A Social Network Analysis

4.1 Introduction

Counterterrorism experts and social media organizations are disrupting the online presence of terrorist groups. Social media platforms such as Facebook and Twitter and forums such as Gab and Stormfront used by extremists and terrorists are monitored and suspended (Behr et al., 2013; J. M. Berger, 2016; Caló & Hartley, 2019; Zelin & Fellow, 2013). As a result, to avoid detection and disruption, the groups have migrated to encrypted platforms such as Telegram and the Dark Web (Malik, 2018; Weimann, 2016a) that are difficult to regulate and disrupt. Most Internet users only access the surface web, also known as the Internet, easily accessible through browsers such as Google and Yahoo. A subsection of the Deep Web, the Dark Web is a hidden part of the Internet that can only be accessed via special browsers such as TOR (Malik, 2018; Mirea et al., 2019; Monk et al., 2018; Weimann, 2016b). The Dark Web or Dark Net offers anonymity, making it attractive for individuals interested in privacy

Appendix GUETLER01

and avoiding censorship and criminals, including terrorist groups. Terrorists use the surface web for communication, spreading propaganda, recruitment, radicalization, and financing. However, with the Dark Web, they can do the same while hiding from law enforcement, avoiding disruption, and storing their propaganda materials (Malik, 2018; Weimann, 2016a).

The Internet's propensity for anonymity, minimal censorship, expanded reach, and influence on global audiences, has transformed the structure and dynamics of violent extremism by changing how terrorists operate and interact. The Internet has created opportunities for self-radicalization, connection, and interaction of like-minded people globally while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence (Behr et al., 2013; Bloom et al., 2019; Conway, 2017; Edwards & Gribbon, 2013; Sageman, 2011). Prior to the growth of the Internet, most of the terrorist networks resulted from face-to-face interactions and strong bonds amongst friends and family (Sageman, 2004, 2011). However, the Internet, especially forums or chat rooms and not passive websites, is crucial in creating social bonds, trust, indoctrination, and eventually radicalization into violent extremism. Active online discussions about ideologies and opinions with friends or strangers have led to the sense of belonging within the 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018). As such, virtual communities have the potential of exposing someone to radical content, thus changing someone's mind rather than just reading impersonal stories. Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community.

Generally, research on terrorists and the Internet has focused on the role of the internet in radicalization and recruitment, specifically through social media accounts such as Twitter, encrypted platforms such as Telegram, websites, and forums such as Stormfront and Gab (Bloom et al., 2019; Bodine-Baron et al., 2016; Bowman-Grieve, 2009; Caló & Hartley, 2019; Gaudette et al., 2020; Koehler, 2014; Piazza & Guler, 2019; Rudner, 2017; Winter et al., 2020). Yet, little is known about terrorist use of the Dark Web and the network structure of the forums used by terrorist groups and their supporters. In this exploratory

study, I use social network analysis to examine the social networks of two international jihadist groups Dark Web forums. I construct undirected networks to analyze the network structure, the interactions between members within the forums, the essential members, and virtual communities. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? And what are the posting activities of key members in the forums?

The paper is organized as follows; the first section describes the technological structure of the Internet, specifically the Dark Web. The second section discusses related work in explaining virtual communities and terrorists' online content, terror on the Dark Web, and social networks of terrorist groups. The last sections explore the research design, methods, data, and results. Finally, the paper concludes with a discussion about the Dark Web forums and their network structures.

4.2 Dark Web's Technological Structure

The Internet is divided into layers, surface, deep and dark. Most people use the surface web or the Internet, which is reachable, indexed, and accessed via browsers such a Google, Bing, and Yahoo. Information from the surface web is visible to all those who want to access it, and there are no restrictions. The Deep Web is the part of the Internet that is hidden, not indexed, and therefore not accessible via search engines. Access to Deep Web content requires sign-in passwords and is behind paywalls. For instance, medical records, membership websites, social media platforms, online banking sites are part of the Deep Web. The Deep Web is approximately 400 to 500 times larger than the surface web, holding 400 times more content and 7,500 terabytes of information than 19 terabytes on the surface web (Chertoff, 2017; Malik, 2018). The Dark Web, or Darknet hereby referred to as Dark Web, exists within the Deep Web. The Dark Web is even harder to access, is unregulated, and

Appendix GUETLER01

contains approximately 0.01% of the content as opposed to the Deep Web's estimated 90% of content hosted on the internet (Malik, 2018; Weimann, 2016a).

The Deep Web is accessed for anonymity, usually for legal purposes, and access is granted via passwords, encryption, or specific software. The Dark Web is a layer deeper in the Deep Web, and it is intentionally hidden, accessed for anonymity, and sometimes illegal activities. Access to the Dark Web is granted through free and unique software programs supporting encrypted channels. Connections and networks can only be accessed by specific software, configurations, or authorizations such as the Tor (The Onion Router) or I2P (Invisible Internet Project). Dark Web sites provide anonymity and prevent monitoring for Internet users through encryption, which is "the process of encoding information using mathematical algorithms, ensuring that communication is obfuscated to protect against unauthorized access" (Malik, 2018, p. 13). For instance, the commonly used Dark Web browser, the Tor browser (The Onion Router), uses multilayered encryption that relays traffic and encrypts it three times. It is also used to block trackers and ads, defend against surveillance and fingerprinting. The idea of onion routing is to route traffic through multiple servers and encrypt all the way (https://www.torproject.org/). Based on multi-layer encryption and routing, users access the Dark Web sites securely, without revealing their location or their identity and anonymizing visited websites and their users (Chertoff, 2017; Grabosky, 2016; Malik, 2018). Ensuring the identity of users and their computers are anonymous in case they are being monitored.

The Dark Web is accessed and used for both legitimate and illegitimate reasons. Legitimate users of the Dark Web include journalists, whistle blowers, and activists, web users seeking to protect their privacy, or users who want to access hidden sites and services that may have been restricted or censored (Chertoff, 2017; Finklea, 2017; Mirea et al., 2019; Monk et al., 2018). Aside from the anonymity users benefit from while on the Dark Web, the site is useful for cybercriminals. However, "the darknet might not be intrinsically criminogenic – it does not naturally increase criminal activities; rather it might be just another tool that

is used by some individuals to carry out illicit activities" (Mirea et al., 2019, p. 104). Dark Web is used by criminals to sell and buy illegal goods. Marketplaces such as the Silk Road, AlphaBay, DreamMarket are used for advertisement and trading of illicit goods such as drugs, weapons, stolen identities, and credit card details, child pornography, hackers, and hacking software for hire (Finklea, 2017; Gupta et al., 2021; Mirea et al., 2019). While law enforcement has shut down the marketplaces, new versions are recreated or moved to forums within the Dark Web where they communicate and perform their illegal transactions (Berton, 2015; Gupta et al., 2021), posing challenges for the detection and swift removal of criminals on the Dark Web.

4.3 Related Work

4.3.1 Virtual Communities and Terrorists Online Content

Most of the studies on terrorists' use and/or misuse of the Internet have identified how terrorists use the internet. The fundamental benefit is to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking, information dissemination, and secure communication (Conway, 2017; Gill et al., 2017; Hoffman, 2006a). As a result, there has been a legitimate concern about the role of the Internet in radicalization. According to McCauley and Moskalenko, radicalization "means change in beliefs, feelings, and behaviors in directions that increasingly justify intergroup violence and demand sacrifice in defense of the ingroup" (McCauley & Moskalenko, 2008, p. 416). Bartlett and Miller (2012) distinguish between violent and non-violent radicalization. In violent radicalization, the end goal is violence where individuals undertake, aid, or abet terrorist activity. In non-violent radicalization, individuals hold radical views but do not undertake, aid, or abet terrorist activity. Thus, radicalization is a social process in which extremist ideologies and beliefs are developed, stemming from exposure to radical content via social ties such as family and peer relationships offline and online via websites, forums, and social me-

Appendix GUETLER01

dia. While radicalization requires face-to-face interaction and trust, social interactions and networks remain influential in the process of radicalization and violent extremism. However, in the absence of broader social ties to groups or terrorist networks, online communications expose radical content and facilitate violent or non-violent radicalization (Bloom & Daymon, 2018; Holt & Steinmetz, 2020). Nearly all lone-wolf terrorists have been radicalized online (Brynielsson et al., 2013).

Forced out of operating and communicating via public spaces such as the mosques and prisons, terrorists and violent extremists have moved their activities to private areas and personal computers for security, safety, accessibility, and anonymity (Edwards & Gribbon, 2013; Sageman, 2011). Interactive platforms such as chat rooms, message boards, and social media allow interaction amongst like-minded individuals to develop, hence binding, bringing together individuals, fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst participants (Bloom & Daymon, 2018; Gaudette et al., 2020; Piazza & Guler, 2019). Interactive platforms have the same influence in the process of radicalization as face-to-face interactions. As such, active online discussions about ideologies and opinions with friends or strangers and the sense of belonging results in a 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018), in the case of jihadists, a 'virtual Muslim community (ummah)' (Erez et al., 2011). Within the virtual communities, participants post and share their views and knowledge, offer support and justification of the use of violence, and validate. Recruitment through the websites and forums is facilitated through community and propaganda efforts (Burris et al., 2000). As a result, a virtual community network is created as extremists and terrorist groups bond and connect with their audiences using these online communication systems, forming a sense of community and belongingness. The virtual networks comprise unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals.

Once in the virtual community, users' language and social interactions change over time, mirroring those within their online community. As time goes by, users in these platforms

have roles including producing and aggregating content, translating and curating content across platforms, or passively consuming the information (Shehabat & Mitew, 2018). For instance, in Telegram, ISIS's outlets, there are three kinds of active Telegram users (i) those seeking information, (ii) those who want to engage more fully with the terrorist group, and (iii) propagandists in search of both (Bloom et al., 2019). As participants feel comfortable and welcomed in these platforms, skeptics and lurkers are encouraged to share their beliefs and become more vocal. In addition, since participation in these forums is an active process, users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). The more time individuals spend online, the more likely they may be radicalized into violent or radical social networks.

To make their platforms attractive, persuasive, and addictive, terrorist groups engage in "soft power" to influence their target audience using persuasive language and arguments without coercion (Bloom et al., 2019). For instance, ISIS' telegram platform has a different selection of materials such as emojis, stickers, and memes, ISIS's pop culture, which are shared and spread in the forum hence sustaining user engagement and mitigating boredom (Bloom & Daymon, 2018; Bloom et al., 2019). The vulnerable, loners, individuals without positive social ties and bonds, are susceptible to these recruitment tactics and become targets of online recruiters.

The online content and strategy are conducive to radicalization, as users seeking justification for joining or intending to join terrorist groups are likely to be indoctrinated by these messages. Additionally, terrorists such as ISIS online recruitment process follows a specific trajectory: (i) discovery of a potential recruit, or a recruit discovers ISIS, (ii) supporters form a community around potential recruits, (iii) potential recruits are encouraged to isolate and cut ties from their social bonds (families, friends), (iv) move to private or encrypted messaging spaces for conversations, and lastly, identify and encourage action to be taken by the recruit, whether to travel and join IS or carry out an attack at home (J. M. Berger, 2016; J. M. Berger & Strathearn, 2013). While the content attracts sympathizers and would-be

extremists to frequent the platforms, recruiters monitor these virtual communities to find potential recruits.

Although the Internet has provided an eager and ready-made audience and acts as a "larger megaphone" (Edwards & Gribbon, 2013) for violent extremists than traditional modes of communication, some scholars are skeptical about the influence of online platforms and communities. Holt and colleagues argue "close to 99%" of individuals exposed to radical messages and content never engage in violent extremism (Holt et al., 2017a). Similarly, Benson (2014) notes the Internet appears to be one influence among many other factors in relation to violent extremism. He claims the Internet is more advantageous to state security forces than terrorist groups. Individuals participating or frequenting these sites are only interested in conversations and not in political violence (Awan, 2017). On the contrary, the Internet is used by violent extremists groups to reinforce the radicalization process as individuals in these online communities experience an echo chamber, where only violent extremists views are expressed and reinforced (Behr et al., 2013; Piazza & Guler, 2019). For instance, as Gaudette et al. (2020) note about the Far-Right, due to the access to extreme right-wing content, exposure, and networks of like-minded individuals online, the internet played a role in facilitating their process of radicalization. Yet, as extremists and terrorists use the Internet and form virtual cliques and communities, terror networks are still a combination of online and offline elements which mutually influence each other (Sageman, 2011).

4.3.2 Terror on the Dark Web

As the terrorist platforms and materials on the surface web are monitored and removed, the groups migrate to the Dark Web, where anonymity is provided if not guaranteed. Private virtual spaces are now encrypted. Administrators must verify members, membership is by referral only, and links and instructions for accessing the Dark Web forums are shared on the social media accounts or encrypted platforms (Berton, 2015; Malik, 2018). The Dark Web

sites provide a 'safe space' for the groups and their supporters to interact, communicate and access the information without being monitored. Media jihad is the production, distribution, promotion, collection, and redistribution of jihadist content as an obligation to incite violence against the enemies (Erez et al., 2011). As such, the 'media mujahideen' (Malik, 2018) amplify terrorists' online messaging campaign by sharing and reposting official content of the groups across social media platforms and the Dark Web forums. At times, "different terrorist groups compete for control of these forums: Shumukh al-Islam, for example, is a forum which oscillates between ISIL and al-Qaeda supporters" (Berton, 2015, p. 2).

Terrorists and their supporters use the Dark Web for communication via forums and emails, where they discuss religion and jihadist ideologies (Abbasi & Chen, 2007; J. R. Scanlon & Gerber, 2015). Forums are the center of online jihadist activism; terrorists control them, and members can ask questions or contribute to discussion topics and private communications (Erez et al., 2011). The forums encourage radicalization as well as promote "a "self-starter" type of terrorism. This is a strategy endorsed by IS both online and offline and motivates vulnerable individuals to commit violence in the organization's name in an attempt to "crowdsource" terrorism" (Malik, 2018, p. 20). The jihadist discourse online focuses on religious and ideological indoctrination, framed by references to the Quran and the early history of Islam (Erez et al., 2011). One can find texts such as the 'Terrorists' Handbook' and 'Explosives Guide' that can be purchased via the Dark Web marketplaces specializing in selling weapons (Berton, 2015). Yet, with the inaccessible nature of the Dark Web and with most people only accessing the surface web, mass recruitment rarely takes place on the Dark Web. Instead, the groups tend to draw interested supporters from the surface web to the more secure Dark Web platform for further interaction, indoctrination, planning, and launching terror attacks (Malik, 2018).

Although terrorists use the surface web to store and share their information, internet companies and social media platforms often remove the materials. Hence, the Dark Web is used as a repository to store its material. For instance, following the November 2015

attacks in Paris, ISIS used the Dark Web to create a new propaganda hub where they shared news and propaganda to their supporters and also to protect its information from being hacked/removed from the surface web by the Anonymous hacktivists (Malik, 2018; Weimann, 2016a). Furthermore, the groups also use Dark Web for financing through cryptocurrencies, where donations are paid via bitcoins that cannot be traced by law enforcement (Berton, 2015). Therefore, the Dark Web is an ideal ecosystem for terrorist groups; compared to the surface web, they can avoid censorship, discuss, communicate with their supporters, protect their identity and supporters, store their content without removal and fundraise anonymously.

4.3.3 Covert Social Networks

The application of social network analysis (SNA) in the study of terrorism is limited (Koschade, 2006; Perliger & Pedahzur, 2011). Social network analysis (SNA) takes relations as the fundamental unit of social analysis rather than individuals, groups, attributes, or categories (Borgatti et al., 2013; Borgatti & Halgin, 2014). Relations measured are the ties and interactions amongst individuals such as family, friendships, religion, politics, finance, in the case of terrorists' ideological links that form a social network. A social network is represented by nodes (vertices, actors, individuals, or events) connected by complex relationships creating a network. Edges are ties, links, or connections between the nodes, and attributes (attitude, opinions, behavior, demographics) are also fundamental for social network analysis (Borgatti & Halgin, 2014; Scott, 2000).

SNA is suitable for analyzing terrorist networks based on the relationships as the attributes (e.g., real names, gender, age) are often difficult to obtain for covert networks (Basu, 2014). Raab and Milward (2003) introduced the concept of 'dark networks' to describe networks of actors and organizations that engage in covert and illegal activities. Several network studies have been applied to these dark networks, including, organized crime and gangs (Bienenstock & Salwen, 2015; Bright et al., 2012; Duijn et al., 2014), terrorism (Almquist

& Bagozzi, 2019; Asal & Rethemeyer, 2008; Bienenstock & Salwen, 2015; Campedelli et al., 2019; Carley et al., 2001; Koschade, 2006; Krebs, 2002; Milla et al., 2020; Perliger & Pedahzur, 2011; Sageman, 2004). Yet only a few have focused on Dark Web forums of terrorist and hacker groups (Monk et al., 2018; Pete et al., 2020; Phillips et al., 2015).

Research has shown that the covert networks of terrorist groups tend to be decentralized (Raab & Milward, 2003; Sageman, 2011) and connected by actors/nodes who function as 'brokers' (Everton, 2008; Raab & Milward, 2003). For instance, Sageman (2004) study of the global Salafi jihad found the network relatively centralized to decentralized with four clusters. The network had more connected nodes (hubs) significant for the terrorist network. It was a small-world network, where influential nodes connected with more people who then influenced others to join the network. Unlike hierarchical networks that can be destroyed by removing their leadership, the small-world network resists fragmentation because of its dense interconnectivity. Thus, the network may be inefficient by identifying and taking out central actors (nodes). However, only small-world networks with targeted attacks at their hub (actors with a high number of ties) are efficient in disrupting the network (Carley et al., 2001; Sageman, 2004).

Network disruption occurs when a network "cannot efficiently diffuse information, goods and knowledge" (Duijn et al., 2014, p. 2). Network destabilization is often based on the critical player approach that seeks to identify the actors with the most considerable impact on the network and remove them. Some actors are central to the network, key influencers, brokers, leaders (Koschade, 2006; Perliger & Pedahzur, 2011). Detecting actors in the network who are critical to the continuous operation of the group is essential as it highlights the power division and the motives within the network. Such actors have roles imperative for the success of the network. For example, Koschade (2006) social network study of the Jemaah Islamiyah cell operating in Bali found that actors most central and active within the network had the ability to access others and had the greatest control over the flow of information. In contrast, actors with the lowest centrality scores were those in the periphery,

isolated, and only called upon for assistance.

Similarly,Milla et al. (2020) study of Indonesian terrorist groups found that operational leaders – those who manage and organize groups, recruit members - are more central than ideological leaders – the preachers. Thus, ideological leaders in operational networks may have less influence than operational leaders. Key members in the network are responsible for coordinating the group's activities, recruitment, and manipulating information flow. In such networks, the removal of operational leaders may disrupt the network. Yet this 'whack-a-mole' strategy (Everton, 2008) may lead to new leaders, and new networks will emerge. Information will continue to flow as covert networks are quick to adapt. Since "covert networks have the ability to heal themselves in the event of loss of a node" (Koschade, 2006, p. 25). Criminal networks become stronger after targeted attacks (Duijn et al., 2014). Similarly, covert networks of Dark Web forums are often decentralized, and the identification and removal of key players usually lead to the networks becoming more decentralized, thus harder to disrupt (Pete et al., 2020). Although removing critical individuals disrupts a network, one of the latent consequences of network destabilization is that decentralized networks tend to reemerge, thus posing challenges in an ultimate and successful removal of covert networks.

Uncovering cohesive subgroups, cliques, groups, or communities is significant for understanding members' network structure and functions. Cliques or communities are formed when each actor in the subgroup is connected to all the other actors (tightly connected) and few connections to the actors outside the subgroup (Newman, 2006; Papadopoulos et al., 2012). Meaning that some actors interact more strongly with members of their communities than they would with members of other subgroups (Fortunato & Hric, 2016). Identifying the subgroups, as Perliger and Pedahzur (2011) note:

> "Allows us to detect different functions of the network (founders, collaborators, passersby), network recruitments paths, operational characteristics (for example, in some networks there is a clear distinction between a suicide bombers subgroup

> and other subgroups, while in others, the suicide bombers are isolated actors in the periphery of the network) and patterns of flow of information. . . . By looking at the attributes of the subgroups, we can evaluate ideological homogeneity and level of solidarity within the network and how this influences the activities and development of the terrorist networks" (Perliger & Pedahzur, 2011, p. 13).

Subgroups or network divisions within the Dark Web forums can be identified and classified based on the related topics and interests of the forum members (Pete et al., 2020). Thus, removing key members in these forums or members who act as brokers between subgroups may disrupt the information flow in that subgroup and the whole network.

Identification of ties, weak ties (relationships with acquaintances), and strong ties (family and friends (Granovetter, 1973) is a significant component for network analyses. The tie strength may be based on the amount of time spent together between a pair of members (Krebs, 2002). Weak ties within terrorist cells are significant for survivability when one of the cells is exposed. On the other hand, strong ties facilitate solidarity and commitment (Perliger & Pedahzur, 2011; Sageman, 2011). Trust is significant within a terrorist network (Krebs, 2002; Milla et al., 2020). Covert or criminal networks are often larger, incomplete, and have fuzzy boundaries as clandestine network borders are unclear; individuals may belong to different groups and networks. The covert networks are dynamic, as the network is constantly changing, with interactions between individuals strengthening or weakening over time (Koschade, 2006). Indeed, as Krebs (2002) noted when collecting data to map the 9/11 airline hijackers, he encountered the problems of incompleteness due to missing nodes and links, difficulty in deciding who to include and who to exclude in the network, and the networks are not static; they are constantly changing.

In sum, research on terrorist use of the Internet, both surface and Dark web, suggest that the groups utilize the platforms to spread propaganda, communicate, recruit, radicalize and finance. Although recruitment often takes place on the surface web, the groups lure potential recruits to the Dark Web for further indoctrination, safer communication, and

planning of attacks. The forums provide a safe space, a sense of belongingness for a virtual community of like-minded individuals to interact. Members discuss, share, support, and justify the jihadist ideology within the virtual communities. Terrorists face the threat of disruption from law enforcement and counterterrorism by removing and disrupting social media platforms, websites and detecting and eliminating critical leaders in their online social networks. However, little is known about the Dark Web forums used by terrorist groups and their supporters. There still exists less research applying social network analysis in the study of terrorist groups, especially the groups and their supporters in virtual covert communities. In this study, I aim to add to the literature knowledge about the network structure, behaviors, and significant members in the Dark Web forums of jihadist groups. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? What are the posting activities of key members in the forums?

4.4 Research Design

This study applies social network analysis (SNA) to explore the network of jihadists and their supporters on the Dark Web. SNA uncovers patterns and behaviors of people's interactions. The aim is to understand the social network structure of the forum users, their posting behavior, and communities. SNA requires relational data to measure the relationships, links, or ties (edges) between individuals or groups (nodes or vertices) (Borgatti et al., 2013; Borgatti & Halgin, 2014). Thus, a social network consists of individuals or groups (nodes) connected (ties) by kinship, friendship, common interest, financial interest, and beliefs. Network analysis is significant for this study as it complements conventional studies on terrorism and provides information about the characteristics of the group structure (influencing motives, behaviors, and outcomes), the recruitment processes, and the division of power among its members (Perliger & Pedahzur, 2011).

Appendix GUETLER01

4.4.1 Data

Data used for this study is from the Dark Web Forum Portal by the Artificial Intelligence Lab, University of Arizona. The Dark Web international jihadist forums are dedicated to Islamic ideology and theology; however, some members sympathize with jihadist groups (H. Chen, 2011). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). The Dark Web Forum Portal contains twenty-eight forums from 2004 to 2012; forums are in English, Arabic, French, German and Russian. I analyze two English forums, the Ansar1, is the English version of the Arabic Ansar AlJihad Network, an invitation-only jihadist forum popular with western jihadists (J. Scanlon & Gerber, 2014; Skillicorn, 2010). The forum has 382 members, 29,492 posts, and 11,244 threads. Previous research (see chapter 3) has identified the topics discussed on this forum include jihad, terrorism, and religion. Myiwc (Islamic Web-Community) forum has 756 members, 25,016 posts, and 6,310 threads. Within the English-language forums, these two are the smallest in terms of members.

Forums	**Year**	**Posts**	**Threads**	**Users**
Ansar1	2008-2010	29,492	11,244	382
Islamic Web-Community	2000-2010	25,016	6,310	756

Table 4.1: Dark Web Forums Data

4.4.2 Analysis

Network Model

Networks are conceptualized as graphs. A graph $G(V, E)$ consists of a set of vertices V (nodes) and a set of edges E (links). G is the whole network. It should be noted that nodes are interchangeably addressed as vertices and ties as edges. Due to different types of vertices and links in networks, it is common to measure sets of vertices and edges (Papadopoulos

Appendix GUETLER01

et al., 2012; Pete et al., 2020). The Dark Web forums are organized around specific topics within threads. The threads are an ordered collection of posts by members. The first post begins a thread and others respond by replying to the author/member of the first post or others. In this case, one could measure the set of vertices V to include the members, the posts, and threads, thus $V = (M, P, T)$. The set of edges E would consist of the member-post, post-threads, member-threads, $E = (MP, PT, MT)$.

In this study, the ties among all pairs of nodes in the network and one-mode and two-mode networks (Borgatti et al., 2013; Borgatti & Halgin, 2014; Wasserman & Faust, 1994) are analyzed. In two-mode, affiliation network models, individuals are connected by attending an event, class, etc. The individuals or events are tied to the extent they share affiliations (Borgatti et al., 2018). In this case, members of the forums are connected when they post on a particular thread. The assumption is that attending an event or participating in the same forum and thread indicates an underlying social relationship between members or a potential opportunity for one to develop (Borgatti et al., 2018; Scott & Carrington, 2014). For this study, I create a bipartite network (two-mode) with two nodes $V = (M, T)$ where $M =$ members, and $T =$ thread and the edges, $E = (MT)$. Two-mode data can be converted to one-mode data or analyzed directly as a bipartite graph (Borgatti et al., 2018). I analyzed the one-mode data (unipartite) consisting of members-members interaction, which measures the interaction when two members post in the same thread and threads-threads, where thread nodes will be connected if they share an affiliation to a member. Interactions occur when two members post in the same thread, thus participating or co-occurring within a particular thread. Members who post frequently or together in the same thread will have multiple ties, which can be used as weights associated with the edges (ties).

Thus, the analysis begins with the raw dataset of the forums being pre-processed to produce the interaction network of members for analysis. First, observations with missing values and duplicates were removed. Next, the network was constructed in RStudio using iGraph packages (Csardi & Nepusz, 2022) and imported to Gephi, an open-source network

graph and analysis tool (Bastian et al., 2009) for visualization. Finally, the undirected interactions network between members of the forums is measured using network statistics outlined below.

Network Measures

I measure and report the network statistics, structure, and node-level characteristics. Measures of network statistics include size (number of nodes and edges), density, average degrees, degree assortativity, network diameter, and average path lengths. Network density measures the extent to which members are connected by direct relations (Scott, 2000), the more nodes are connected, the more dense the graph will be. Density can indicate how covert a network is (Koschade, 2006). However, most covert networks balance between efficiency (low number of unconnected ties) and survivability (high density and high number of redundant ties) (Basu, 2014; Perliger & Pedahzur, 2011). Covert networks and Dark Web forum users are interested in privacy; thus, high density levels and group centrality increase the chance of being exposed and monitored. On the other hand, high density facilitates indoctrination as information spreads, and diffusion is faster in densely interconnected networks (Koschade, 2006; Perliger & Pedahzur, 2011; Pete et al., 2020). Thus, analyzing the network density of the Dark Web forums of jihadists is useful in understanding the information spread and user engagement within the forums.

Similarly, the average degree measures the number of edges attached to a node. The network diameter measures the length of the longest path between any two vertices in a network. The average path length measures the steps taken on average to reach any node in the graph—the average distance between all pairs of nodes. Degree assortativity measures the similarity of nodes based on their degree, how nodes tend to connect to similar nodes (homophily). When high degree nodes are connected to high degree nodes, and low degree nodes are connected to low degree nodes, the network is considered assortative. Some nodes with a high degree may be connected to low degree nodes; these are disassortative networks

(Barrenas et al., 2009; Foster et al., 2010). Assortativity provides information on the pattern of connections between nodes with a similar degree and whether they are likely to interact (Pete et al., 2020). Assortative networks are likely to remain connected with the removal of a node. In contrast, disassortative networks limit the effects of node removal as important nodes tend to be isolated from each other (Foster et al., 2010).

The following network properties are measured to uncover the large-scale network structure of the Dark Web forums. Hubs and small-world property, and community detection. Hubs in a network are nodes with a larger number of connections. A network with hubs is considered decentralized as opposed to a hierarchical network. Such graphs are likely to have the small-world property (Kleinberg, 2001; Pete et al., 2020; Watts & Strogatz, 1998) where two individuals can reach each other in the network through a short path length, a short sequence of acquaintances. The presence of hubs makes it likely to connect many nodes in the network.

In community detection, vertices or nodes are organized into communities or clusters (Fortunato & Hric, 2016; Papadopoulos et al., 2012). Within a community, a set of vertices or nodes are tightly connected in a set and loosely connected to vertices outside of the community. Thus, highly interconnected but with few ties or links to other nodes (Pons & Latapy, 2006). Community detection algorithms are valuable for analyzing large complex networks by identifying the groups of more densely connected nodes than the rest of the network. Detecting the communities or subgroups in the Dark Web forums is significant for analyzing the interaction network's structure, organization, and function. Thus, identifying the groups that tend to participate or interact in forums and threads about a specific topic, e.g., religion, ideologies, or other forum topics. To detect densely connected communities, I utilize a variety of community detection algorithms from the iGraph package and report the results of the Louvain algorithm (Blondel et al., 2008), one of the most popular, efficient, and faster algorithms for uncovering community structure in larger networks. In addition, the algorithm reports the modularity score, which measures how effective a given clustering

algorithm is; higher modularity means the algorithm has identified distinct groups and the extent to which the network can be divided into groups.

At the node level, I next measure network centrality to identify the most influential forum members using eigenvector/degree centrality, betweenness, and closeness (Borgatti et al., 2018; Freeman, 1978). Degree centrality measures the number of connections a node has, eigenvector centrality measures the importance of a node if its neighbors are influential as well. Betweenness centrality identifies the 'bridge' nodes, which are significant in connecting subgroups and fall along the shortest path between two nodes. Closeness centrality measures a node's distance to others how close a node is to others in the network. A node with high closeness score is a short distance to others, meaning that information flows quickly to other nodes and the central nodes.

Below I present the results of the interaction networks of the two Dark Web forums, Islamic Web-Community (Mywic) and Ansar1.

4.5 Results

The networks examined from the Dark Web forums are small; the Islamic Web-Community has 638 nodes and 4438 edges, Ansar1 interaction network has 353 nodes and 8765 edges (see Table 4.2). For both networks, to facilitate further analysis and graph visualization, nodes were pruned by removing nodes with zero connections (isolates) and parallel edges; several edges between one pair of nodes were merged and averaged. The Islamic Web-Community network has a smaller density compared to the Ansar1 network. The density score is an indicator of the connectedness of a network, in the case of the Dark Web forums, the covertness and the ability of members to interact by creating a thread or posting in a thread. Smaller density scores indicate a highly covert network, and higher density scores indicate efficiency (Koschade, 2006). Therefore as ideas spread faster in a denser connected network (Pete et al., 2020), a sparsely connected network may not be as effective. Members

Appendix GUETLER01

Network Measures	Islamic Web-Community	Ansar1
Number of nodes (members)	638	353
Number of edges (interaction)	4438	8765
Density	0.022	0.141
Average degree	13.912	49.66
Network diameter	6	5
Assortativity (homophily)	0.022	-0.026
Average path length	2.437	1.939
Average clustering coefficient	0.831	0.73

Table 4.2: Interaction Network Statistics

with the highest degree scores are connected to other nodes on average by 14 nodes for the Islamic Web-Community, and 50 nodes for the Ansar1 network. Therefore, the Ansar1 forum is highly active, and members participate in many threads and interact with others. Assortativity measures homophily, the tendency of nodes with the same degree to connect, thus forming an assortative network. Some nodes with a high degree may be connected to low degree nodes; these are disassortative networks (Barrenas et al., 2009; Foster et al., 2010). The Islamic Web-Community network is assortative, meaning that members with high degree nodes tend to connect with other high degree nodes in the network and low degree nodes to low degree nodes ($r = 0.022$). The Ansar1 forum has a negative disassortative network structure, meaning nodes with a high degree tend to connect with lower degree nodes ($r =$ -0.026). Hence, low degree nodes tend to participate in the same thread as nodes with a higher degree.

The clustering coefficient and average path length analyze the small world effect. A higher clustering coefficient indicates that many connections of a node are also connected. Both networks have a high clustering coefficient, Islamic Web-Community (0.831) and Ansar1 (0.73). The average path length measures the steps to get from one node to another in the network. On average, members of the Islamic Web-Community and Ansar1 forums are about two steps from another member interacting in any other thread. Meaning that any two random members participating in thread1 and another participating in thread2 are connected to at least one member who participates in both thread1 and thread2 (Pete et

al., 2020). The network diameter of the Islamic Web-Community is 6, and for Ansar1, the longest path is 5. Diameter indicates how far information must spread from one node from one end to the other end of the network. Therefore, lower average paths facilitate the faster spread of information and ideas.

Community detection algorithms are applied to find tightly connected nodes in a sub-group/community and loosely connected to nodes outside the community. The modularity algorithm (Newman, 2006) produces an overall modularity score which measures the network division into distinct modules or groups. A higher modularity score indicates the presence of densely connected nodes within a community. Below, I present the network graph showing the two forums' communities (Figures 4.1 and 4.2).

The Islamic Web-Community (Myiwc) resulted in eleven communities with a modularity score of 0.361. 37% of the nodes are part of the largest community in pink, 31% are part of the green cluster, while 12 % and 10% of the nodes were part of the blue and black clusters, respectively. Between 5% and 0.31%, the rest of the nodes were part of the smaller communities. The percentage of nodes in each group suggests that nodes within each forum were classified in the densely connected communities. Since the forums are based on threads with specific topics, the network division seen here could be due to members participating in threads they are interested in discussing.

Network centrality measures highlight the influential members and their positions (see Tables 5.1 and 5.2 in the Appendix). The top 3 influential members of the Islamic Web-Community network are Tayeb, Lulua, and OmMohammed. They all have a higher score in degree (more connections), eigenvector (influential neighbors), betweenness (bridge nodes), and closeness centrality. Further analysis of the influential members posting behavior shows that they tended to post frequently. For instance, in this network, Tayeb, the most central node in the pink community (see network graph), had the most posts (2848 out of 25,016) in the forum, participating in 1258 threads (out of 6,310) by discussing issues related to religion (Islam vs. Christianity), terrorists (Taliban, Al-Qaeda), and relationships. This

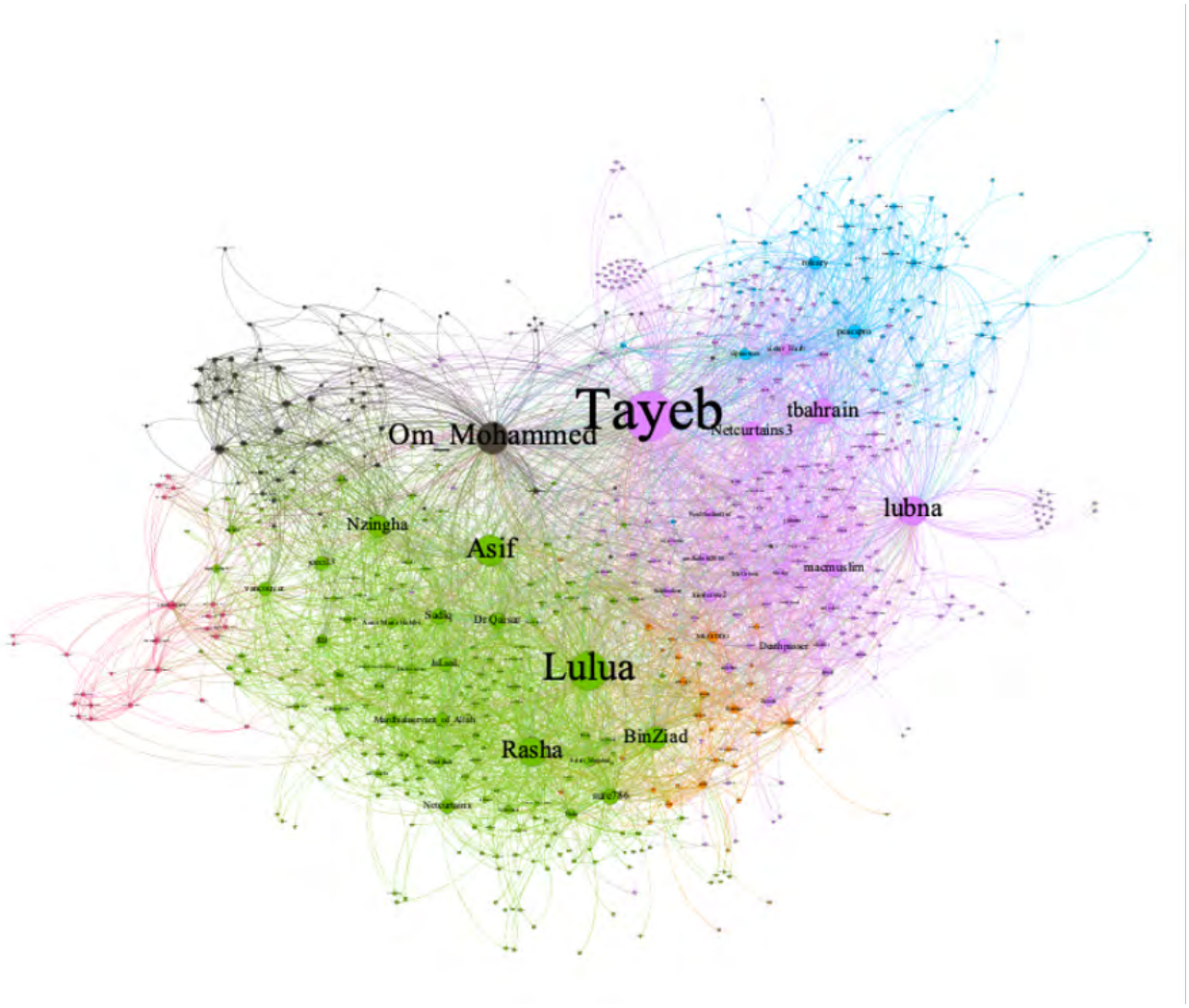


Figure 4.1: Islamic Web-Community (Myiwc) undirected network with eleven communities. 638 nodes, 4438 edges, node size represents the node degree, colors represent communities.

node mainly participated in the thread titled 'Logic of some Christians. . . not all!' (62 posts) and 'Historical Jesus vs historical Muhammad' (24 posts) discussing Christianity and Islam. The second node with high centrality score is Lulua belonging to the community in green. Lulua had 1659 posts in 1114 threads and responded to threads related to religion, poverty, relationships, politics, beauty and mental illness, news. The node frequently posted in the threads 'Comparative Study between Quran and Bible. . . " (35 times), 'News!' (25 times) and 'Islamic Quest' (14 times). The node OmMohammed in the black community posted less frequently than the other two. This node has 530 posts in 355 threads. Topics discussed

include jihad, terrorists and extremism, religion, women's rights, relationships. The node frequently posted in threads about the 'Can Salvation be reached through Christ?' (9 times) and 'A message for People of the Book...' (7 times) and posted a few times in the rest of the threads. One could note the nodes participating in the same threads or mentioning other members with high centrality scores. For instance, OmMohammed responded to a threat titled 'Welcome back bro Tayeb.' Both OmMohammed and Tayeb participated in similar threads. While the node Tayeb frequently posted in this forum, other nodes with lower centrality scores had the most posts (see Appendix). The nodes, Lubna had 1868 posts, netcurtain3 had 1776 posts, while the node Lulua with 1659 posts, is ranked second in all the centrality measures.

The communities in this network are based on topics discussed. The division stems from the specific discussions surrounding religion – Islam vs. Christianity. While the central nodes in the pink and green communities tended to discuss issues related to terrorism, the central node in the green subgroup tended to focus primarily on Islamic topics and discussions related to relationships. Other prominent members in the network tended to discuss similar topics, and some participated in forums about computer viruses and software, religion and poverty, homosexuality, food, and marriage.

The Ansar1 network (see figure 4.2) resulted in four communities with a modularity score of 0.296. In addition, 64% of the nodes are part of the largest community in pink, 17% are part of the green cluster, while 9.35 % are in the orange cluster, and 8.22% of the nodes were part of the blue community. Similarly, the network centrality measures identified the influential members of this network; the top 3 are Asadullah A. (pink community), Insurgent (green community), and Asad'Allah (pink community).

When looking at their communities and the threads/topics, one could see similarities among influential members. For example, the node Asadullah posted 2402 times in 1674 threads. Topics included terrorism - Taliban, terrorist media center Al-Ansar media center, Pakistan under Taliban, Mujahideen, Al-Qaeda, US soldiers' death in Iraq, and religion. The

Appendix GUETLER01

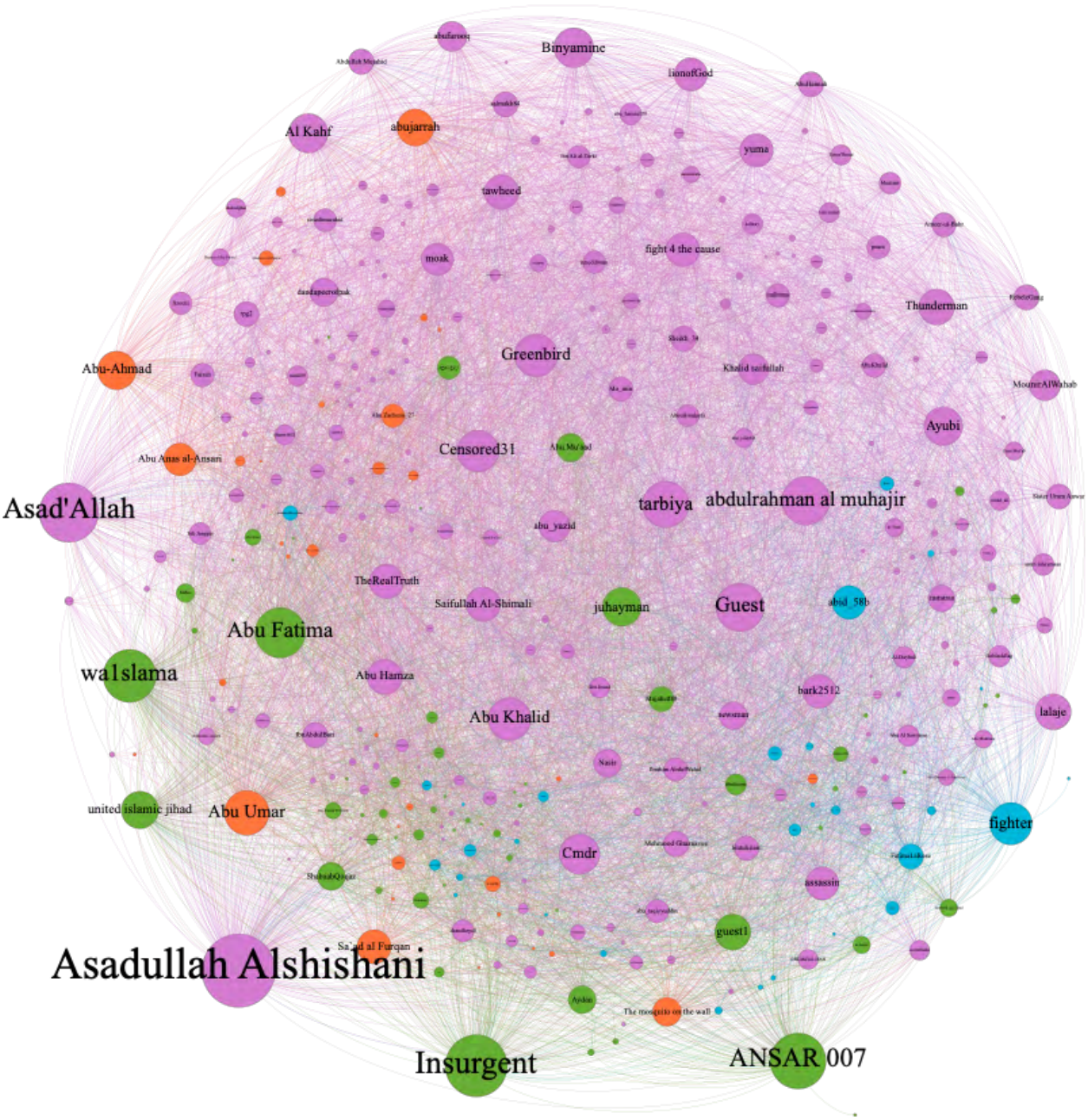


Figure 4.2: Ansar1 undirected network. 353 nodes, 8765 edges with 4 communities. Node size represents the node degree, colors represent communities.

top thread this node participated in was the '13 dead, 30 wounded in dual attacks at Fort Hood, Texas' (38 times), 'Clashes between HAMAS & Jund Ansar Allah kill 24' (28 times), and 'Feds: Leader of radical Islam group killed in raid' (22 times) in this thread, one could see prayers and support for the killed and arrested radical leaders. On the other hand, the node Insurgent posted 2101 times in 1795 threads. Topics ranged from terrorists – Islamic

Emirate of Afghanistan, Al-Ansar media, death of Iraqi police, Taliban, jihad, Al-Qaida, and religion. The thread Insurgent posted mostly is titled 'Glad Tiding for the Believers and shaking for the disbelievers…" (13 times) and 'As-Sahab: how to prevent a repeat of Gaza holocaust' (10 times) but participated less frequently in the rest of the threads.

The node Asad'Allah posted 521 posts in 324 threads, topics include terrorism – Taliban, Afghan attacks, martyrdom bomber, wills of 9/11 martyrs, youth mujahideen movement, jihad, ansar media center, Tamil Tiger rebels, Al-Qaida, Al Shabab in Somalia. The top thread the node participated in was titled "As-Sahab: An address to the American people by the Lion of Islam Sheikh Usama bin Laden (May Allah protect him)' posted eight times and 'As-Sahab: 9/11 Surprise, The West…and the Gloomy Tunnel' posted eight times as well. The node posted less in the rest of the threads. It should be noted the important nodes identified here do not necessarily mean they post the most. For instance, while the node ANSAR 007 is ranked fourth in all the centrality measures, the node posted more frequently (2213 posts) than Insurgent and Asad'Allah.

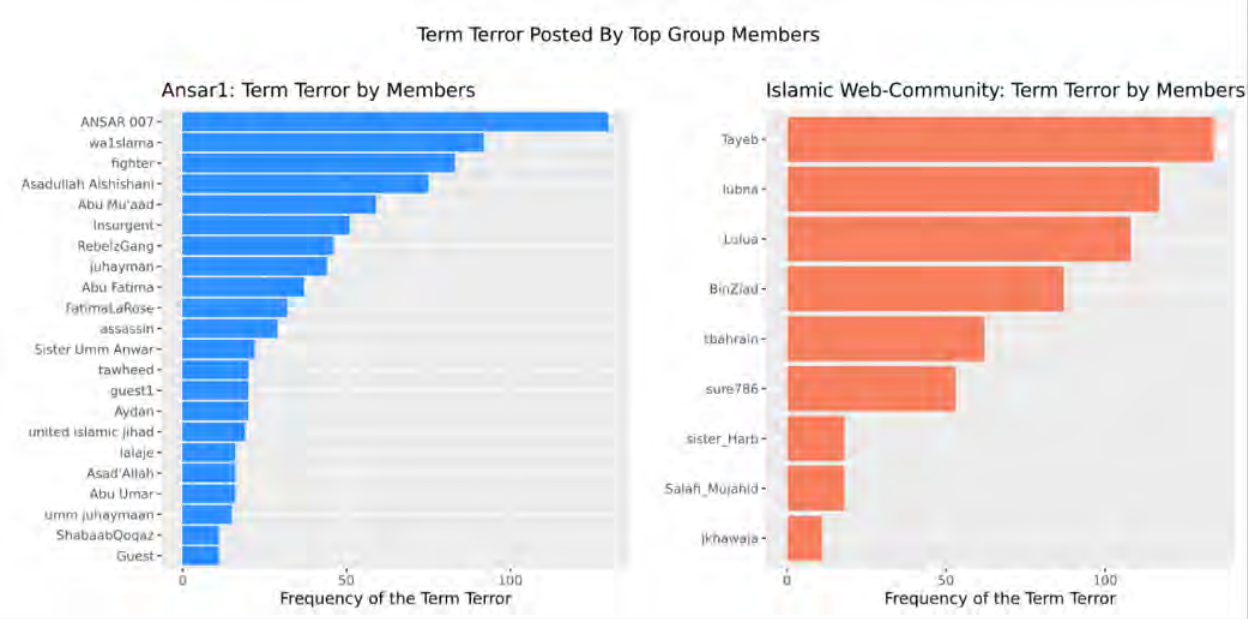


Figure 4.3: Frequency of the term terror posted by key members in the network

The division of the nodes in this network stems from the topics discussed. All three central nodes addressed issues related to terrorism, offering support and sharing news about

terrorists and the attacks. Two of the central nodes are part of the large community in pink, while one was in the green community. Other central members in the larger and small subgroups also posted about terrorism. The nodes posting behavior indicate some level of specialization, interests, or roles of the key members in the communities. For example, the central members of the Ansar1 network tend to frequently post about terrorism (figure 4.3), whereas some members of the Islamic Web-Community posted about terrorism. Findings suggest members with high centrality scores may be forum moderators, terrorist recruiters, supporters, or just frequent users with information to share.

4.6 Discussion and Conclusion

Terrorist groups use the Internet for communication, spreading propaganda, recruitment, radicalization, financing. The Internet has created opportunities for self-radicalization, connection, and interaction of like-minded people while facilitating the promotion, popularization, and operation of radical ideologies and ideologically motivated violence. Online interaction occurs in active forums such as the Dark Web forums or chat rooms, creating social bonds, trust, indoctrination, and eventually radicalization into violent extremism. Active online discussions about ideologies and opinions with friends or strangers have led to the sense of belonging within the 'virtual community' (Bowman-Grieve, 2009; Scrivens et al., 2018) and in the case of jihadists, a 'virtual Muslim community (ummah)' (Erez et al., 2011). The virtual communities change users' behavior, mirroring those within the online community.

The network structure of the two jihadist Dark Web forums, Islamic Web-Community and Ansar1, are sparsely connected, smaller density. The density score is an indicator of the connectedness of a network. Members of the Dark Web forums are less connected thus have fewer interactions. However, most covert networks balance efficiency, security, and survivability (Basu, 2014; Perliger & Pedahzur, 2011). Dark Web forum users, in general,

Appendix GUETLER01

are interested in privacy. Members join the forums via invitation or shared links; thus, high density levels increase the chance of being exposed and monitored. Forum users supporting terrorist groups are interested in secrecy as well. On the other hand, high density facilitates indoctrination and information spread; diffusion is faster in densely interconnected networks (Koschade, 2006; Perliger & Pedahzur, 2011; Pete et al., 2020).

Despite the sparse density in these forums, members with the same interests and characteristics, i.e., those who frequently post in various threads, tend to interact with similar members. For example, members in the Islamic Web-Community tended to interact with similar members, i.e., experienced or old members only interacted with experienced members. Thus, this network's information and resource flow were between similar users; one could argue trusted users. However, in the Ansar1 network, some forum members who frequently posted in almost all the threads also interacted with those who posted less or in a few threads suggesting information and resource flow between dissimilar users, the interaction between experienced and new users. Interactions between new and experienced members can shape opinions and influence new members' behavior (Pete et al., 2020). This participating pattern suggests that the Ansar1 network welcomed new members more. On the other hand, some members join the forums to seek information passively, engage with others, or share terrorist propaganda (Bloom et al., 2019; Shehabat & Mitew, 2018). As participants feel comfortable and welcomed in these platforms, skeptics and lurkers are emboldened to share their beliefs and become more vocal. In addition, since participation in these forums is an active process, users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011).

Detecting the communities or subgroups in the Dark Web forums is significant for analyzing the interaction network's structure, organization, and function. Community detection identifies the groups and members who tend to participate in forums and threads about a specific topic, e.g., religion, than they would with members discussing other issues like terrorism or relationships. Therefore, there is a division of members within the forum; com-

munities can overlap, members can participate in different threads and topics depending on their interests. Participation in forums is an active process; users select platforms most compatible with their views (Piazza & Guler, 2019; Sageman, 2011). For example, the Islamic Web-Community (Myiwc) resulted in eleven communities, and the Ansar1 network resulted in four communities.

Further analysis of the key members in these groups shows the network division is due to the members participating in threads and topics they are interested in discussing. Since the forums are based on threads with specific topics, the topics discussed by the key members in the Islamic Web-Community were mainly about the Islamic religion and a few discussions on terrorism, marriage and relationships, computer viruses, and software. The Ansar1 topics tended to be mostly about terrorism, similar to the results found in a previous paper (see chapter 3). Most of the jihadist discourse online involves religious/ideological indoctrination framed by references to the Quran, Hadith, and the early history of Islam (Erez et al., 2011). The posting activities suggest that some members tended to post in various threads. Some of the central nodes participated in the same forums or mentioned other major nodes in their discussions. Identifying the communities allows us to detect different functions of the network, such as the collaborators, key posters, recruitment patterns (Perliger & Pedahzur, 2011), and in this case, topics of interest and supporters of terrorist groups. The key members identified in these communities influence the topic discussion and information flow within the subgroups and the forums. These could be 'media mujahideen' (Malik, 2018), who amplify terrorists' online messaging campaigns by sharing and reposting official content of the groups across social media platforms and the Dark Web forums. In the posts, it was interesting to note explicit and implicit support for jihad, terrorist groups, and news shared from terrorist media such as Al-Qaeda's as-Sahab, and the Global Islamic Media Front led by supporters of jihad but unaffiliated with any terrorist organization (Erez et al., 2011).

Network destabilization, a common approach in counterterrorism, begins with identifying and removing key members in a network. Some of the actors are central to the net-

work, influencers, brokers, leaders, therefore crucial for the flow of information and resources (Koschade, 2006; Perliger & Pedahzur, 2011). Detecting actors in the network who are key to the continuous operation of the group is essential as it highlights the power division and the motives within the network. Such actors have roles imperative for the network's success; in the case of Jihadists Dark Web forums, they could be unaffiliated sympathizers, propagandists, fighters, recruiters, and rivals. While the content attracts sympathizers and would-be extremists to frequent the platforms, recruiters monitor these virtual communities to find potential recruits.

The effectiveness of disruption strategies is based on network topology and network resilience (Duijn et al., 2014). Disruption strategies target the central members, the hubs. Degree centrality identifies the actor with most direct contacts; actors with high betweennesses centrality scores are the bridge; they control the flow of information and resources within the network. For example, the removal of the top 3 influential members of the Islamic Web-Community network, Tayeb, Lulua, and OmMohammed and the Ansar1 members Asadullah A., Insurgent, and Asad'Allah may disrupt the flow of information in these forums. The key members are responsible for coordinating the group's activities, sharing the content, and starting and responding to a thread. However, the disruption approach may not be effective in the long run as new leaders, and new networks will emerge, information will continue to flow, as covert networks are quick to adapt (Everton, 2008; Koschade, 2006). Similarly, covert networks of Dark Web forums are often decentralized, and the identification and removal of key players often lead the networks to become more decentralized, thus harder to disrupt (Duijn et al., 2014; Pete et al., 2020). It should be noted while an individuals' centrality indicates importance, this does not mean that they hold power in the network (Koschade, 2006). Additionally, some members deliberately operate from the peripheries, thus not easily detected as key members for disruption. Yet, disruption strategies render covert networks more exposed, leading to the practice of network disruption a long-term effort (Duijn et al., 2014). As a result, although network disruption has some challenges

as new leaders and networks reemerge, network destabilization is still beneficial. It may temporarily destabilize the network, providing the opportunity to monitor and await new leaders or those in the peripheries to surface.

In summary, this study analyzed two Dark Web forums of jihadist groups and their supporters using social network analysis techniques. The study investigates the Dark Web forum's network structure, behavior, and patterns. The network structure is decentralized, sparse, and divided based on topics (religion, terrorism, current events, and relationships) and the members' interests in participating in these threads. Some members are essential and influential in the information and resources flow within the networks. Identifying key members is significant for counterterrorism, as mapping network structures and key users are important for destabilizing terrorist networks. Ethical implications such privacy violations, lack of informed consent and confidentiality arise from applying SNA to identify and detect key members especially in a forum used by terrorist sympathizers and supporters. While the data analyzed is considered public information therefore requiring no consent from the forum users or IRB approval, the results of this study do not imply the key members identified to be terrorists or supporters of terrorist groups.

Nevertheless, SNA is conducive for predictive and explanatory studies on terrorism and counterterrorism. Future studies perhaps may analyze current Dark Web forums for comparison as one of the limitations of this study was the analysis of forums collected from 2008 to 2012. An investigation that includes member attributes such as time posted, and events may highlight the changes over time to further understand the networks of Dark Web forums of covert networks.

Chapter 5

Conclusions

5.1 Summary

This three-article dissertation sought to explore three distinct but related topics. First was the exploration of cybercrime and cyberterrorism from a criminological perspective. Secondly, to highlight the utility of innovative computational techniques, specifically topic modeling in detecting Jihadist topics and discussion on the Dark Web forums. Thirdly, illuminate and map Jihadist Dark Web forums' social structure and behavior using social network analysis.

In the first paper, I explore existing research on cybercrime and cyberterrorism. It is known that the development of the Internet has provided terrorists with access to a global audience and target. However, criminology researchers examining cybercrime and cyberterrorism are often impeded with the challenge of defining cybercrime, terrorism, cyberterrorism, and access to data that can be used to analyze and understand cybercriminals. Similarly, terrorism and violent extremism scholars are divided as the term cyberterrorism has sparked an intense debate over whether terrorist groups can attack cyberspace and critical infrastructures to further their goals. For some researchers, terrorists have to engage in cyberattacks that result in violence against individuals (death, injury) or property (ex-

Appendix GUETLER01

plosions, economic loss) to be considered cyberterrorism, the convergence of cyberspace and terrorism. Furthermore, terrorists do not possess the technical capability required to attack cyberspace. However, some scholars argue terrorist groups engage in cyberterrorism by targeting government or businesses' online critical infrastructures. Cyberterrorism does not have to result in violence to be taken seriously.

Indeed, while the groups have not yet attacked a critical infrastructure such as power plants, they are attacking computer systems and networks, stealing data and information, thus increasingly posing serious threats to national security, the economy, and public policy. Research shows that the ability to attack cyberspace has a far-reaching impact on terrorist groups' enemies, unlike bombings restricted to a specific physical location and communities. To address their lack of technical sophistication, some group members are tech-savvy, or they recruit tech-savvy supporters to attack cyberspace through website defacement, distributed denial-of-service (DDoS), cyber-enabled financing, credit card fraud, stealing data, and selling on the Dark Web.

The existing research reveals that due to the rapid technological advancement and emerging technologies, including artificial intelligence, and the unintended consequences of these technologies, it is challenging to predict cybercriminals and cyberterrorists' technical capabilities and cyberattacks. Indeed, while there have been few incidents of cyberterrorism globally, mostly attacks against governments, organizations, and business networks, we cannot ignore the threat of severe cyberattacks and critical infrastructures posed by terrorist groups. Countries face a growing risk of cybercrime and cyberterrorism as we rely more on the Internet and technologies for communication, business, health, economy, governing, and critical infrastructures.

The second article is methodological and aims to address the issue of terrorism content on the Dark Web forums and automatic techniques, such as topic modeling that can be applied to classify and detect the topics. Terrorists use the Internet to spread propaganda/publicity, psychological warfare, data mining, fundraising, recruitment and mobilization, networking,

and secure communication facilitating the promotion, popularization, and operation of radical ideologies and violence. Interactive platforms such as chat rooms, forums, message boards, and social media allow interaction and communication amongst like-minded individuals to develop, thus fostering a sense of belonging and self, loyalty, emotional attachment, and meaning amongst the virtual community. The Jihadist content on these platforms focuses on religious and ideological indoctrination, inciting violence against their enemies, share information about attacks, strategies, and justifications.

Although criminologists have applied classical research methods to analyze terrorism and contributed to the development of novel research methods, crime theories, findings, and policies, few have yet to embrace computational approaches to the study of crime and crime-related problems, specifically terrorism and extremism studies. In the second paper, I show how natural language processing computational techniques, specifically LDA topic models, can be applied to inductively analyze and uncover topics from the large-scale information generated by terrorists and their supporters. I applied topic models and sentiment analysis on three Dark Web forums of Jihadist groups and their supporters, the Ansar1 (Ansar AlJihad), the Gawaher forum, which is dedicated to discussions of Islam and the Islamic world, and the Islamic Network forum, which is also devoted to the theology and world events. Some members of these forums sympathize with and support terrorist organizations.

Results show that the most prevalent topic in all forums was religion. Topics about religion included Islam, Christianity, and Judaism, discussion about Allah and God, believers and disbelievers, Quran recitations, religious rituals, and philosophical discussions about religion and science. These topics are expected, as the forums were dedicated to the Islamic faith and issues concerning Muslims. The forums also discussed terrorism, mainly within the Ansar1 forum. The discussion was about terrorists and attacks, support for the Mujahideen fighters, jihad, Al-Qaeda, Taliban, the U.S in Iraq, Afghanistan, the arrests and torture of terrorists and foreign fighters. The support of terrorists and sharing news about the attacks signify the media strategy of the groups and their supporters to communicate and justify

Appendix GUETLER01

their ideological goals. Indeed, one could argue the discussion or posts about terror groups were created by the terrorist group's media centers and supporters, also known as media mujahideen. A few of the discussions were related to relationships and marriages, advice, seeking help, health, food, selling electronics, and identity cards.

It is known that Jihadists' ideologies and narratives are based on the misinterpretation of the Islamic religious traditions and historical events framed by references to the Quran, Hadith, and the early history of Islam. Therefore, participating in forum discussions is also considered a form of jihad. LDA topic modeling has several advantages over the manual annotation of text data, especially when analyzing a large body of text such as the Dark Web forums. Topic models quickly discover hidden topics, providing real-time analysis of terrorist online content. The topics are inductively uncovered, thus requiring little or no previous knowledge of the content from the text. Additionally, the keywords discovered from the topic models can further be utilized to organize and search for content in other text data. For instance, the keywords derived from the three Dark Web forums can be used to analyze or search other Dark Web forums or websites for similar topics.

The topic models suggest that terrorist content and discussions occur in Dark Web forums that are created for the discussion of religion. Therefore, counterterrorism measures should also focus on data-driven applications such as topic modeling in detecting and analyzing terroristic content from large-scale textual data. Additionally, monitor Dark Web forums and especially forums that discuss religion, as terrorist groups are known to use religion to justify their goals and recruit in such forums for supporters. Topic modeling for counterterrorism is significant as it is quicker to detect and classify terrorist content for disruption than manual analysis. In addition, the detection of extremist content online using topic models is significant for analyzing users' beliefs and countering/disrupting their online narrative. The application of sentiment analysis provides insight into the opinions and attitudes of forum users toward the prevalent topics of religion and terrorism. For instance, the forum Ansar1 tend to portray negative attitudes towards the war, military, police, and government, while

the Gawaher and Islamic Network forums were mostly positive about topics surrounding the Islamic religion.

The third article builds on article two, and it is an exploratory study providing an insight into the network structure and behavior of Dark Web forum users. In this paper, I use social network analysis to examine the networks of two international Jihadist groups' Dark Web forums. I construct interaction networks amongst members of the forums to map, visualize and measure the interactions between members within the forums, illuminating the network structure, how members interact with each other, the key members, virtual communities, and how the network behaves. The research questions guiding this study are: what is the network structure of jihadist Dark Web forums, and how does the structure affect interactions? What communities can be detected from the network? And what are the posting activities of key members in the forums?

Results suggest that the network structure of the two Jihadist Dark Web forums, Islamic Web-Community and Ansar1, are sparsely connected have smaller densities. Thus, implying a decentralized network, a 'leaderless' (Sageman, 2011) network where all members communicate equally, with not one leader but several important members spread across the forums. In such networks, individuals are likely to act independently or in smaller groups. Members of the Dark Web forums are less connected thus have fewer interactions; this is expected in a covert network. Dark Web users, in general, are interested in privacy and anonymity. Members join the forums via invitation or shared links. A well-connected network would increase the chance of being exposed and monitored but also successfully facilitate indoctrination and the spread of information as diffusion is faster in densely interconnected networks. Despite the sparse density in these forums, members in the Islamic Web-Community tended to interact with similar members, i.e., experienced, or old members only interacted with experienced members. Thus, this network's information and resource flow were between similar users; one could argue trusted users. However, in the Ansar1 network, some forum members who frequently posted in almost all the threads also interacted with those who posted less

or in a few threads suggesting information and resource flow between dissimilar users, the interaction between experienced and new users. Interactions between new and experienced members can shape opinions and influence new members' behavior.

Furthermore, as participation in forums is an active process, users tend to select platforms most compatible with their views, forming a subgroup or community. For example, the Islamic Web-Community (Myiwc) resulted in eleven communities, and the Ansar1 network resulted in four communities. Further analysis of the key members in these groups shows the network division is due to the members participating in threads and topics they are interested in discussing. Since the forums are based on threads with specific topics, the topics discussed by the key members in the Islamic Web-Community were mainly about the Islamic religion and a few discussions on terrorism, marriage and relationships, computer viruses, and software. The Ansar1 topics tended to be mostly about terrorism, similar to the results found in article two. Some members are essential and influential in the information and resources flow within the networks. Identifying key members is significant for counterterrorism, as mapping network structures and key users are important for destabilizing terrorist networks. However, the results of this study do not imply the key members identified are terrorists or supporters of terrorist groups. The analysis of the interaction networks provides an understanding of terrorist behavior, thus leading to more effective counterterrorism measures such as disruption and quicker intelligence gathering of online terrorist networks and operations.

5.2 Limitations and Future Research

This dissertation has explored the problem of cyberterrorism and the applicability of automatic computational methods in analyzing Dark Web forums and networks. However, this dissertation is not without limitations. First, as noted elsewhere, the lack of access to official data on terrorism and cyberterrorism is one of the limitations for any research

Appendix GUETLER01

on terrorism and cyberterrorism (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Diamond & Bachmann, 2015; Holt, Lee, et al., 2020). Official data on terrorism is limited; if available, it is censored or influenced by political considerations. Some countries do not collect terrorism data, and those who keep terrorism databases are unwilling to share it with the research community (Hodwitz, 2019; LaFree & Ackerman, 2009). Additionally, in some countries such as the U.S, terrorists are often not prosecuted for terrorism acts but other related offenses; therefore, gathering data of those arrested, convicted, or prosecuted for terrorism acts is challenging (LaFree & Dugan, 2015; Rosenfeld, 2002). Industry or open-source government data are rarely available due to victims underreporting over fears of economic loss (Holt, Cale, et al., 2021; Holt & Steinmetz, 2020). Furthermore, cyberattack incidents are often omitted from terrorism databases "due to their failure to qualify as a "violent" incident" (Freilich et al., 2014; Holt, Cale, et al., 2021; Holt, Navarro, et al., 2020). As such, most scholars have resorted to examining a sample of college and school students, mainly surveying their hacking and digital piracy experiences, analyzing bulletin boards and blogs, and field experiments (A. M. Bossler & Berenblum, 2019; Bossler A.M., 2017; Burruss et al., 2012; Holt, Lee, et al., 2020). Others have collected data from simulated computer systems created to be attacked to analyze system intrusions (Maimon et al., 2013; Testa et al., 2017).

Due to the challenge of accessing terrorist online content, in this study, I am analyzing a secondary dataset, the Dark Web Portal, collected by the University of Arizona, Artificial Intelligence Lab. Therefore, the analysis is limited to the data and variables available in the collection. As a result, I may not have been able to analyze other essential variables not available in the database that would be conducive to exploring topics and users of Dark Web forums, such as demographics and locations of the Dark Web forum members. It is important to note while it is easier and faster to collect behavioral data of Internet users online, it is challenging to collect and assess demographic data from Internet users. The Dark Web forums include username, posts, threads, date, and year of the posts. The names are expected to be aliases of the users interested in anonymity. However, including the username

and dates they participated in the forums in future research would provide insight into the members' topics and time spent interacting with others. The analysis of time spent on the forums, the topics discussed, and interactions may classify radical, non-radical, or neutral users.

Secondly, although the dataset analyzed of forums was from 2004 – 2012, it still provided insight into the Dark Web forums topics and network behavior. A comparison with a current dataset would provide insight into the changes in topics and posting behaviors of the forum users. A current dataset may also be used to compare the language use of forum users, as language is known to evolve. Thirdly, this dissertation has taken a computational social science approach that integrates social sciences with statistics and computer science techniques focusing on topic modeling, specifically LDA. However, other topic modeling algorithms such as structural topic modeling and author-topic models can be applied to identify the authors' topics. Additionally, different computational approaches can be used to analyze text data, such as semi-supervised machine learning algorithms. For instance, a semi-supervised study of Dark Web forums would entail manually annotating a percentage of the forum posts as radical, non-radical, or neutral and using those annotations as dependent variables that can be applied to classify the rest of the Dark Web forums and users. Another application would be, instead of manually classifying the Dark Web forums, to use the results of topic models and machine learning classifiers to analyze the rest of the Dark Web forums. Additionally, topic models can be combined with sentiment analysis, topic models highlight the topics discussed in the Dark Web forums, and sentiment analysis highlights the attitudes of forum users based on their posts and topics i.e., negative, positive, or neutral. A further analysis of the crime-terror nexus on the Dark Web would shed light into the partnership between cybercriminals and cyberterrorists and their behaviors.

Thirdly, social networks analysis of the Dark Web forums was exploratory. Future directions may include a deeper analysis of the topics identified by topic modeling and key members to highlight what they post and how they behave and interact in the Dark Web

forums. Another future direction will employ time-series analysis to measure the network changes, identifying key members and groups and how they change over time. Another approach would be to apply Exponential Random Graph Models (ERGM) to analyze and predict the relationship of forum members based on their posts and time spent on the forums.

Lastly, although this research applied computational social science methods such as topic models and social network analysis, the ethical concerns arising from the computational techniques should be noted. Applying topic models to classify text as radical or non-radical and identifying key forum users based on their network position and posting behavior may lead to unnecessary profiling and/or mislabeling Internet users who may have accessed the Dark Web forums for legitimate purposes.

Appendix GUETLER01

References

Abbasi, A., & Chen, H. (2007). Affect Intensity Analysis of Dark Web Forums. *2007 IEEE Intelligence and Security Informatics*, 282–288. https://doi.org/10.1109/ISI.2007.379486

Adam, A. (2005). Hacking into Hacking: Gender and the Hacker Phenomenon. In A. Adam (Ed.), *Gender, Ethics and Information Technology* (pp. 128–146). Palgrave Macmillan UK. https://doi.org/10.1057/9780230000520_7

Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Awan, I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human-centric Computing and Information Sciences*, *9*(1), 24. https://doi.org/10.1186/s13673-019-0185-6

Akers, R. L., Krohn, M. D., Lanza-Kaduce, L., & Radosevich, M. (1979). Social Learning and Deviant Behavior: A Specific Test of a General Theory. *American Sociological Review*, *44*(4), 636–655. https://doi.org/10.2307/2094592

Akers, R. L., Silverman, A. L., & Silverman, A. L. (2014). Toward a Social Learning Model of Violence and Terrorism. Retrieved September 24, 2020, from http://www.taylorfrancis.com/

Akins, J., & Winfree, J., Latham. (2016). Social Learning Theory and Becoming a Terrorist: New Challenges for a General Theory. *The Handbook of the Criminology of Terrorism* (pp. 133–149).

Ali, M. Y., & Bwana, O. M. (2015). Peace-building and Conflict Prevention Training Manual and Resource Guide for Building Resilience Against Violent Extremism. https://cscrcenter.org/

Almquist, Z. W., & Bagozzi, B. E. (2019). Using Radical Environmentalist Texts to Uncover Network Structure and Network Features. *Sociological Methods & Research*, *48*(4), 905–960. https://doi.org/10.1177/0049124117729696

Aly, A., Macdonald, S., Jarvis, L., & Chen, T. M. (2017). Introduction to the Special Issue: Terrorist Online Propaganda and Radicalization. *Studies in Conflict & Terrorism*, *40*(1), 1–9. https://doi.org/10.1080/1057610X.2016.1157402

Aly, A., Weimann-Saks, D., & Weimann, G. (2014). Making 'Noise' Online:An Analysis of the Say No to Terror Online Campaign. *Perspectives on Terrorism*, *8*(5). Retrieved September 24, 2020, from http://www.terrorismanalysts.com/pt/index.php/pot/article/view/376

Anzalone, C. (2016). Continuity and Change: The Evolution and Resilience of Al-Shabab's Media Insurgency, 2006–2016. Retrieved March 11, 2022, from https://www.belfercenter.org/publication/continuity-and-change-evolution-and-resilience-al-shababs-media-insurgency-2006-2016

Archer, N. (2012). Consumer identity theft prevention and identity fraud detection behaviours. *Journal of Financial Crime*, *19*(1), 20–36. https://doi.org/10.1108/13590791211190704

Armborst, A. (2010). Modelling Terrorism and Political Violence [Publisher: SAGE Publications Ltd]. *International Relations*, *24*(4), 414–432. https://doi.org/10.1177/0047117810385779

Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in Knowledge*

Discovery and Data Mining (pp. 391–402). Springer. https://doi.org/10.1007/978-3-642-13657-3_43

Asal, V., & Rethemeyer, R. K. (2008). The Nature of the Beast: Organizational Structures and the Lethality of Terrorist Attacks. *The Journal of Politics*, *70*(2), 437–449. https://doi.org/10.1017/S0022381608080419

Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Society*, *54*(2), 138–149. https://doi.org/10.1007/s12115-017-0114-0

Bachmann, M. (2010). The Risk Propensity and Rationality of Computer Hackers. *International Journal of Cyber Criminology*, *4*(1), 14.

Back, S., Soor, S., & LaPrade, J. (2018). Juvenile Hackers: An Empirical Test of Self-Control Theory and Social Bonding Theory. *International Journal of Cybersecurity Intelligence & Cybercrime*, *1*, 17.

Barrenas, F., Chavali, S., Holme, P., Mobini, R., & Benson, M. (2009). Network Properties of Complex Human Disease Genes Identified through Genome-Wide Association Studies. *PLOS ONE*, *4*(11), e8090. https://doi.org/10.1371/journal.pone.0008090

Bartlett, J., & Miller, C. (2012). The Edge of Violence: Towards Telling the Difference Between Violent and Non-Violent Radicalization. *Terrorism and Political Violence*, *24*(1), 1–21. https://doi.org/10.1080/09546553.2011.594923

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*.

Basu, A. (2014). Social Network Analysis: A Methodology for Studying Terrorism. In M. Panda, S. Dehuri, & G.-N. Wang (Eds.), *Social Networking: Mining, Visualization, and Security* (pp. 215–242). Springer International Publishing. https://doi.org/10.1007/978-3-319-05164-2_9

Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of*

the Association for Information Science and Technology, *68*(6), 1397–1410. https://doi.org/10.1002/asi.23786

Behr, I. V., Reding, A., Edwards, C., & Gribbon, L. (2013). Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism. *RAND*, 76.

Bender, E. (2019). The #BenderRule: On Naming the Languages We Study and Why It Matters. Retrieved February 17, 2022, from https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W., & Müller, C. (2021). Quanteda: Quantitative Analysis of Textual Data. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=quanteda

Benson, D. C. (2014). Why the Internet Is Not Increasing Terrorism. *Security Studies*, *23*(2), 293–328. https://doi.org/10.1080/09636412.2014.905353

Berger, J. M. (2016). Nazis vs. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks, 32.

Berger, J. M. (2018). *Extremism*. The MIT Press.

Berger, J. M., & Strathearn, B. (2013). Measuring influence, evaluating content and countering violent extremism in online social networks. *The International Centre for the Study of Radicalisation and Political Violence*, 56.

Berger, J. (2017). Extremist Construction of Identity: How Escalating Demands for Legitimacy Shape and Define In-Group and Out-Group Dynamics. *Terrorism and Counter-Terrorism Studies*. https://doi.org/10.19165/2017.1.07

Bermingham, A., Conway, M., McInerney, L., O'Hare, N., & Smeaton, A. (2009). Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation. https://doi.org/10.1109/ASONAM.2009.31

Berton, B. (2015). The dark side of the web: ISIL's one-stop shop? — European Union Institute for Security Studies. Retrieved March 1, 2022, from https://www.iss.europa.eu/content/dark-side-web-isil%E2%80%99s-one-stop-shop

Bienenstock, E. J., & Salwen, M. (2015). Covert Network Analysis: An Exchange Network Theory Perspective. In L. M. Gerdes (Ed.), *Illuminating Dark Networks: The Study of Clandestine Groups and Organizations* (pp. 8–18). Cambridge University Press. https://doi.org/10.1017/CBO9781316212639.002

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. "O'Reilly Media, Inc."

Bisgin, H., Arslan, H., & Korkmaz, Y. (2019). Analyzing the Dabiq Magazine: The Language and the Propaganda Structure of ISIS. In R. Thomson, H. Bisgin, C. Dancy, & A. Hyder (Eds.), *Social, Cultural, and Behavioral Modeling* (pp. 1–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-21741-9_1

Black, D. (2004). Terrorism as Social Control. In M. Deflem (Ed.), *Terrorism and Counter-Terrorism* (pp. 9–18). Emerald Group Publishing Limited. https://doi.org/10.1108/S1521-6136(2004)0000005003

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation [Publisher: Microtome Publishing, Brookline, MA]. *Journal of Machine Learning Research (JMLR)*, *3*(4-5), 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bloom, M., & Daymon, C. (2018). Assessing the Future Threat: ISIS's Virtual Caliphate. *Orbis*, *62*(3), 372–388. https://doi.org/10.1016/j.orbis.2018.05.007

Bloom, M., Tiflati, H., & Horgan, J. (2019). Navigating ISIS's Preferred Platform: Telegram1. *Terrorism and Political Violence*, *31*(6), 1242–1254. https://doi.org/10.1080/09546553.2017.1339695

Bodine-Baron, E., Helmus, T., Magnuson, M., & Winkelman, Z. (2016). *Examining ISIS Support and Opposition Networks on Twitter*. RAND Corporation. https://doi.org/10.7249/RR1328

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks* (1st edition). SAGE Publications Ltd.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing Social Networks*. SAGE.

Borgatti, S. P., & Halgin, D. S. (2014). Analyzing Affiliation Networks. *The SAGE Handbook of Social Network Analysis* (pp. 417–433). SAGE Publications Ltd. https://doi.org/10.4135/9781446294413.n28

Borum, R. (2011). Radicalization into Violent Extremism II: A Review of Conceptual Models and Empirical Research. *Journal of Strategic Security*, *4*(4). https://doi.org/⟨p⟩http://dx.doi.org/10.5038/1944-0472.4.4.2⟨/p⟩

Bossler, A., & Burruss, G. (2011). The General Theory of Crime and Computer Hacking: Low Self-control Hackers? *Corporate Hacking and Technology-driven Crime: Social Dynamics and Implications*, 38–67. https://digitalcommons.georgiasouthern.edu/crimjust-criminology-facpubs/247

Bossler, A., & Holt, T. (2009). On-line Activities, Guardianship, and Malware Infection: An Examination of Routine Activities Theory. *International Journal of Cyber Criminology*, *3*(1), 400–420. https://digitalcommons.georgiasouthern.edu/crimjust-criminology-facpubs/259

Bossler, A. M., & Berenblum, T. (2019). Introduction: New directions in cybercrime research [Publisher: Routledge]. *Journal of Crime and Justice*, *42*(5), 495–499. https://doi.org/10.1080/0735648X.2019.1692426

Bossler A.M. (2017). Need for Debate on the Implications of Honeypot Data for Restrictive Deterrence Policies in Cyberspace. *Criminology and Public Policy*, *16*(3), 681–688. https://doi.org/10.1111/1745-9133.12322

Bowman-Grieve, L. (2009). Exploring "Stormfront": A Virtual Community of the Radical Right. *Studies in Conflict & Terrorism*, *32*(11), 989–1007. https://doi.org/10.1080/10576100903259951

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press. https://doi.org/10.1093/oso/9780190684099.001.0001

Brayne, S., & Christin, A. (2021). Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. *Social Problems*, *68*(3), 608–624. https://doi.org/10.1093/socpro/spaa004

Bright, D. A., Hughes, C. E., & Chalmers, J. (2012). Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate. *Crime, Law and Social Change*, *57*(2), 151–176. https://doi.org/10.1007/s10611-011-9336-z

Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Mårtenson, C., & Svenson, P. (2013). Harvesting and analysis of weak signals for detecting lone wolf terrorists. *Security Informatics*, *2*(1), 11. https://doi.org/10.1186/2190-8532-2-11

Burris, V., Smith, E., & Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, *33*(2), 215–235. https://doi.org/10.1080/00380237.2000.10571166

Burruss, G. W., Bossler, A. M., & Holt, T. J. (2012). Assessing the Mediation of a Fuller Social Learning Model on Low Self-Control's Influence on Software Piracy: *Crime & Delinquency*. https://doi.org/10.1177/0011128712437915

Caló, B., & Hartley, E. (2019). ISIL recruiters as social media influencers: Mechanisms of legitimation by Australian Muslim men. *Contemporary Voices: St Andrews Journal of International Relations*, *1*, 2. https://doi.org/10.15664/jtr.1497

Campedelli, G. M., Bartulovic, M., & Carley, K. M. (2019). Pairwise similarity of jihadist groups in target and weapon transitions. *Journal of Computational Social Science*, *2*(2), 245–270. https://doi.org/10.1007/s42001-019-00046-8

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Carley, K., Lee, J.-S., & Krackhardt, D. (2001). Destabilizing Networks. *Connections*, *24*, 79–92.

Carlin, J. P., & Graff, G. M. (2018). *Dawn of the Code War: America's Battle Against Russia, China, and the Rising Global Cyber Threat*. PublicAffairs.

Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, *20*(1), 21–39. https://doi.org/10.1177/1362480615586614

Chang, J., & Blei, D. (2009). Relational Topic Models for Document Networks. *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, 81–88. Retrieved February 18, 2022, from https://proceedings.mlr.press/v5/chang09a.html

Chen, H. (2011). *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Springer Science & Business Media.

Chen, H., Chung, W., Qin, J., Reid, E., Sageman, M., & Weimann, G. (2008). Uncovering the dark Web: A case study of Jihad on the Web. *Journal of the American Society for Information Science and Technology*, *59*(8), 1347–1359. https://doi.org/10.1002/asi.20838

Chen, T. M., Jarvis, L., & Macdonald, S. (Eds.). (2014). *Cyberterrorism*. Springer New York. https://doi.org/10.1007/978-1-4939-0962-9

Chen, Y., Wu, X., Hu, A., He, G., & Ju, G. (2021). Social prediction: A new research paradigm based on machine learning. *The Journal of Chinese Sociology*, *8*(1), 15. https://doi.org/10.1186/s40711-021-00152-z

Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy*, *2*(1), 26–38. https://doi.org/10.1080/23738871.2017.1298643

Choi, J., Kruis, N. E., & Choo, K.-S. (2021). Explaining Fear of Identity Theft Victimization Using a Routine Activity Approach. *Journal of Contemporary Criminal Justice*, *37*(3), 406–426. https://doi.org/10.1177/10439862211001627

Choi, K.-s., Lee, C., & Cadigan, R. (2018). Spreading Propaganda in Cyberspace: Comparing Cyber-Resource Usage of Al Qaeda and ISIS. *International Journal of Cybersecurity Intelligence & Cybercrime*, *1*(1), 21–39. https://vc.bridgew.edu/ijcic/vol1/iss1/4

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588–608.

Conway, M. (2003). Hackers as terrorists? Why it doesn't compute. *Computer Fraud & Security*, *2003*(12), 10–13.

Conway, M. (2014). Reality Check: Assessing the (Un)Likelihood of Cyberterrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 103–121). Springer. Retrieved January 24, 2022, from https://doi.org/10.1007/978-1-4939-0962-9_6

Conway, M. (2017). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies in Conflict & Terrorism*, *40*(1), 77–98. https://doi.org/10.1080/1057610X.2016.1157408

Crenshaw, M. (2011). *Explaining terrorism: Causes, processes, and consequences* [OCLC: 457164350].

Csardi, G., & Nepusz, T. (2022). Igraph: Network Analysis and Visualization. Retrieved March 4, 2022, from https://CRAN.R-project.org/package=igraph

Danilak, M. M. (n.d.). Langdetect: Language detection library ported from Google's language-detection. Retrieved February 24, 2022, from https://github.com/Mimino666/langdetect

Deflem, M. (Ed.). (2004). *Terrorism and counter-terrorism: Criminological perspectives*. Elsevier.

Denning, D. E. (2000). Cyberterrorism: The logic bomb versus the truck bomb. *Global Dialogue*, *2*(4), 29.

Denning, D. E. (2001). Activism, hacktivism, and cyberterrorism: The Internet as a tool for influencing foreign policy. *Networks and netwars: The future of terror, crime, and militancy*, *239*, 288.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, *17*(1), 61–84. https://doi.org/10.3166/dn.17.1.61-84

Diamond, B., & Bachmann, M. (2015). Out Of The Beta Phase: Obstacles, Challenges, And Promising Paths In The Study Of Cyber Criminology. https://doi.org/10.5281/ZENODO.22196

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Dion, M. (2010). Advance Fee Fraud Letters as Machiavellian/Narcissistic Narratives. *4*(1), 13.

Duijn, P. A. C., Kashirin, V., & Sloot, P. M. A. (2014). The Relative Ineffectiveness of Criminal Network Disruption. *Scientific Reports*, *4*(1), 4238. https://doi.org/10.1038/srep04238

Edwards, C., & Gribbon, L. (2013). Pathways to Violent Extremism in the Digital Era. *The RUSI Journal*, *158*(5), 40–47. https://doi.org/10.1080/03071847.2013.847714

Erez, E., Weimann, G., & Weisburd, A. A. (2011). Jihad, Crime, and the Internet: Content Analysis of Jihadist Forum Discussions, 179.

Europol. (2021). Internet Organised Crime Threat Assessment (IOCTA) 2021. Retrieved February 7, 2022, from https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021

Evan, T., Leverett, E., Ruffle, S., Coburn, A. W., Bourdeau, J., Gunaratna, R., & Ralph, D. (2017). Cyber Terrorism: Assessment of the Threat to Insurance. Retrieved September 24, 2020, from https://www.researchgate.net/publication/328530893_Cyber_Terrorism_Assessment_of_the_Threat_to_Insurance

Everton, S. S. (2008). Tracking, Destabilizing and Disrupting Dark Networks with Social Networks Analysis. Retrieved March 2, 2022, from https://calhoun.nps.edu/handle/10945/34415

FBI. (2021). Strategic Intelligence Assessment and Data on Domestic Terrorism. Retrieved February 9, 2022, from https://www.fbi.gov/file-repository/fbi-dhs-domestic-terrorism-strategic-report.pdf/view

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press.

Finklea, K. (2017). Dark Web. *Congressional Research Service*, 19.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Forum, W. E. (2020). The Global Risks Report 2020. Retrieved February 11, 2022, from https://www.weforum.org/reports/the-global-risks-report-2020/

Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10815–10820. https://doi.org/10.1073/pnas.0912671107

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Freilich, J. D., Chermak, S. M., Belli, R., Gruenewald, J., & Parkin, W. S. (2014). Introducing the United States Extremis Crime Database (ECDB). *Terrorism and Political Violence*, *26*(2), 372–384. https://doi.org/10.1080/09546553.2012.713229

Furnell, S. M., & Warren, M. J. (1999). Computer hacking and cyber terrorism: The real threats in the new millennium? *Computers & Security*, *18*(1), 28–34. https://doi.org/10.1016/S0167-4048(99)80006-6

Furnell, S. (2003). Cybercrime: Vandalizing the Information Society. In J. M. C. Lovelle, B. M. G. Rodríguez, J. E. L. Gayo, M. del Puerto Paule Ruiz, & L. J. Aguilar (Eds.), *Web Engineering* (pp. 8–16). Springer. https://doi.org/10.1007/3-540-45068-8_2

Furnell, S., Emm, D., & Papadaki, M. (2015). The challenge of measuring cyber-dependent crimes. *Computer Fraud & Security*, *2015*(10), 5–12. https://doi.org/10.1016/S1361-3723(15)30093-2

Gartenstein-Ross, D., & Barr, N. (2016). Fixing How We Fight the Islamic State's Narrative. Retrieved March 11, 2022, from https://warontherocks.com/2016/01/fixing-how-we-fight-the-islamic-states-narrative/

Gaudette, T., Scrivens, R., & Venkatesh, V. (2020). The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists. *Terrorism and Political Violence*, *0*(0), 1–18. https://doi.org/10.1080/09546553.2020.1784147

Gies, S. V., Piquero, N. L., Piquero, A. R., Green, B., & Bobnis, A. (2021). Wild, Wild Theft: Identity Crimes in the Digital Frontier. *Criminal Justice Policy Review*, *32*(6), 592–617. https://doi.org/10.1177/0887403420949650

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist Use of the Internet by the Numbers. *Criminology & Public Policy*, *16*(1), 99–117. https://doi.org/10.1111/1745-9133.12249

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Grabosky, P. (2016). The evolution of cybercrime, 2006–2016. *Cybercrime Through an Interdisciplinary Lens*. Routledge.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. Retrieved March 3, 2022, from https://www.jstor.org/stable/2776392

Greenberg, K. J. (2016). Counter-Radicalization via the Internet. *The ANNALS of the American Academy of Political and Social Science*, *668*(1), 165–179. https://doi.org/10.1177/0002716216672635

Greene, K. T., & Lucas, C. (2020). Once more, with feeling: Using sentiment analysis to improve models of relationships between non-state actors. *International Interactions*, *46*(1), 150–162. https://doi.org/10.1080/03050629.2019.1684913

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(Supplement 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Gruen, B., & Hornik, K. (2011). Topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, *40*(13), 1–30. https://doi.org/10.18637/jss.v040.i13

Gupta, A., Maynard, S. B., & Ahmad, A. (2021). The Dark Web Phenomenon: A Review and Research Agenda. *arXiv:2104.07138 [cs]*. Retrieved March 2, 2022, from http://arxiv.org/abs/2104.07138

Hamm, M. S. (2007). *Terrorism As Crime: From Oklahoma City to Al-Qaeda and Beyond*. NYU Press.

Hamm, M. S., & Van de Voorde, C. (2005). Crimes committed by terrorist groups: Theory, research, and prevention. *Trends in Organized Crime*, *9*(2), 18–50. https://doi.org/10.1007/s12117-005-1023-y

Hardy, K., & Williams, G. (2014). What is 'Cyberterrorism'? Computer and Internet Technology in Legal Definitions of Terrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 1–23). Springer. https://doi.org/10.1007/978-1-4939-0962-9_1

Hay, C., Meldrum, R., & Mann, K. (2010). Traditional Bullying, Cyber Bullying, and Deviance: A General Strain Theory Approach. *Journal of Contemporary Criminal Justice*, *26*, 130–147. https://doi.org/10.1177/1043986209359557

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, *17*(2), 209–233. https://doi.org/10.1177/1741659020917434

Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Perez-Meana, H., Olivares-Mercado, J., & Sanchez, V. (2018). Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using 1 Regularization. *Sensors*, *18*(5), 1380. https://doi.org/10.3390/s18051380

Higgins, G. E., Wolfe, S. E., & Marcum, C. D. (2008). Music Piracy and Neutralization: A Preliminary Trajectory Analysis from Short-Term Longitudinal Data. *2*(2), 13.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing [Publisher: American Association for the Advancement of Science]. *Science*, *349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

Hirschi, T. (1969). *Causes of Delinquency*. University of California Press.

Hodwitz, O. (2019). Recognizing & Resolving Issues in Terrorism Research, Data Collection, & Analysis. Retrieved March 13, 2022, from https://www.resolvenet.org/research/recognizing-resolving-issues-terrorism-research-data-collection-analysis

Hoffman, B. (2006a). *Inside Terrorism*. Columbia University Press.

Hoffman, B. (2006b). The Use of the Internet by Islamic Extremists. Retrieved September 24, 2020, from https://www.rand.org/pubs/testimonies/CT262-1.html

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Holt, T. J. (2007). Subcultural evolution? examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior*, *28*(2), 171–198. https://doi.org/10.1080/01639620601131065

Holt, T. J. (2010). Examining the Role of Technology in the Formation of Deviant Subcultures. *Social Science Computer Review*, *28*(4), 466–481. https://doi.org/10.1177/0894439309351344

Holt, T. J. (2012). Exploring the Intersections of Technology, Crime, and Terror. *Terrorism and Political Violence*, *24*(2), 337–354. https://doi.org/10.1080/09546553.2011.648350

Holt, T. J. (2013). *Cybercrime and criminological theory: Fundamental readings on hacking, piracy, theft, and harassment*.

Holt, T. J. (2016). *Cybercrime Through an Interdisciplinary Lens*. Taylor & Francis Group.

Holt, T. J., & Bossler, A. M. (2009). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. *Deviant Behavior*, *30*(1), 1–25. https://doi.org/10.1080/01639620701876577

Holt, T. J., & Bossler, A. M. (2014). An Assessment of the Current State of Cybercrime Scholarship. *Deviant Behavior*, *35*(1), 20–40. Retrieved February 4, 2022, from https://heinonline.org/HOL/P?h=hein.journals/devbh35&i=21

Holt, T. J., Cale, J., Brewer, R., & Goldsmith, A. (2021). Assessing the Role of Opportunity and Low Self-Control in Juvenile Hacking. *Crime & Delinquency*, *67*(5), 662–688. https://doi.org/10.1177/0011128720978730

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2017a). Exploring the Subculture of Ideologically Motivated Cyber-Attackers: *Journal of Contemporary Criminal Justice*. https://doi.org/10.1177/1043986217699100

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2017b). Internet-Based Radicalization as Enculturation to Violent Deviant Subcultures. *Deviant Behavior*, *38*(8), 855–869. https://doi.org/10.1080/01639625.2016.1197704

Holt, T. J., Freilich, J. D., Chermak, S. M., Mills, C., & Silva, J. (2019). Loners, Colleagues, or Peers? Assessing the Social Organization of Radicalization. *American Journal of Criminal Justice*, *44*(1), 83–105. https://doi.org/10.1007/s12103-018-9439-5

Holt, T. J., & Graves, D. C. (2007). A Qualitative Analysis of Advance Fee Fraud E-mail Schemes. *1*(1), 18.

Holt, T. J., Lee, J. R., Freilich, J. D., Chermak, S. M., Bauer, J. M., Shillair, R., & Ross, A. (2020). An Exploratory Analysis of the Characteristics of Ideologically Motivated Cyberattacks. *Terrorism and Political Violence*, 1–16. https://doi.org/10.1080/09546553.2020.1777987

Holt, T. J., Navarro, J. N., & Clevenger, S. (2020). Exploring the Moderating Role of Gender in Juvenile Hacking Behaviors. *Crime & Delinquency*, *66*(11), 1533–1555. https://doi.org/10.1177/0011128719875697

Holt, T. J., & Steinmetz, K. F. (2020). Examining the Role of Power-Control Theory and Self-Control to Account for Computer Hacking. *Crime & Delinquency*, 0011128720981892. https://doi.org/10.1177/0011128720981892

Holt, T. J., Strumsky, D., Smirnova, O., & Kilger, M. (2012). Examining the Social Networks of Malware Writers and Hackers. *6*(1), 13.

Holt, T. J., Turner, N. D., Freilich, J. D., & Chermak, S. M. (2021). Examining the Characteristics That Differentiate Jihadi-Associated Cyberattacks Using Routine Activities Theory. *Social Science Computer Review*, 08944393211023324. https://doi.org/10.1177/08944393211023324

Howell, C., Burruss, G., Maimon, D., & Sahani, S. (2019). Website Defacement and Routine Activities: Considering the Importance of Hackers' Valuations of Potential Targets. *EBCS Articles*. https://scholarworks.gsu.edu/ebcs_articles/10

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews, 168–177. https://doi.org/10.1145/1014052.1014073

Hussain, G., & Saltman, E. (2016). Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it. Retrieved March 11, 2022, from https://preventviolentextremism.info/jihad-trending-comprehensive-analysis-online-extremism-and-how-counter-it

Hutchings, A., & Chua, Y. T. (2016). Gendering cybercrime. *Cybercrime Through an Interdisciplinary Lens*. Routledge.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. Retrieved April 14, 2022, from https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Hvitfeldt, E., & Silge, J. (2021). *Supervised Machine Learning for Text Analysis in R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781003093459

IC3. (2020). Internet Crime Report 2020. https://www.ic3.gov/

ICE. (2020). Global disruption of 3 terror finance cyber-enabled campaigns. Retrieved February 10, 2022, from https://www.ice.gov/news/releases/global-disruption-3-terror-finance-cyber-enabled-campaigns

Jaishankar, K. (2007). Establishing a Theory of Cyber Crimes. *International Journal of Cyber Criminology*, *1*(2), 3.

Jang, H., Song, J., & Kim, R. (2014). Does the offline bully-victimization influence cyberbullying behavior among youths? Application of General Strain Theory. *Computers in Human Behavior*, *31*, 85–93. https://doi.org/10.1016/j.chb.2013.10.007

Jarvis, L., & Macdonald, S. (2014). Locating Cyberterrorism: How Terrorism Researchers Use and View the Cyber Lexicon. *Perspectives on Terrorism*, *8*(2), 52–65. Retrieved February 4, 2022, from https://www.jstor.org/stable/26297136

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jordan, T., & Taylor, P. (1998). A Sociology of Hackers. *The Sociological Review*, *46*(4), 757–780. https://doi.org/10.1111/1467-954X.00139

Kemp, S. (2019). Digital 2019: Global Digital Overview. Retrieved February 11, 2022, from https://datareportal.com/reports/digital-2019-global-digital-overview

Kemp, S. (2022). Digital 2022: Global Overview Report. Retrieved February 23, 2022, from https://datareportal.com/reports/digital-2022-global-overview-report

Kerstens, J., & Veenstra, S. (2016). Cyber Bullying In The Netherlands: A Criminological Perspective. https://doi.org/10.5281/ZENODO.55055

Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, *1*(1), 3–14. https://doi.org/10.1007/s42001-017-0006-5

Kigerl, A. (2011). Routine Activity Theory and the Determinants of High Cybercrime Countries. *Social Science Computer Review*, 0894439311422689. https://doi.org/10.1177/0894439311422689

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, *26*(1), 89–120. https://doi.org/10.1007/s11948-018-00081-0

Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 431–438.

Koehler, D. (2014). The Radical Online: Individual Radicalization Processes and the Role of the Internet. *Journal for Deradicalization*, (1), 116–134. Retrieved January 24, 2022, from https://journals.sfu.ca/jd/index.php/jd/article/view/8

Kopp, C., Layton, R., Sillitoe, J., & Gondal, I. (2016). The Role Of Love Stories In Romance Scams: A Qualitative Analysis Of Fraudulent Profiles. https://doi.org/10.5281/ZENODO.56227

Koschade, S. (2006). A Social Network Analysis of Jemaah Islamiyah: The Applications to Counterterrorism and Intelligence. *Studies in Conflict & Terrorism*, *29*(6), 559–575. https://doi.org/10.1080/10576100600798418

Krebs, V. E. (2002). Mapping Networks of Terrorist Cells. *Connections*, *24*(3), 10.

LaFree, G., & Ackerman, G. (2009). The Empirical Study of Terrorism: Social and Legal Research. *Annual Review of Law and Social Science*, *5*(1), 347–374. https://doi.org/10.1146/annurev.lawsocsci.093008.131517

LaFree, G., & Dugan, L. (2009). Research on Terrorism and Countering Terrorism. *Crime and Justice*, *38*(1), 413–477. https://doi.org/10.1086/599201

LaFree, G., & Dugan, L. (2015). How Has Criminology Contributed to the Study of Terrorism since 9/11? *Terrorism and Counterterrorism Today* (pp. 1–23). Emerald Group Publishing Limited. Retrieved September 24, 2020, from https://doi.org/10.1108/S1521-613620150000020002

Laqueur, W. (2004). *No End to War: Terrorism in the Twenty-First Century*. A&C Black.

Lazer, D., Pentland, A., Watts, D., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science (New York, N.Y.)*, *369*, 1060–1062. https://doi.org/10.1126/science.aaz8170

Ledesma, J. (2021). What is RaaS and why is it so dangerous? Retrieved February 7, 2022, from https://businessinsights.bitdefender.com/what-is-raas-and-why-is-it-so-dangerous

Lee, J. R., & Holt, T. J. (2020). Assessing the Factors Associated With the Detection of Juvenile Hacking Behaviors. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00840

Leukfeldt, E. R., & Yar, M. (2016). Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. *Deviant Behavior*, *37*(3), 263–280. https://doi.org/10.1080/01639625.2015.1012409

Li, C. K. W., Holt, T. J., Bossler, A. M., & May, D. C. (2016). Examining the Mediating Effects of Social Learning on the Low Self-Control—Cyberbullying Relationship in a Youth Sample. *Deviant Behavior*, *37*(2), 126–138. https://doi.org/10.1080/01639625.2014.1004023

Ligon, G. S., Logan, M., Hall, M., Derrick, D., Fuller, J., & Church, S. (2017). The Jihadi Industry: Assessing the Organizational, Leadership, and Cyber Profiles. *START*.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1), 1608. https://doi.org/10.1186/s40064-016-3252-8

Louderback, E. R., & Antonaccio, O. (2020). New Applications of Self-Control Theory to Computer-Focused Cyber Deviance and Victimization: A Comparison of Cognitive and Behavioral Measures of Self-Control and Test of Peer Cyber Deviance and Gender as Moderators. *Crime & Delinquency*, 0011128720906116. https://doi.org/10.1177/0011128720906116

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, *23*(2), 254–277. https://doi.org/10.1093/pan/mpu019

Maalem Lahcen, R. A., Caulkins, B., Mohapatra, R., & Kumar, M. (2020). Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*, *3*(1), 10. https://doi.org/10.1186/s42400-020-00050-w

Macdonald, S., Correia, S. G., & Watkin, A.-L. (2019). Regulating terrorist content on social media: Automation and the rule of law. *International Journal of Law in Context*, *15*(2), 183–197. https://doi.org/10.1017/S1744552319000119

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2-3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Maimon, D., & Louderback, E. R. (2019). Cyber-Dependent Crimes: An Interdisciplinary Review. *Annual Review of Criminology*, *2*(1), 191–216. https://doi.org/10.1146/annurev-criminol-032317-092057

Maimon, D., Sobesto, B., & Cukier, M. (2013). Restrictive Deterrent Effects of a Warning Banner in an Attacked Computer System. *EBCS Articles*. https://scholarworks.gsu.edu/ebcs_articles/7

Malik, N. (2018). Terror in The Dark: How Terrorists use Encryption, the Darknet and Cryptocurrencies. Retrieved February 17, 2022, from https://henryjacksonsociety.org/publications/terror-in-the-dark-how-terrorists-use-encryption-the-darknet-and-cryptocurrencies/

Manning, C., & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Marcum, C. D., & Higgins, G. E. (2012). Battle of the sexes: An examination of male and female cyber bullying. *6*(1), 8.

Marcum, C. D., Higgins, G. E., Ricketts, M. L., & Wolfe, S. E. (2014). Hacking in High School: Cybercrime Perpetration by Juveniles. *Deviant Behavior*, *35*(7), 581–591. https://doi.org/10.1080/01639625.2013.867721

Marwick, A. E., & boyd danah, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

McCauley, C., & Moskalenko, S. (2008). Mechanisms of Political Radicalization: Pathways Toward Terrorism. *Terrorism and Political Violence*, *20*(3), 415–433. https://doi.org/10.1080/09546550802073367

Milla, M. N., Hudiyana, J., Cahyono, W., & Muluk, H. (2020). Is the Role of Ideologists Central in Terrorist Networks? A Social Network Analysis of Indonesian Terrorist Groups. *Frontiers in Psychology*, *11*. Retrieved February 28, 2022, from https://www.frontiersin.org/article/10.3389/fpsyg.2020.00333

Miller, B., & Morris, R. G. (2016). Virtual Peer Effects in Social Learning Theory. *Crime & Delinquency*, *62*(12), 1543–1569. https://doi.org/10.1177/0011128714526499

Mirea, M., Wang, V., & Jung, J. (2019). The not so dark side of the darknet: A qualitative study. *Security Journal*, *32*(2), 102–118. https://doi.org/10.1057/s41284-018-0150-5

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *arXiv:1308.6297*. Retrieved December 2, 2018, from http://arxiv.org/abs/1308.6297

Molina, M., & Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology*, *45*(1), 27–45. https://doi.org/10.1146/annurev-soc-073117-041106

Monk, B., Mitchell, J., Frank, R., & Davies, G. (2018). Uncovering Tor: An Examination of the Network Structure. *Security and Communication Networks*, *2018*, e4231326. https://doi.org/10.1155/2018/4231326

Morgan, S. (2021). Cybercrime To Cost The World $10.5 Trillion Annually By 2025. Retrieved February 11, 2022, from https://cybersecurityventures.com/cyberwarfare-report-intrusion/

Morris, R. G. (2011). Computer Hacking and the Techniques of Neutralization: An Empirical Assessment. Retrieved February 5, 2022, from https://www.igi-global.com/

chapter/computer-hacking-techniques-neutralization/www.igi-global.com/chapter/computer-hacking-techniques-neutralization/46417

Morris, R. G., & Higgins, G. E. (2010). Criminological theory in the digital age: The case of social learning theory and digital piracy. *Journal of Criminal Justice*, *38*(4), 470–480. https://doi.org/10.1016/j.jcrimjus.2010.04.016

Mullins, S. (2013). Social network analysis and terrorism: An introduction to the special issue. *Behavioral Sciences of Terrorism and Political Aggression*, *5*(2), 67–69. https://doi.org/10.1080/19434472.2012.731697

Németh, R., & Koltai, J. (2021). The Potential of Automated Text Analytics in Social Knowledge Building. In T. Rudas & G. Péli (Eds.), *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations* (pp. 49–70). Springer International Publishing. https://doi.org/10.1007/978-3-030-54936-7_3

Newman, M. (2006). Modularity and community structure in networks. Retrieved March 3, 2022, from https://www.pnas.org/doi/abs/10.1073/pnas.0601602103

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv:1103.2903 [cs]*. Retrieved April 14, 2022, from http://arxiv.org/abs/1103.2903

Nikita, M., & Chaney, N. (2020). Ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=ldatuning

Nodeland, B., Belshaw, S., & Saber, M. (2018). Teaching Cybersecurity to Criminal Justice Majors. *Journal of Criminal Justice Education*, *30*, 1–20. https://doi.org/10.1080/10511253.2018.1439513

Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in Social Media. *Data Mining and Knowledge Discovery*, *24*(3), 515–554. https://doi.org/10.1007/s10618-011-0224-z

Payne, B. K., & Hadzhidimova, L. (2020). Disciplinary and Interdisciplinary Trends in Cybercrime Research: An Examination. https://doi.org/10.5281/ZENODO.3741131

Pelletier, I. R., Lundmark, L., Gardner, R., Ligon, G. S., & Kilinc, R. (2016). Why ISIS's Message Resonates: Leveraging Islam, Sociopolitical Catalysts, and Adaptive Messaging. *Studies in Conflict & Terrorism*, *39*(10), 871–899. https://doi.org/10.1080/1057610X.2016.1139373

Perliger, A., & Pedahzur, A. (2011). Social Network Analysis in the Study of Terrorism and Political Violence. *PS: Political Science and Politics*, *44*(1), 45–50. Retrieved February 25, 2021, from https://www.jstor.org/stable/40984482

Pete, I., Hughes, J., Chua, Y. T., & Bada, M. (2020). A Social Network Analysis and Comparison of Six Dark Web Forums. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, 484–493. https://doi.org/10.1109/EuroSPW51379.2020.00071

Phillips, E., Nurse, J., Goldsmith, M., & Creese, S. (2015). Extracting Social Structure from DarkWeb Forums.

Piazza, J. A., & Guler, A. (2019). The Online Caliphate: Internet Usage and ISIS Support in the Arab World. *Terrorism and Political Violence*, *0*(0), 0–20. https://doi.org/10.1080/09546553.2019.1606801

Pollitt, M. M. (1998). Cyberterrorism — fact or fancy? *Computer Fraud & Security*, *1998*(2), 8–10. https://doi.org/10.1016/S1361-3723(00)87009-8

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks (long version). *arXiv:physics/0512106*. Retrieved February 25, 2022, from http://arxiv.org/abs/physics/0512106

Pons, P., & Latapy, M. (2006). Computing Communities in Large Networks Using Random Walks. *J. Graph Algorithms Appl.*, *10*, 191–218. https://doi.org/10.7155/jgaa.00124

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society*. Routledge. https://doi.org/10.4324/9781315205786

Pozzi, F., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment analysis in social networks*. Morgan Kaufmann.

Puschmann, C., & Powell, A. (2018). Turning Words Into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media. *Social Media + Society*, *4*(3), 2056305118797724. https://doi.org/10.1177/2056305118797724

Queiroz, G. D., Fay, C., Hvitfeldt, E., Keyes, O., Misra, K., Mastny, T., Erickson, J., Robinson, D., & Silge, J. (2021). Tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=tidytext

Raab, J., & Milward, H. B. (2003). Dark Networks as Problems. *Journal of Public Administration Research and Theory: J-PART*, *13*(4), 413–439. Retrieved March 2, 2022, from https://www.jstor.org/stable/3525656

Reyns, B. W., Henson, B., & Fisher, B. S. (2011). Being Pursued Online: Applying Cyberlifestyle–Routine Activities Theory to Cyberstalking Victimization. *Criminal Justice and Behavior*, *38*(11), 1149–1169. https://doi.org/10.1177/0093854811421448

Roberts, M. E., Tingley, D., Stewart, B. M., & Airoldi, E. M. (2013). The Structural Topic Model and Applied Social Science. *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application and Evaluation.*, 4.

Romaniuk, P. (2015). Does CVE Work? Lessons Learned From the Global Effort to Counter Violent Extremism. Retrieved March 11, 2022, from https://www.globalcenter.org/publications/does-cve-work-lessons-learned-from-the-global-effort-to-counter-violent-extremism/

Rosenfeld, R. (2002). Why Criminologists Should Study Terrorism. *The Criminologist*, *27*(6), 1–4.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 487–494.

Rowe, M., & Saif, H. (2016). Mining Pro-ISIS Radicalisation Signals from Social Media Users. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, 329–338. Retrieved November 5, 2020, from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13023/12752

Roy, A., & Rambo-Hernandez, K. E. (2021). There's So Much to Do and Not Enough Time to Do It! A Case for Sentiment Analysis to Derive Meaning From Open Text Using Student Reflections of Engineering Activities. *American Journal of Evaluation*, *42*(4), 559–576. https://doi.org/10.1177/1098214020962576

Rudner, M. (2008). Misuse of Passports: Identity Fraud, the Propensity to Travel, and International Terrorism. *Studies in Conflict & Terrorism*, *31*(2), 95–110. https://doi.org/10.1080/10576100701812803

Rudner, M. (2017). "Electronic Jihad": The Internet as Al Qaeda's Catalyst for Global Terror. *Studies in Conflict & Terrorism*, *40*(1), 10–23. https://doi.org/10.1080/1057610X.2016.1157403

Sageman, M. (2004). *Understanding Terror Networks* (First Printing edition). University of Pennsylvania Press.

Sageman, M. (2011). *Leaderless Jihad: Terror Networks in the Twenty-First Century*. University of Pennsylvania Press.

Salganik, M. J. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Saltman, E., Kooti, F., & Vockery, K. (2021). New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis. *Studies in Conflict & Terrorism*, *0*(0), 1–24. https://doi.org/10.1080/1057610X.2021.1888404

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The Risk of Racial Bias in Hate Speech Detection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678. https://doi.org/10.18653/v1/P19-1163

Scanlon, J., & Gerber, M. (2014). Automatic detection of cyber-recruitment by violent extremists. *Security Informatics*, *3*. https://doi.org/10.1186/s13388-014-0005-5

Scanlon, J. R., & Gerber, M. S. (2015). Forecasting Violent Extremist Cyber Recruitment. *IEEE Transactions on Information Forensics and Security*, *10*(11), 2461–2470. https://doi.org/10.1109/TIFS.2015.2464775

Schell, B. H., & Dodge, J. L. (2002). *The hacking of America: Who's doing it, why, and how*. Greenwood Publishing Group Inc.

Schmid, A. P. (2004). Frameworks for Conceptualising Terrorism. *Terrorism and Political Violence*, *16*(2), 197–221. https://doi.org/10.1080/09546550490483134

Schmid, A. P. (2014). Al-Qaeda's "Single Narrative" and Attempts to Develop Counter-Narratives: The State of Knowledge. *International Centre for Counter-Terrorism*, 40. http://www.jstor.org/stable/resrep29395

Schofield, A., & Mimno, D. (2016). Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics*, *4*, 287–300. https://doi.org/10.1162/tacl_a_00099

Schuurman, B. (2019). Topics in terrorism research: Reviewing trends and gaps, 2007-2016. *Critical Studies on Terrorism*, *12*(3), 463–480. https://doi.org/10.1080/17539153.2019.1579777

Schweinberger, M. (2022). Sentiment Analysis in R. Retrieved April 19, 2022, from https://slcladal.github.io/sentiment.html

Scott, J., & Spaniel, D. (2016). *The Anatomy of Cyber-Jihad: Cyberspace is the New Great Equalizer*. CreateSpace Independent Publishing Platform.

Scott, J. (2000). *Social Network Analysis: A Handbook*. SAGE.

Scott, J., & Carrington, P. (2014). *The SAGE Handbook of Social Network Analysis*. SAGE Publications Ltd. https://doi.org/10.4135/9781446294413

Scrivens, R., Chermak, S. M., Freilich, J. D., Wojciechowski, T. W., & Frank, a. R. (2021). Detecting Extremists Online: Examining Online Posting Behaviors of Violent and

Non-Violent Right-Wing Extremists. Retrieved September 17, 2021, from https://www.resolvenet.org/research/detecting-extremists-online-examining-online-posting-behaviors-violent-and-non-violent

Scrivens, R., Davies, G., & Frank, R. (2018). Searching for signs of extremism on the web: An introduction to Sentiment-based Identification of Radical Authors. *Behavioral Sciences of Terrorism and Political Aggression*, *10*(1), 39–59. https://doi.org/10.1080/19434472.2016.1276612

Shehabat, A., & Mitew, T. (2018). Black-boxing the Black Flag: Anonymous Sharing Platforms and ISIS Content Distribution Tactics. *Perspectives on Terrorism*, *12*(1), 81–99. Retrieved September 21, 2020, from http://www.jstor.org/stable/26343748

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach* (1st edition). O'Reilly Media.

Silke, A. (Ed.). (2004). *Research on Terrorism*. Routledge.

Singer, P. W., & Friedman, A. (2013). *Cybersecurity and Cyberwar: What Everyone Needs to Know*. Oxford University Press.

Skillicorn, D. (2010). Applying Interestingness Measures to Ansar Forum Texts. *Association for Computing Machinery*, 7. https://doi.org/10.1145/1938606.1938613

Skinner, W. F., & Fream, A. M. (1997). A Social Learning Theory Analysis of Computer Crime among College Students. *Journal of Research in Crime and Delinquency*, *34*(4), 495–518. https://doi.org/10.1177/0022427897034004005

Smith, G., Moses, L., & Chan, J. (2017). The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach. *British Journal of Criminology*, *57*, 259–274. https://doi.org/10.1093/bjc/azw096

Steinmetz, K. F., Holt, T. J., & Holt, K. M. (2020). Decoding the Binary: Reconsidering the Hacker Subculture through a Gendered Lens. *Deviant Behavior*, *41*(8), 936–948. https://doi.org/10.1080/01639625.2019.1596460

Stockton, P. N., & Golabek-Goldman, M. (2014). Prosecuting Cyberterrorists: Applying Traditional Jurisdictional Frameworks to a Modern Threat. *Policy Review*, *25*, 58.

Su, C., & Holt, T. J. (2010). Cyber bullying in Chinese Web Forums: An examination of nature and extent. *4*(1), 13.

Testa, A., Maimon, D., Sobesto, B., & Cukier, M. (2017). Illegal Roaming and File Manipulation on Target Computers: Assessing the Effect of Sanction Threats on System Trespassers' Online Behaviors. *Criminology & Public Policy*, *16*. https://doi.org/10.1111/1745-9133.12312

Tuck, H., & Silverman, T. (2016). The Counter-narrative Handbook. Retrieved March 11, 2022, from https://www.isdglobal.org/isd-publications/the-counter-narrative-handbook/

Turkle, S. (2005). *The Second Self: Computers and the Human Spirit*. MIT Press.

Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. "O'Reilly Media, Inc."

Verton, D., & Brownlow, J. (2003). *Black Ice: The Invisible Threat of Cyber-Terrorism* (1st ed.). Osborne.

Wall, D. S. (2001a). *Crime and the Internet*. Routledge.

Wall, D. S. (2001b). Cybercrimes and the Internet. *Crime and the Internet*, 1–17.

Wall, D. S. (2015). Dis-Organised Crime: Towards a Distributed Model of the Organization of Cybercrime. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2677113

Walter, D., & Ophir, Y. (2019). News Frame Analysis: An Inductive Mixed-method Computational Approach. *Communication Methods and Measures*, *13*(4), 248–266. https://doi.org/10.1080/19312458.2019.1639145

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1st edition). Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weimann, G. (n.d.). Going Darker? The Challenge of Dark Net Terrorism. *Wilson Center*, 13.

Weimann, G. (2004). *Cyberterrorism: How real is the threat?* (Vol. 31). United States Institute of Peace.

Weimann, G. (2005). Cyberterrorism: The sum of all fears? *Studies in Conflict & Terrorism*, *28*(2), 129–149.

Weimann, G. (2006). *Terror on the Internet: The New Arena, the New Challenges*. US Institute of Peace Press.

Weimann, G. (2011). Cyber-Fatwas and Terrorism. *Studies in Conflict & Terrorism*, *34*(10), 765–781. https://doi.org/10.1080/1057610X.2011.604831

Weimann, G. (2016a). Going Dark: Terrorism on the Dark Web. *Studies in Conflict & Terrorism*, *39*(3), 195–206. https://doi.org/10.1080/1057610X.2015.1119546

Weimann, G. (2016b). Terrorist Migration to the Dark Web. *Perspectives on Terrorism*, *10*(3), 40–44. Retrieved September 24, 2020, from http://www.jstor.org/stable/26297596

Welch, T. (2018). Theology, heroism, justice, and fear: An analysis of ISIS propaganda magazines Dabiq and Rumiyah. *Dynamics of Asymmetric Conflict*, *11*(3), 186–198. https://doi.org/10.1080/17467586.2018.1517943

Weulen Kranenbarg, M., Holt, T. J., & van der Ham, J. (2018). Don't shoot the messenger! A criminological and computer science perspective on coordinated vulnerability disclosure. *Crime Science*, *7*(1), 16. https://doi.org/10.1186/s40163-018-0090-8

Weulen Kranenbarg, M., Holt, T. J., & van Gelder, J.-L. (2019). Offending and Victimization in the Digital Age: Comparing Correlates of Cybercrime and Traditional Offending-Only, Victimization-Only and the Victimization-Offending Overlap. *Deviant Behavior*, *40*(1), 40–55. https://doi.org/10.1080/01639625.2017.1411030

Whittaker, J. M., & Button, M. (2020). Understanding pet scams: A case study of advance fee and non-delivery fraud using victims' accounts. *Australian & New Zealand Journal of Criminology*, *53*(4), 497–514. https://doi.org/10.1177/0004865820957077

Wickham, H. (2021). Tidyverse: Easily Install and Load the 'Tidyverse'. Retrieved February 24, 2022, from https://CRAN.R-project.org/package=tidyverse

Wilson, T., Maimon, D., Sobesto, B., & Cukier, M. (2015). The Effect of a Surveillance Banner in an Attacked Computer System: Additional Evidence for the Relevance of Restrictive Deterrence in Cyberspace. *Journal of Research in Crime and Delinquency*, *52*(6), 829–855. https://doi.org/10.1177/0022427815587761

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J. (2020). Online Extremism: Research Trends in Internet Activism, Radicalization, and Counter-Strategies. *International Journal of Conflict and Violence (IJCV)*, *14*, 1–20. https://doi.org/10.4119/ijcv-3809

Wozniak, J. S. G., Woods, J., & Lee, Y. S. (2020). The evolving self-presentation of the Islamic State, from Dabiq to Rumiyah. *The Social Science Journal*, *0*(0), 1–13. https://doi.org/10.1080/03623319.2020.1727242

Yannakogeorgos, P. A. (2014). Rethinking the Threat of Cyberterrorism. In T. M. Chen, L. Jarvis, & S. Macdonald (Eds.), *Cyberterrorism: Understanding, Assessment, and Response* (pp. 43–62). Springer. Retrieved May 12, 2021, from https://doi.org/10.1007/978-1-4939-0962-9_3

Zelin, A. Y., & Fellow, R. B. (2013). The State of Global Jihad Online: A Qualitative, Quantitative, and Cross-Lingual Analysis, 24.

Zhang, S., Leidner, D., Cao, X., & Liu, N. (2021). Workplace cyberbullying: A criminological and routine activity perspective. *Journal of Information Technology*, 02683962211027888. https://doi.org/10.1177/02683962211027888

Appendix

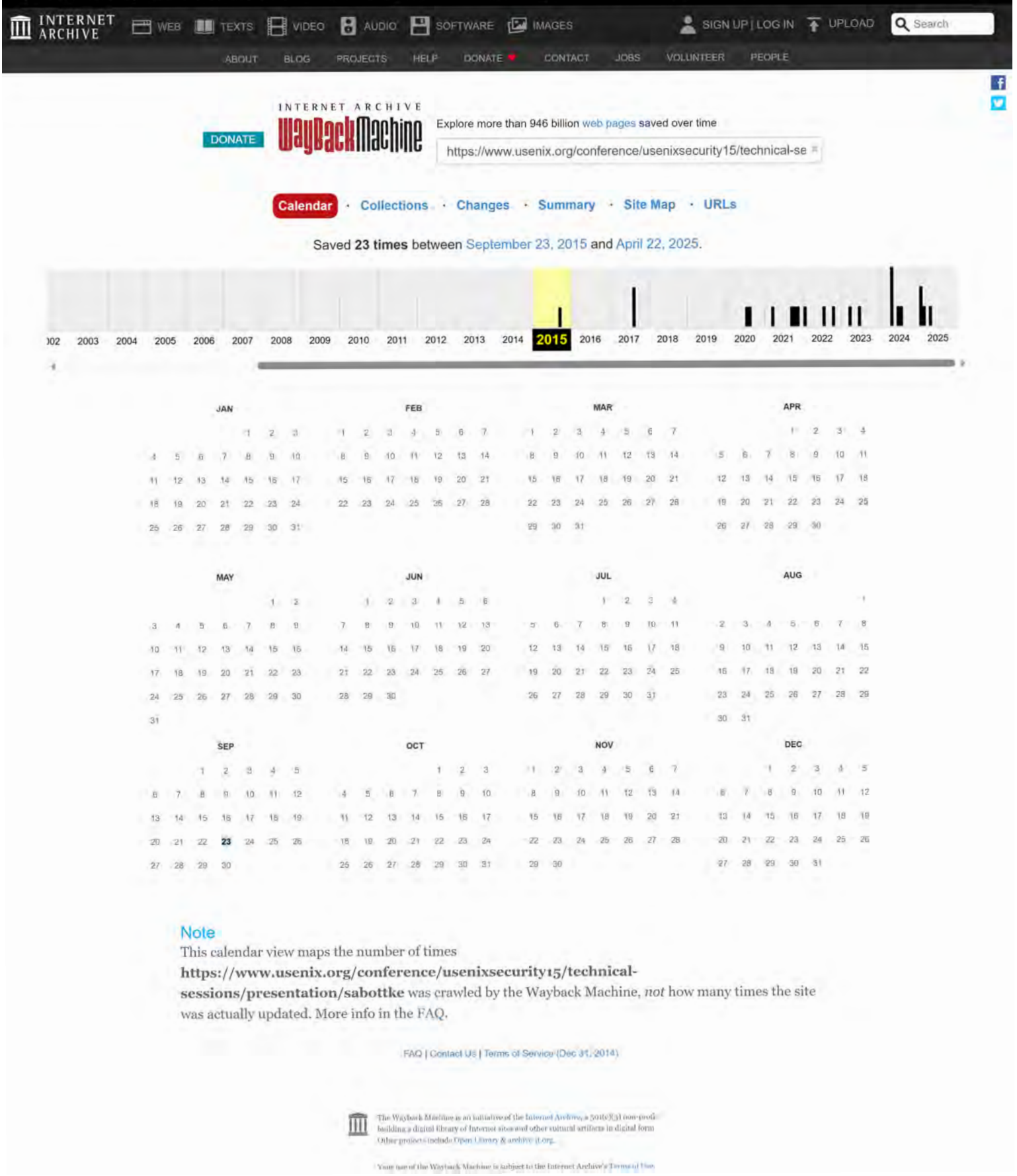
Nodes	Centrality	Eigen	Closeness	Betweenness
Tayeb	338	1	0.679266	61034.9
Lulua	255	0.896858	0.621542	29190.0
OmMohammed	207	0.775857	0.590056	20844.7
Asif	199	0.814905	0.585661	13367.7
lubna	188	0.646704	0.577594	20910.7
Rasha	182	0.73589	0.573382	13920.5
tbahrain	160	0.643549	0.56362	10398.4
BinZiad	145	0.667047	0.554674	10093.8
Nzingha	135	0.608455	0.549345	9062.2
Netcurtains3	123	0.569296	0.543178	5031.5

Table 5.1: Top 10 Nodes Centrality Scores – Myiwc Interaction Network

Nodes	Centrality	Eigen	Closeness	Betweenness
Asadullah Alshishani	294	1	0.854369	7388.6
Insurgent	247	0.90149	0.768559	5173.2
Asad'Allah	235	0.930799	0.747346	2647.5
ANSAR 007	221	0.889646	0.727273	2925.5
wa1slama	209	0.863618	0.708249	3188.9
Abu Fatima	199	0.855034	0.69428	1332.2
Guest	192	0.851984	0.684825	1106.1
abdulrahman al muhajir	191	0.819765	0.683495	1813.0
tarbiya	184	0.82232	0.67433	1057.4
Abu Umar	175	0.784745	0.6629	995.4

Table 5.2: Top 10 Nodes Centrality Scores – Ansar1 Interaction Network

https://web.archive.org/web/20151001000000*/https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke June 23, 2025



https://web.archive.org/web/20150923100812/https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke June 23, 2025

INTERNET ARCHIVE Wayback Machine https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke Go AUG SEP 23 AUG
23 captures 23 Sep 2015 - 22 Apr 2025 2014 2015 2017 About this capture



24th USENIX Security Symposium

AUGUST 12-14, 2015 • WASHINGTON, D.C.

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits

Authors:
Carl Sabottke, Octavian Suciu, and Tudor Dumitraş, *University of Maryland*

Open Access Content

USENIX is committed to Open Access to the research presented at our events. Papers and proceedings are freely available to everyone once the event begins. Any video, audio, and/or slides that are posted after the event are also free and open to everyone. Support USENIX and our commitment to Open Access.

Sabottke PDF View the slides BibTeX

Abstract:
In recent years, the number of software vulnerabilities discovered has grown significantly. This creates a need for prioritizing the response to new disclosures by assessing which vulnerabilities are likely to be exploited and by quickly ruling out the vulnerabilities that are not actually exploited in the real world. We conduct a quantitative and qualitative exploration of the vulnerability-related information disseminated on Twitter. We then describe the design of a Twitter-based exploit detector, and we introduce a threat model specific to our problem. In addition to response prioritization, our detection techniques have applications in risk modeling for cyber-insurance and they highlight the value of information provided by the victims of attacks.

Log in or Register to post comments

CONNECT WITH US
Twitter
Facebook
LinkedIn
Google+
YouTube

HELP PROMOTE
24th USENIX Security Symposium
AUGUST 12-14, 2015
WASHINGTON, D.C.
www.usenix.org/sec15
Get more Help Promote graphics!

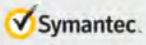
SPONSORS
Silver Sponsor
Google

TWITTER
Tweets by @usenix

USENIX CONFERENCE POLICIES
- Event Code of Conduct
- Conference Network Policy
- Statement on Environmental Responsibility Policy

Platinum Sponsors
facebook

Gold Sponsors

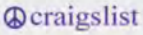



Silver Sponsors



Microsoft Research

Bronze Sponsors




General Sponsors

OXFORD
UNIVERSITY PRESS

Media Sponsors & Industry Partners











no starch press



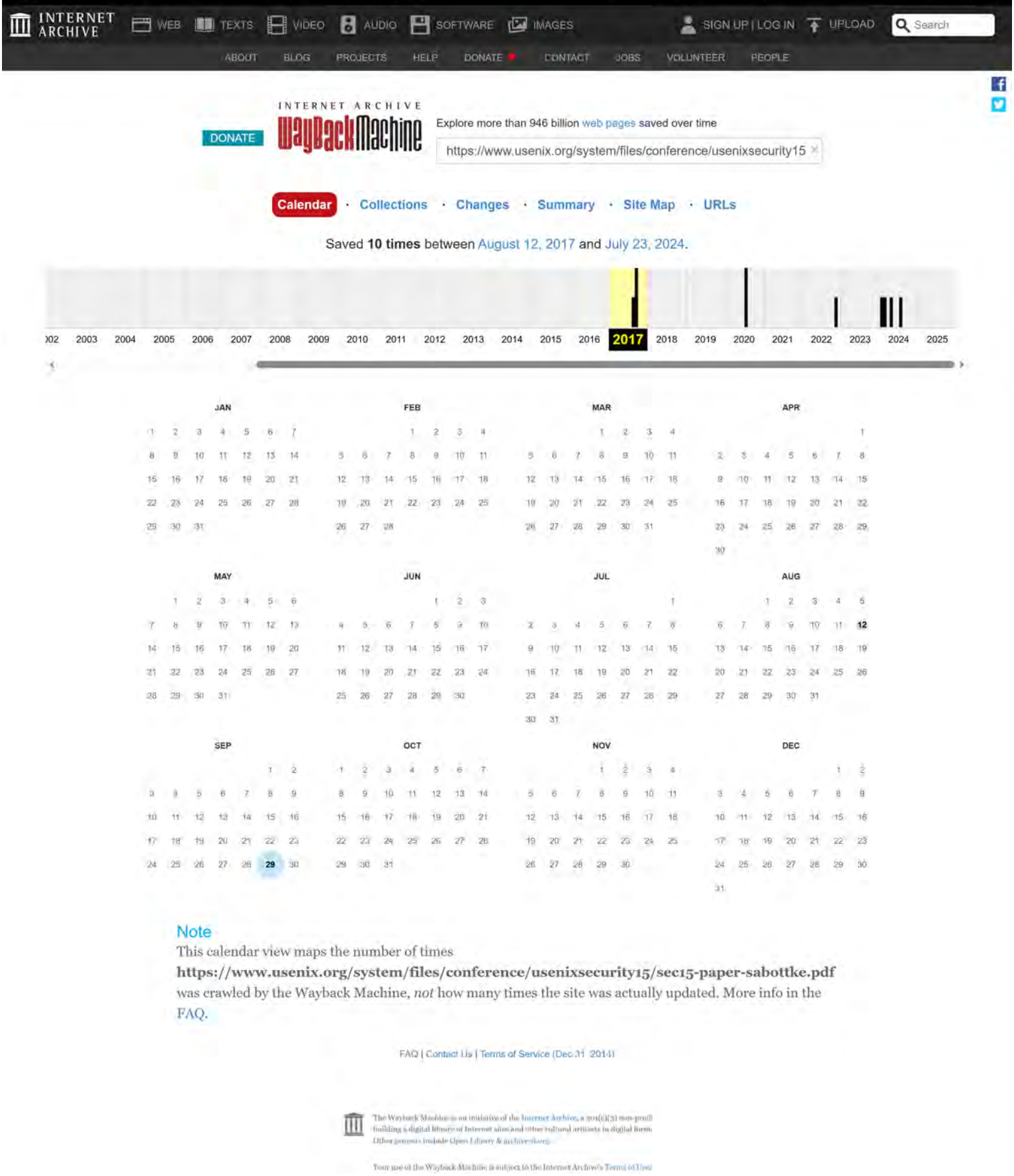
Open Access Publishing Partner





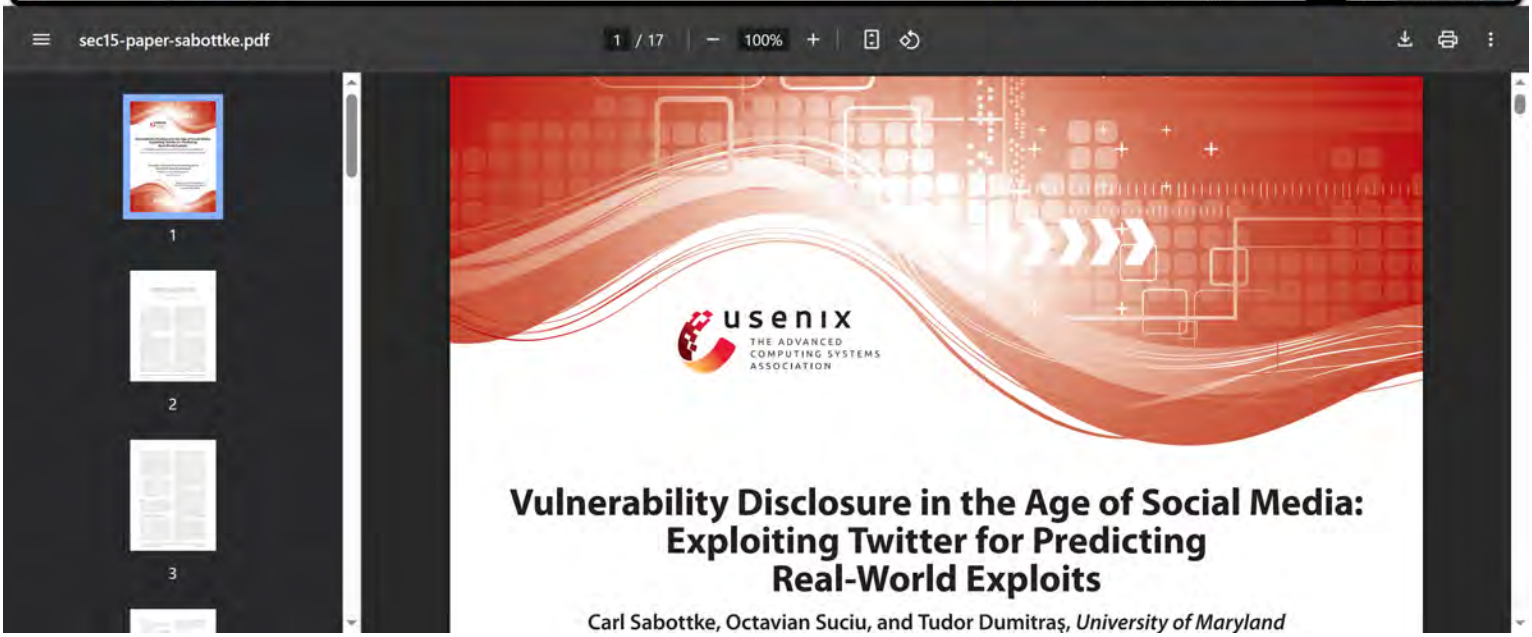
© USENIX 2015

https://web.archive.org/web/20170201000000*/https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-sabottke.pdf June 23, 2025



https://web.archive.org/web/20170812225231/https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-sabottke.pdf June 23, 2025

INTERNET ARCHIVE WayBack Machine https://www.usenix.org/system/files/conference/usenixsecurity15/sec15-paper-sabottke.pdf Go JUL AUG SEP
10 captures 12 Aug 2017 - 23 Jul 2024 2016 12 2017 2020 About this capture





Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits

Carl Sabottke, Octavian Suciu, and Tudor Dumitraş, ***University of Maryland***

https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/sabottke

This paper is included in the Proceedings of the 24th USENIX Security Symposium

August 12–14, 2015 • Washington, D.C.

ISBN 978-1-931971-232

Open access to the Proceedings of the 24th USENIX Security Symposium is sponsored by USENIX

Appendix SABOTTKE01

Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits

Carl Sabottke Octavian Suciu Tudor Dumitraș
University of Maryland

Abstract

In recent years, the number of software vulnerabilities discovered has grown significantly. This creates a need for prioritizing the response to new disclosures by assessing which vulnerabilities are likely to be exploited and by quickly ruling out the vulnerabilities that are not actually exploited in the real world. We conduct a quantitative and qualitative exploration of the vulnerability-related information disseminated on Twitter. We then describe the design of a Twitter-based exploit detector, and we introduce a threat model specific to our problem. In addition to response prioritization, our detection techniques have applications in risk modeling for cyber-insurance and they highlight the value of information provided by the victims of attacks.

1 Introduction

The number of software vulnerabilities discovered has grown significantly in recent years. For example, 2014 marked the first appearance of a 5 digit CVE, as the CVE database [46], which assigns unique identifiers to vulnerabilities, has adopted a new format that no longer caps the number of CVE IDs at 10,000 per year. Additionally, many vulnerabilities are made public through a *coordinated disclosure process* [18], which specifies a period when information about the vulnerability is kept confidential to allow vendors to create a patch. However, this process results in multi-vendor disclosure schedules that sometimes align, causing a flood of disclosures. For example, 254 vulnerabilities were disclosed on 14 October 2014 across a wide range of vendors including Microsoft, Adobe, and Oracle [16].

To cope with the growing rate of vulnerability discovery, the security community must prioritize the effort to respond to new disclosures by assessing the risk that the vulnerabilities will be exploited. The existing scoring systems that are recommended for this purpose, such as FIRST's Common Vulnerability Scoring System (CVSS) [54], Microsoft's exploitability index [21] and Adobe's priority ratings [19], err on the side of caution by marking many vulnerabilities as likely to be exploited [24]. The situation in the real world is more nuanced. While the disclosure process often produces *proof of concept* exploits, which are publicly available, recent empirical studies reported that only a small fraction of vulnerabilities are exploited *in the real world*, and this fraction has decreased over time [22,47]. At the same time, some vulnerabilities attract significant attention and are quickly exploited; for example, exploits for the Heartbleed bug in OpenSSL were detected 21 hours after the vulnerability's public disclosure [41]. To provide an adequate response on such a short time frame, the security community must quickly determine which vulnerabilities are exploited in the real world, while minimizing false positive detections.

The security vendors, system administrators, and hackers, who discuss vulnerabilities on social media sites like Twitter, constitute rich sources of information, as the participants in coordinated disclosures discuss technical details about exploits and the victims of attacks share their experiences. This paper explores the opportunities for *early exploit detection* using information available on Twitter. We characterize the exploit-related discourse on Twitter, the information posted before vulnerability disclosures, and the users who post this information. We also reexamine a prior experiment on predicting the development of proof-of-concept exploits [36] and find a considerable performance gap. This illuminates the threat landscape evolution over the past decade and the current challenges for early exploit detection.

Building on these insights, we describe techniques for detecting exploits that are active in the real world. Our techniques utilize *supervised machine learning* and ground truth about exploits from ExploitDB [3], OSVDB [9], Microsoft security advisories [21] and the descriptions of Symantec's anti-virus and intrusion-protection signatures [23]. We collect an unsampled cor-

pus of tweets that contain the keyword "CVE," posted between February 2014 and January 2015, and we extract features for training and testing a support vector machine (SVM) classifier. We evaluate the false positive and false negative rates and we assess the detection lead time compared to existing data sets. Because Twitter is an open and free service, we introduce a *threat model*, considering realistic adversaries that can poison both the training and the testing data sets but that may be resource-bound, and we conduct simulations to evaluate the resilience of our detector to such attacks. Finally, we discuss the implications of our results for building security systems without secrets, the applications of early exploit detection and the value of sharing information about successful attacks.

In summary, we make three contributions:

- We characterize the landscape of threats related to information leaks about vulnerabilities before their public disclosure, and we identify features that can be extracted automatically from the Twitter discourse to detect exploits.
- To our knowledge, we describe the first technique for early detection of real-world exploits using social media.
- We introduce a threat model specific to our problem and we evaluate the robustness of our detector to adversarial interference.

Roadmap. In Sections 2 and 3 we formulate the problem of exploit detection and we describe the design of our detector, respectively. Section 4 provides an empirical analysis of the exploit-related information disseminated on Twitter, Section 5 presents our detection results, and Section 6 evaluates attacks against our exploit detectors. Section 7 reviews the related work, and Section 8 discusses the implications of our results.

2 The problem of exploit detection

We consider a *vulnerability* to be a software bug that has security implications and that has been assigned a unique identifier in the CVE database [46]. An *exploit* is a piece of code that can be used by an attacker to subvert the functionality of the vulnerable software. While many researchers have investigated the techniques for creating exploits, the utilization patterns of these exploits provide another interesting dimension to their security implications. We consider *real-world exploits* to be the exploits that are being used in real attacks against hosts and networks worldwide. In contrast, *proof-of-concept (PoC) exploits* are often developed as part of the vulnerability disclosure process and are included in penetration testing suites. We further distinguish between *public PoC exploits*, for which the exploit code is publicly available, and *private PoC exploits*, for which we can find reliable information that the exploit was developed, but it was not released to the public. A PoC exploit may also be a real-world exploit if it is used in attacks.

The existence of a real-world or PoC exploit gives urgency to fixing the corresponding vulnerability, and this knowledge can be utilized for prioritizing remediation actions. We investigate the opportunities for *early detection* of such exploits by using information that is available publicly, but is not included in existing vulnerability databases such as the National Vulnerability Database (NVD) [7] or the Open Sourced Vulnerability Database (OSVDB) [9]. Specifically, we analyze the Twitter stream, which exemplifies the information available from social media feeds. On Twitter, a community of hackers, security vendors and system administrators discuss security vulnerabilities. In some cases, the victims of attacks report new vulnerability exploits. In other cases, information leaks from the *coordinated disclosure* process [18] through which the security community prepares the response to the impending public disclosure of a vulnerability.

The vulnerability-related discourse on Twitter is influenced by trend-setting vulnerabilities, such as Heartbleed (CVE-2014-0160), Shellshock (CVE-2014-6271, CVE-2014-7169, and CVE-2014-6277) or Drupalgeddon (CVE-2014-3704) [41]. Such vulnerabilities are mentioned by many users who otherwise do not provide actionable information on exploits, which introduces a significant amount of noise in the information retrieved from the Twitter stream. Additionally, adversaries may inject fake information into the Twitter stream, in an attempt to poison our detector. Our *goals* in this paper are (i) to identify the good sources of information about exploits and (ii) to assess the opportunities for early detection of exploits in the presence of benign and adversarial noise. Specifically, we investigate techniques for minimizing *false-positive detections*—vulnerabilities that are not actually exploited—which is critical for prioritizing response actions.

Non-goals. We do not consider the detection of *zero-day attacks* [32], which exploit vulnerabilities before their public disclosure; instead, we focus on detecting the use of exploits against known vulnerabilities. Because our aim is to assess the value of publicly available information for exploit detection, we do not evaluate the benefits of incorporating commercial or private data feeds. The design of a complete system for early exploit detection, which likely requires mechanisms beyond the realm of Twitter analytics (e.g., for managing the reputation of data sources to prevent poisoning attacks), is also out of scope for this paper.

2.1 Challenges

To put our contributions in context, we review the three primary challenges for predicting exploits in the absence of adversarial interference: class imbalance, data scarcity, and ground truth biases.

Class imbalance. We aim to train a classifier that produces binary predictions: each vulnerability is classified as either exploited or not exploited. If there are significantly more vulnerabilities in one class than in the other class, this biases the output of supervised machine learning algorithms. Prior research on predicting the existence of proof-of-concept exploits suggests that this bias is not large, as over half of the vulnerabilities disclosed before 2007 had such exploits [36]. However, few vulnerabilities are exploited in the real world and the exploitation ratios tend to decrease over time [47]. In consequence, our data set exhibits a severe class imbalance: we were able to find evidence of real-world exploitation for only 1.3% of vulnerabilities disclosed during our observation period. This class imbalance represents a significant challenge for simultaneously reducing the false positive and false negative detections.

Data scarcity. Prior research efforts on Twitter analytics have been able to extract information from millions of tweets, by focusing on popular topics like movies [27], flu outbreaks [20, 26], or large-scale threats like spam [56]. In contrast, only a small subset of Twitter users discuss vulnerability exploits (approximately 32,000 users), and they do not always mention the CVE numbers in their tweets, which prevents us from identifying the vulnerability discussed. In consequence, 90% of the CVE numbers disclosed during our observation period appear in fewer than 50 tweets. Worse, when considering the known real-world exploits, close to half have fewer than 50 associated tweets. This data scarcity compounds the challenge of class imbalance for reducing false positives and false negatives.

Quality of ground truth. Prior work on Twitter analytics focused on predicting quantities for which good predictors are already available (modulo a time lag): the Hollywood Stock Exchange for movie box-office revenues [27], CDC reports for flu trends [45] and Twitter's internal detectors for highjacked accounts, which trigger account suspensions [56]. These predictors can be used as ground truth for training high-performance classifiers. In contrast, there is no comprehensive data set of vulnerabilities that are exploited in the real world. We employ as ground truth the set of vulnerabilities mentioned in the descriptions of Symantec's anti-virus and intrusion-protection signatures, which is, reportedly, the best available indicator for the exploits included in exploit kits [23, 47]. However, this dataset has coverage biases, since Symantec does not cover all platforms and products uniformly. For example, since Symantec does not provide a security product for Linux, Linux kernel vulnerabilities are less likely to appear in our ground truth dataset than exploits targeting software that runs on the Windows platform.

2.2 Threat model

Research in adversarial machine learning [28, 29], distinguishes between exploratory attacks, which poison the testing data, and causative attacks, which poison both the testing and the training data sets. Because Twitter is an open and free service, causative adversaries are a realistic threat to a system that accepts inputs from all Twitter users. We assume that these adversaries cannot prevent the victims of attacks from tweeting about their observations, but they can inject additional tweets in order to compromise the performance of our classifier. To test the ramifications of these causative attacks, we develop a threat model with three types of adversaries.

Blabbering adversary. Our weakest adversary is not aware of the statistical properties of the training features or labels. This adversary simply sends tweets with random CVEs and random security-related keywords.

Word copycat adversary. A stronger adversary is aware of the features we use for training and has access to our ground truth (which comes from public sources). This adversary uses fraudulent accounts to manipulate the word features and total tweet counts in the training data. However, this adversary is resource constrained and cannot manipulate any user statistics which would require either more expensive or time intensive account acquisition and setup (e.g., creation date, verification, follower and friend counts). The copycat adversary crafts tweets by randomly selecting pairs of non-exploited and exploited vulnerabilities and then sending tweets, so that the word feature distributions between these two classes become nearly identical.

Full copycat adversary. Our strongest adversary has full knowledge of our feature set. Additionally, this adversary has sufficient time and economic resources to purchase or create Twitter accounts with arbitrary user statistics, with the exception of verification and the account creation date. Therefore, the full copycat adversary can use a set of fraudulent Twitter accounts to fully manipulate almost all word and user-based features, which creates scenarios where relatively benign CVEs and real-world exploit CVEs appear to have nearly identical Twitter traffic at an abstracted statistical level.

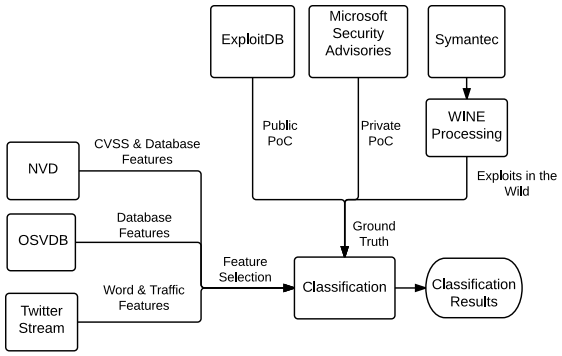


Figure 1: Overview of the system architecture.

3 A Twitter-based exploit detector

We present the design of a Twitter-based exploit detector, using supervised machine learning techniques. Our detector extracts vulnerability-related information from the Twitter stream, and augments it with additional sources of data about vulnerabilities and exploits.

3.1 Data collection

Figure 1 illustrates the architecture of our exploit detector. Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets", which then become publicly available. For collecting tweets mentioning vulnerabilities, the system monitors occurrences of the "CVE" keyword using Twitter's Streaming API [15]. The policy of the Streaming API implies that a client receives all the tweets matching a keyword as long as the result does not exceed 1% of the entire Twitter hose, when the tweets become samples of the entire matching volume. Because the CVE tweeting volume is not high enough to reach 1% of the hose (as the API signals rate limiting), we conclude that our collection contains all references to CVEs, except during the periods of downtime for our infrastructure.

We collect data over a period of one year, from February 2014 to January 2015. Out of the 1.1 billion tweets collected during this period, 287,717 contain explicit references to CVE IDs. We identify 7,560 distinct CVEs. After filtering out the vulnerabilities disclosed before the start of our observation period, for which we have missed many tweets, we are left with 5,865 CVEs.

To obtain context about the vulnerabilities discussed on Twitter, we query the National Vulnerability Database (NVD) [7] for the CVSS scores, the products affected and additional references about these vulnerabilities. Additionally, we crawl the Open Sourced Vulnerability Database (OSVDB) [9] for a few additional attributes, including the disclosure dates and categories of the vulnerabilities in our study.[1] Our data collection infrastructure consists of Python scripts, and the data is stored using Hadoop Distributed File System. From the raw data collected, we extract multiple features using Apache PIG and Spark, which run on top of a local Hadoop cluster.

Ground truth. We use three sources of ground truth. We identify the set of vulnerabilities *exploited in the real world* by extracting the CVE IDs mentioned in the descriptions of Symantec's anti-virus (AV) signatures [12] and intrusion-protection (IPS) signatures [13]. Prior work has suggested that this approach produces the best available indicator for the vulnerabilities targeted in exploits kits available on the black market [23, 47]. Considering only the vulnerabilities included in our study, this data set contains 77 vulnerabilities targeting products from 31 different vendors. We extract the creation date from the descriptions of AV signatures to estimate the date when the exploits were discovered. Unfortunately, the IPS signatures do not provide this information, so we query Symantec's Worldwide Intelligence Network Environment (WINE) [40] for the dates when these signatures were triggered in the wild. For each real-world exploit, we use the earliest date across these data sources as an estimate for the date when the exploit became known to the security community.

However, as mentioned in Section 2.1, this ground truth does not cover all platforms and products uniformly. Nevertheless, we expect that some software vendors, which have well established procedures for coordinated disclosure, systematically notify security companies of impending vulnerability disclosures to allow them to release detection signatures on the date of disclosure. For example, the members of Microsoft's MAPP program [5] receive vulnerability information in advance of the monthly publication of security advisories. This practice provides defense-in-depth, as system administrators can react to vulnerability disclosures either by deploying the software patches or by updating their AV or IPS signatures. To identify which products are well covered in this data set, we group the exploits by the vendor of the affected product. Out of the 77 real-world exploits, 41 (53%) target products from Microsoft and Adobe, while no other vendor accounts for more than 3% of exploits. This suggests that our ground truth provides the best coverage for vulnerabilities in Microsoft and Adobe products.

We identify the set of vulnerabilities with *public proof-of-concept exploits* by querying ExploitDB [3], a collaborative project that collects vulnerability exploits. We

[1] In the past, OSVDB was called the Open Source Vulnerability Database and released full dumps of their database. Since 2012, OSVDB no longer provides public dumps and actively blocks attempts to crawl the website for most of the information in the database.

identify exploits for 387 vulnerabilities disclosed during our observation period. We use the date when the exploits were added to ExploitDB as an indicator for when the vulnerabilities were exploited.

We also identify the set of vulnerabilities in Microsoft's products for which *private proof-of-concept* exploits have been developed by using the Exploitability Index [21] included in Microsoft security advisories. This index ranges from 0 to 3: 0 for vulnerabilities that are known to be exploited in the real world at the time of release for a security bulletin,[2] and 1 for vulnerabilities that allowed the development of exploits with consistent behavior. Vulnerabilities with scores of 2 and 3 are considered less likely and unlikely to be exploited, respectively. We therefore consider that the vulnerabilities with an exploitability index of 0 or 1 have an private PoC exploit, and we identify 218 such vulnerabilities. 22 of these 218 vulnerabilities are considered real-world exploits in our Symantec ground truth.

3.2 Vulnerability categories

To quantify how these vulnerabilities and exploits are discussed on Twitter, we group them into 7 categories, based on their utility for an attacker: Code Execution, Information Disclosure, Denial of Service, Protection Bypass, Script Injection, Session Hijacking and Spoofing. Although heterogeneous and unstructured, the summary field from NVD entries provides sufficient information for assigning a category to most of the vulnerabilities in the study, using regular expressions comprised of domain vocabulary.

Table 2 and Section 4 show how these categories intersect with POC and real-world exploits. Since vulnerabilities may belong to several categories (a code execution exploit could also be used in a denial of service), the regular expressions are applied in order. If a match is found for one category, the subsequent categories would not be matched.

Aditionally, the Unknown category contains vulnerabilities not matched by the regular expressions and those whose summaries explicitly state that the consequences are unknown or unspecified.

3.3 Classifier feature selection

The features considered in this study can be classified in 4 categories: Twitter Text, Twitter Statistics, CVSS Information and Database Information.

For the Twitter features, we started with a set of 1000 keywords and 12 additional features based on the distribution of tweets for the CVEs, e.g. the total number of tweets related to the CVE, the average age of the accounts posting about the vulnerability and the number of retweets associated to the vulnerability. For each of these initial features, we compute the mutual information (MI) of the set of feature values X and the class labels $Y \in \{\text{exploited}, \text{not exploited}\}$:

$$MI(Y,X) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \ln \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Mutual information, expressed in nats, compares the frequencies of values from the joint distribution $p(x,y)$ (i.e. values from X and Y that occur together) with the product of the frequencies from the two distributions $p(x)$ and $p(y)$. MI measures how much knowing X reduces uncertainty about Y, and can single out useful features suggesting that the vulnerability is exploited as well as features suggesting it is not. We prune the initial feature set by excluding all features with mutual information below 0.0001 nats. For numerical features, we estimate probability distributions using a resolution of 50 bins per feature. After this feature selection process, we are left with 38 word features for real-world exploits.

Here, rather than use a wrapper method for feature selection, we use this mutual information-based filter method in order to facilitate the combination of automatic feature selection with intuition-driven manual pruning. For example, keywords that correspond to trend-setting vulnerabilities from 2014, like Heartbleed and Shellshock, exhibit a higher mutual information than many other potential keywords despite their relation to only a small subset of vulnerabilities. Yet, such highly

Keyword	MI Wild	MI PoC	Keyword	MI Wild	MI PoC
advisory	0.0007	0.0005	ok	0.0015	0.0002
beware	0.0007	0.0005	mcafee	0.0005	0.0002
sample	0.0007	0.0005	windows	0.0012	0.0011
exploit	0.0026	0.0016	w	0.0004	0.0002
go	0.0007	0.0005	microsoft	0.0007	0.0005
xp	0.0007	0.0005	info	0.0007	X
ie	0.0015	0.0005	rce	0.0007	X
poc	0.0004	0.0006	patch	0.0007	X
web	0.0015	0.0005	piyolog	0.0007	X
java	0.0007	0.0005	tested	0.0007	X
working	0.0007	0.0005	and	X	0.0005
fix	0.0012	0.0002	rt	X	0.0005
bug	0.0007	0.0005	eset	X	0.0005
blog	0.0007	0.0005	for	X	0.0005
pc	0.0007	0.0005	redhat	X	0.0002
reading	0.0007	0.0005	kali	X	0.0005
iis	0.0007	0.0005	0day	X	0.0009
ssl	0.0005	0.0003	vs	X	0.0005
post	0.0007	0.0005	linux	X	0.0009
day	0.0015	0.0005	new	X	0.0002
bash	0.0015	0.0009			

Table 1: Mutual information provided by the reduced set of keywords with respect to both sources of ground truth data. The "X" marks in the table indicate that the respective words were excluded from the final feature set due to MI below 0.0001 nats.

[2] We do not use this score as an indicator for the existence of real-world exploits because the 0 rating is available only since August 2014, toward the end of our observation period.

specific keywords are undesirable for classification due to their transitory utility. Therefore, in order to reduce susceptibility to concept drift, we manually prune these word features for our classifiers to generate a final set of 31 out of 38 word features (listed in Table 1 with additional keyword intuition described in Section 4.1).

In order to improve performance and increase classifier robustness to potential Twitter-based adversaries acting through account hijacking and Sybil attacks, we also derive features from NVD and OSVDB. We consider all 7 CVSS score components, as well as features that proved useful for predicting proof-of-concept exploits in prior work [36], such as the number of unique references, the presence of the token BUGTRAQ in the NVD references, the vulnerability category from OSVDB and our own vulnerability categories (see Section 3.2). This gives us 17 additional features. The inclusion of these non-Twitter features is useful for boosting the classifier's resilience to adversarial noise. Figure 2 illustrates the most useful features for detecting real-world or PoC exploits, along with the corresponding mutual information.

3.4 Classifier training and evaluation

We train linear support vector machine (SVM) classifiers [35, 38, 39, 43] in a feature space with 67 dimensions that results from our feature selection step (Section 3.3). SVMs seek to determine the maximum margin hyperplane to separate the classes of exploited and non-exploited vulnerabilities. When a hyperplane cannot perfectly separate the positive and negative class samples based on the feature vectors used in training, the basic SVM cost function is modified to include a regularization penalty, C, and non-negative slack variables, ξ_i. By varying C, we explore the trade-off between false negatives and false positives in our classifiers.

We train SVM classifiers using multiple rounds of stratified random sampling. We perform sampling because of the large imbalance in the class sizes between vulnerabilities exploited and vulnerabilities not exploited. Typically, our classifier training consists of 10 random training shuffles where 50% of the available data is used for training and the remaining 50% is used for testing. We use the `scikit-learn` Python package [49] to train our classifiers.

An important caveat, though, is that our one year of data limits our ability to evaluate concept drift. In most cases, our cross-validation data is temporally intermixed with the training data, since restricting sets of training and testing CVEs to temporally adjacent blocks confounds performance losses due to concept drift with performance losses due to small sample sizes. Furthermore, performance differences between the vulnerability database features of our classifiers and those explored in [36] emphasize the benefit of periodically repeating previous experiments from the security literature in order to properly assess whether the results are subject to long-term concept drift.

Category	# CVEs	Real-World	PoC	Both
	All Data / Good Coverage			
Code Execution	1249/322	66/39	192/14	28/8
Info Disclosure	1918/59	4/0	69/5	4/0
Denial Of Service	657/17	0/0	16/1	0/0
Protection Bypass	204/34	0/0	3/0	0/0
Script Injection	683/14	0/0	40/0	0/0
Session Hijacking	167/1	0/0	25/0	0/0
Spoofing	55/4	0/0	0/0	0/0
Unknown	981/51	7/0	42/6	5/0
Total	5914/502	77/39	387/26	37/8

Table 2: CVEs Categories and exploits summary. The first sub-column represents the whole dataset, while the second sub-column is restricted to Adobe and Microsoft vulnerabilities, for which our ground truth of real-world exploits provides good coverage.

Performance metrics. When evaluating our classifiers, we rely on two standard performance metrics: *precision* and *recall*.[3] Recall is equivalent to the true positive rate: Recall $= \frac{TP}{TP+FN}$, where TP is the number of true positive classifications and FN is the number of false negatives. The denominator is the total number of positive samples in the testing data. Precision is defined as: Precision $= \frac{TP}{TP+FP}$ where FP is the total number of false positives identified by the classifier. When optimizing classifier performance based on these criteria, the relative importance of these quantities is dependent on the intended applications of the classifier. If avoiding false negatives is priority, then recall must be high. However, if avoiding false positives is more critical, then precision is the more important metric. Because we envision utilizing our classifier as a tool for prioritizing the response to vulnerability disclosures, we focus on improving the precision rather than the recall.

4 Exploit-related information on Twitter

Table 2 breaks down the vulnerabilities in our study according to the categories described in Section 3.2. 1249 vulnerabilities allowing code execution were disclosed during our observation period. 66 have real-world exploits, and 192 have public proof-of-concept exploits; the intersection of these two sets includes 28 exploits. If we consider only Microsoft and Adobe vulnerabilities, for which we expect that our ground truth has good coverage (see Section 3.1), the table shows that 322 code-

[3] We choose precision and recall because Receiver Operating Characteristic (ROC) curves can present an overly optimistic view of a classifier's performance when dealing with skewed data sets [**?**].

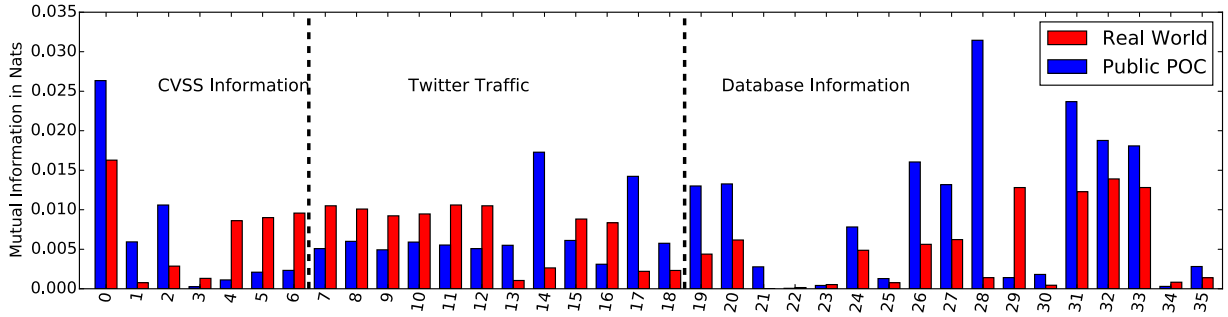


Figure 2: Mutual information between real world and public proof of concept exploits for CVSS, Twitter user statistics, NVD, and OSVDB features. **CVSS Information** - 0: CVSS Score, 1: Access Complexity, 2: Access Vector, 3: Authentication, 4: Availability Impact, 5: Confidentiality Impact, 6: Integrity Impact. **Twitter Traffic** - 7: Number of tweets, 8/9: # users with minimum T followers/friends, 10/11: # retweets/replies, 12: # tweets favorited, 13/14/15: Avg # hashtags/URLs/user mentions per tweet, 16: # verified accounts, 17: Avg age of accounts, 18: Avg # of tweets per account. **Database Information** - 19: # references in NVD, 20: # sources in NVD, 21/22: BUGTRAQ/SECUNIA in NVD sources, 23: allow in NVD summay, 24: NVD last modified date - NVD published date, 25: NVD last modified date - OSVDB disclosed date, 26: Number of tokens in OSVDB title, 27: Current date - NVD last modified date, 28: OSVDB in NVD sources, 29: code in NVD summay, 30: # OSVDB entries, 31: OSVDB Category, 32: Regex Category, 33: First vendor in NVD, 34: # vendors in NVD, 35: # affected products in NVD.

execution vulnerabilities were disclosed, 39 have real-world exploits, 14 have public PoC exploits and 8 have both real-world and public PoC exploits.

Information disclosure is the largest category of vulnerabilities from NVD and it has a large number of PoC exploits, but we find few of these vulnerabilities in our ground truth of real-world exploits (one exception is Heartbleed). Instead, most of the real-world exploits focus on *code execution* vulnerabilities. However, many proof-of-concept exploits for such vulnerabilities do not seem to be utilized in real-world attacks. To understand the factors that drive the differences between real-world and proof-of-concept exploits, we examine the CVSS base metrics, which describe the characteristics of each vulnerability. This analysis reveals that most of the real-world exploits allow *remote* code execution, while some PoC exploits require local host or local network access. Moreover, while some PoC vulnerabilities require authentication before a successful exploit, real-world exploits focus on vulnerabilities that do not require *bypassing authentication* mechanisms. In fact, this is the only type of exploit we found in the segment of our ground truth that has good coverage, suggesting that remote code-execution exploits with no authentication required are strongly favored by real-world attackers. Our ground truth does not provide good coverage of web exploits, which explains the lack of Script Injection, Session Highjacking and Spoofing exploits from our real-world data set.

Surprisingly, we find that, among the remote execution vulnerabilities for which our ground truth provides good coverage, there are *more real-world exploits than public PoC exploits*. This could be explained by the increasing prevalence of obfuscated disclosures, as reflected in NVD vulnerability summaries that mention the possibility of exploitation "via unspecified vectors" (for example, CVE-2014-8439). Such disclosures make it more difficult to create PoC exploits, as the technical information required is not readily available, but they may not thwart determined attackers who have gained experience in hacking the product in question.

4.1 Exploit-related discourse on Twitter

The Twitter discourse is dominated by a few vulnerabilities. Heartbleed (CVE-2014-0160) received the highest attention, with more than 25,000 tweets (8,000 posted in the first day after disclosure). 24 vulnerabilities received more than 1,000 tweets. 16 of these vulnerabilities were exploited: 11 in the real-world, 12 in public proofs of concept and 8 in private proofs of concept. The median number of tweets across all the vulnerabilities in our data set is 14.

The terms that Twitter users employ when discussing exploits also provide interesting insights. Surprisingly, the distribution of the keyword "0day" exhibits a high mutual information with public proof-of-concept exploits, but not with real-world exploits. This could be explained by confusion over the definition of the term *zero-day vulnerability*: many Twitter users understand this to mean simply a new vulnerability, rather than a vulnerability that was exploited in real-world attacks before its public disclosure [32]. Conversely, the distribution of the keyword "patch" has high mutual information only

with the real-world exploits, because a common reason for posting tweets about vulnerabilities is to alert other users and point them to advisories for updating vulnerable software versions after exploits are detected in the wild. Certain words like "exploit" or "advisory" are useful for detecting both real-world and PoC exploits.

4.2 Information posted before disclosure

For the vulnerabilities in our data set, Figure 3 compares the dates of the earliest tweets mentioning the corresponding CVE numbers with the public disclosure dates for these vulnerabilities. Using the disclosure date recorded in OSVDB, we identify 47 vulnerabilities that were mentioned in the Twitter stream before their public disclosure. We investigate these cases manually to determine the sources of this information. 11 of these cases represent misspelled CVE IDs (e.g. users mentioning 6172 but talking about 6271 – Shellshock), and we are unable to determine the root cause for 5 additional cases owing to the lack of sufficient information. The remaining cases can be classified into 3 general categories of information leaks:

Disagreements about the planned disclosure date. The vendor of the vulnerable software sometimes posts links to a security bulletin ahead of the public disclosure date. These cases are typically benign, as the security advisories provide instructions for patching the vulnerability. A more dangerous situation occurs when the party who discovers the vulnerability and the vendor disagree about the disclosure schedule, resulting in the publication of vulnerability details a few days before a patch is made available [6, 14, 16, 17]. We have found 13 cases of disagreements about the disclosure date.

Coordination of the response to vulnerabilities discovered in open-source software. The developers of open-source software sometimes coordinate their response to new vulnerabilities through social media, e.g. mailing lists, blogs and Twitter. An example for this behavior is a tweet about a `wget` patch for CVE-2014-4877 posted by the patch developer, followed by retweets and advice to update the binaries. If the public discussion starts before a patch is completed, then this is potentially dangerous. However, in the 5 such cases we identified, the patching recommendations were first posted on Twitter and followed by an increased retweet volume.

Leaks from the coordinated disclosure process. In some cases, the participants in the coordinated disclosure process leak information before disclosure. For example, security researchers may tweet about having confirmed

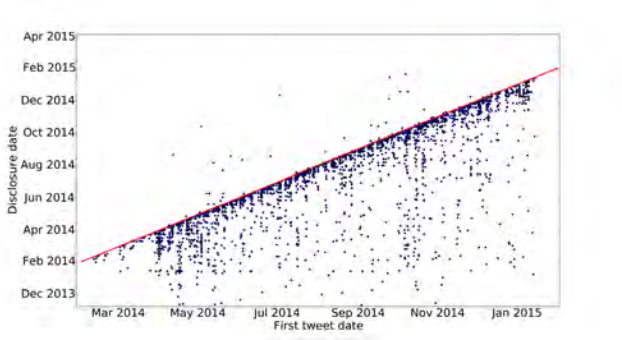


Figure 3: Comparison of the disclosure dates with the dates when the first tweets are posted for all the vulnerabilities in our dataset. Plotted in red is the identity line where the two dates coincide.

that a vulnerability is exploitable, along with the software affected. This is the most dangerous situation, as attackers may then contact the researcher with offers to purchase the exploit, before the vendor is able to release a patch. We have identified 13 such cases.

4.3 Users with information-rich tweets

The tweets we have collected were posted by approximately 32,000 unique users, but the messages posted by these users are not equally informative. Therefore, we quantify utility on a per user basis by computing the ratio of CVE tweets related to real-world exploits as well as the fraction of unique real-world exploits that a given user tweets about. We rank user utility based on the harmonic mean of these two quantities. This ranking penalizes users that tweet about many CVEs indiscriminately (e.g. a security news bot) as well as the thousands of users that only tweet about the most popular vulnerabilities (e.g. Shellshock and Heartbleed). We create a whitelist with the top 20% most informative users, and we use this whitelist in our experiments in the following sections as a means of isolating our classifier from potential adversarial attacks. Top ranked whitelist users include computer repair servicemen posting about the latest viruses discovered in their shops and security researchers and enthusiasts sharing information about the latest blog and news postings related to vulnerabilities and exploits.

Figure 4 provides an example of how the information about vulnerabilities and exploits propagates on Twitter amongst all users. The "Futex" bug, which enables unauthorized root access on Linux systems, was disclosed on June 6 as CVE-2014-3153. After identifying users who posted messages that had retweets counting for at least 1% of the total tweet counts for this vulnerability and applying a thresholding based on the number of retweets,

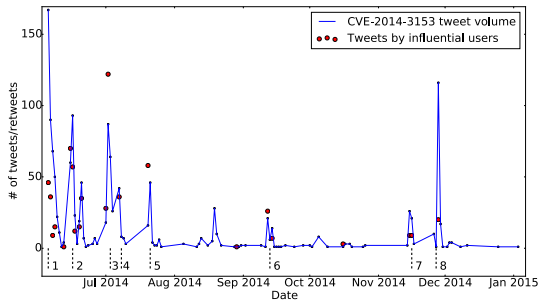


Figure 4: Tweet volume for CVE-2014-3153 and tweets from the most influential users. These influential tweets shape the volume of Twitter posts about the vulnerability. The 8 marks represent important events in the vulnerability's lifecycle: 1 - disclosure, 2 - Android exploit called Towelroot is reported and exploitation attempts are detected in the wild, 3,4,5 - new technical details emerge, including the Towelroot code, 6 - new mobile phones continue to be vulnerable to this exploit, 7 - advisory about the vulnerability is posted, 8 - exploit is included in ExploitDB.

we identify 20 influential tweets that shaped the volume of Twitter messages. These tweets correspond to 8 important milestones in the vulnerability's lifecycle, as marked in the figure.

While CVE-2014-3153 is known to be exploited in the wild, it is not included in our ground truth for real-world exploits, which does not cover the Linux platform. This example illustrates that monitoring a subset of users can yield most of the vulnerability- and exploit-related information available on Twitter. However, over reliance on a small number of user accounts, even with manual analysis, can increase susceptibility to data manipulation via adversarial account hijacking.

5 Detection of proof-of-concept and real-world exploits

To provide a baseline for our ability to classify exploits, we first examine the performance of a classifier that uses only the CVSS score, which is currently recommended as the reference assessment method for software security [50]. We use the total CVSS score and the exploitability subscore as a means of establishing baseline classifier performances. The exploitability subscore is calculated as a combination of the CVSS access vector, access complexity, and authentication components. Both the total score and exploitability subscore range from 0-10. By varying a threshold across the full range of values for each score, we can generate putative labels

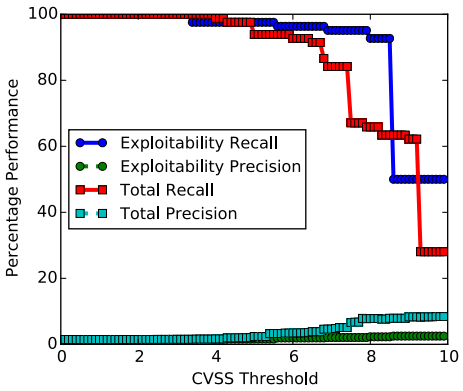
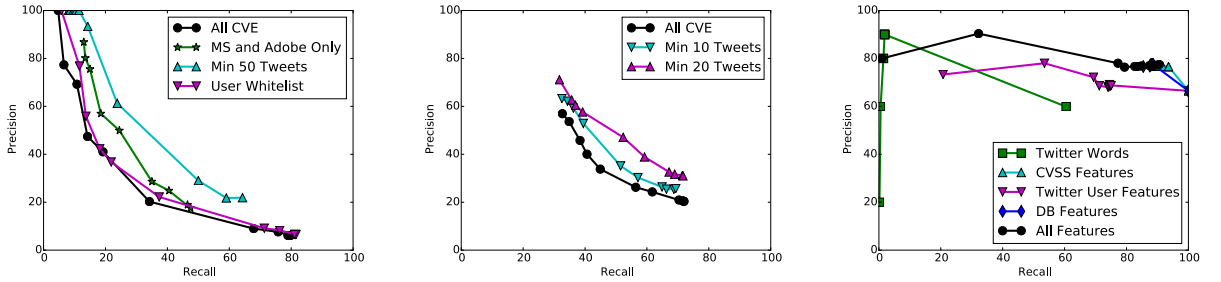


Figure 5: Precision and recall for classifying real world exploits with CVSS score thresholds.

where vulnerabilities with scores above the threshold are marked as "real-world exploits" and vulnerabilities below the threshold are labeled as "not exploited". Unsurprisingly, since CVSS is designed as a high recall system which errs on the side of caution for vulnerability severity, the maximum possible precision for this baseline classifier is less than 9%. Figure 5 shows the recall and precision values for both total CVSS score thresholds and CVSS exploitability subscore thresholds.

Thus, this high recall, low precision vulnerability score is not useful by itself for real-world exploit identification, and boosting precision is a key area for improvement.

Classifiers for real-world exploits. Classifiers for real-world exploits have to deal with a severe class imbalance: we have found evidence of real-world exploitation for only 1.3% of the vulnerabilities disclosed during our observation period. To improve the classification precision, we train linear SVM classifiers on a combination of CVSS metadata features, features extracted from security-related tweets, and features extracted from NVD and OSVDB (see Figure 2). We tune these classifiers by varying the regularization parameter C, and we illustrate the precision and recall achieved. Values shown are for cross-validation testing averaged across 10 stratified random shuffles. Figure 6a shows the average cross-validated precision and recall that are simultaneously acheivable with our Twitter-enhanced feature set. These classifiers can achieve higher precision than a baseline classifier that uses only the CVSS score, but there is still a tradeoff between precision and recall. We can tune the classifier with regularization to decrease the number of false positives (increasing precision), but this comes at the cost of a larger number of false negatives (decreasing recall).



(a) Real-world exploits.

(b) Public proof-of-concept exploits.

(c) Private proof-of-concept exploits for vulnerabilities in Microsoft products.

Figure 6: Precision and recall of our classifiers.

This result partially reflects an additional challenge for our classifier: the fact that our ground truth is imperfect, as Symantec does not have products for all the platforms that are targeted in attacks. If our Twitter-based classifier predicts that a vulnerability is exploited and the exploit exists, but is absent from the Symantec signatures, we will count this instance as a false positive, penalizing the reported precision. To assess the magnitude of this problem, we restrict the training and evaluation of the classifier to the 477 vulnerabilities in Microsoft and Adobe products, which are likely to have a good coverage in our ground truth for real-world exploits (see Section 3.1). 41 of these 477 vulnerabilities are identified as real-world exploits in our ground truth. For comparison, we include the performance of this classifier in Figure 6a. Improving the quality of the ground truth allows us to bolster the values of precision and recall which are simultaneously achievable, while still enabling classification precision an order of magnitude larger than a baseline CVSS score-based classifier. Additionally, while restricting the training of our classifier to a whitelist made up of the top 20% most informative Twitter users (as described in Section 4.3) does not enhance classifier performance, it does allow us to achieve a precision comparable to the previous experiments (Figure 6a). This is helpful for preventing an adversary from poisoning our classifier, as discussed in Section 6.

These results illustrate the current potential and limitations for predicting real-world exploits using publicly-available information. Further improvements in the classification performance may be achieved through a broader effort for sharing information about exploits active in the wild, in order to assemble a high-coverage ground truth for training of classifiers.

Classifiers for proof-of-concept exploits. We explore two classification problems: predicting *public proof-of-concept exploits*, for which the exploit code is publicly available, and predicting *private proof-of-concept exploits*, for which we can find reliable information that the exploit was developed, but it was not released to the public. We consider these problems separately, as they have different security implications and the Twitter users are likely to discuss them in distinct ways.

First, we train a classifier to predict the availability of exploits in ExploitDB [3], the largest archive of *public* exploits. This is similar to the experiment reported by Bozorgi et al. in [36], except that our feature set is slightly different—in particular, we extract word features from Twitter messages, rather than from the textual descriptions of the vulnerabilities. However, we include the most useful features for predicting proof-of-concept exploits, as reported in [36]. Additionally, Bozorgi et al. determined the availability of proof-of-concept exploits using information from OSVDB [9], which is typically populated using references to ExploitDB but may also include vulnerabilities for which the exploit is rumored or private (approximately 17% of their exploit data set). After training a classifier with the information extracted about the vulnerabilities disclosed between 1991–2007, they achieved a precision of 87.5%.

Surprisingly, we are not able to reproduce their performance results, as seen in Figure 6b, when analyzing the vulnerabilities that appear in ExploitDB in 2014. The figure also illustrates the performance of a classifier trained with exclusion threshold for CVEs that lack sufficient quantities of tweets. These volume thresholds improve performance, but not dramatically.[4] In part, this is due to our smaller data set compared to [36], made up of vulnerabilities disclosed during one year rather than a 16-year period. Moreover, our ground truth for public proof-of-concept exploits also exhibits a high class im-

[4] In this case, we do not restrict the exploits to specific vendors, as ExploitDB will incorporate any exploit submitted.

balance: only 6.2% of the vulnerabilities disclosed during our observation period have exploits available in ExploitDB. This is in stark contrast to the prior work, where more than half of vulnerabilities had proof of concept exploits.

This result provides an interesting insight into the evolution of the threat landscape: today, proof-of-concept exploits are less centralized than in 2007, and ExploitDB does not have total coverage of public proof-of-concept exploits. We have found several tweets with links to PoC exploits, published on blogs or mailing lists, that were missing from ExploitDB (we further analyze these instances in Section 4). This suggests that information about public exploits is increasingly dispersed among social media sources, rather than being included in a centralized database like ExploitDB,[5] which represents a hurdle for prioritizing the response to vulnerability disclosures.

To address this concern, we explore the potential for predicting the existence of *private* proof-of-concept exploits by considering only the vulnerabilities disclosed in Microsoft products and by using Microsoft's Exploitability Index to derive our ground truth. Figure 6c illustrates the performance of a classifier trained with a conservative ground truth in which we treat vulnerabilities with scores of 1 or less as exploits. This classifier achieves precision and recall higher than 80%, even when only relying on database feature subsets. Unlike in our prior experiments, for the Microsoft Exploitability Index the classes are more balanced: 67% of the vulnerabilities (218 out of 327) are labeled as having a private proof-of-concept exploit. However, only 8% of the Microsoft vulnerabilities in our dataset are contained within our real-world exploit ground truth. Thus, by using a conservative ground truth that labels many vulnerabilities as exploits we can achieve high precision and recall, but this classifier performance does not readily translate to real-world exploit prediction.

Contribution of various feature groups to the classifier performance. To understand how our features contribute to the performance of our classifiers, in Figure 7 we compare the precision and recall of our real-world exploit classifier when using different subgroups of features. In particular, incorporating Twitter data into the classifiers allows for improving the precision beyond the levels of precision achievable with data that is currently available publicly in vulnerability databases. Both user features and word features generated based on tweets are capable of bolstering classifier precision in comparison to CVSS and features extracted from NVD and OSVDB. Consequently, the analysis of social media streams like Twitter is useful for boosting identification of exploits active in the real-world.

[5] Indeed, OSVDB no longer seems to provide the exploitation availability flag.

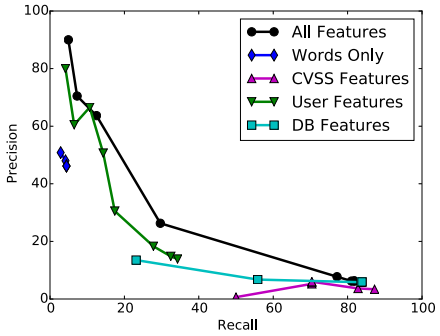


Figure 7: Precision and recall for classification of real world exploits with different feature subsets. Twitter features allow higher-precision classification of real-world exploits.

5.1 Early detection of exploits

In this section we ask the question: *How soon can we detect exploits active in the real world by monitoring the Twitter stream?* Without rapid detecton capabilities that leverage the real-time data availability inherent to social media platforms, a Twiter-based vulnerability classifier has little practical value. While the first tweets about a vulnerability precede the creation of IPS or AV signatures by a median time of 10 days, these first tweets are typically not informative enough to determine that the vulnerability is likely exploited. We therefore simulate a scenario where our classifier for real-world exploits is used in an online manner, in order to identify the dates when the output of the classifier changes from "not exploited" to "exploited" for each vulnerability in our ground truth.

We draw 10 stratified random samples, each with 50% coverage of "exploited" and "not exploited" vulnerabilities, and we train a separate linear SVM classifier with each one of these samples ($C = 0.0003$). We start testing each of our 10 classifiers with a feature set that does not include features extracted from tweets, to simulate the activation of the online classifier. We then continue to test the ensemble of classifiers incrementally, by adding one tweet at a time to the testing set. We update the aggregated prediction using a moving average with a window of 1000 tweets. Figure 8a highlights the tradeoff between precision and early detection for a range of aggregated SVM prediction thresholds. Notably, though, large precision sacrifices do not necessarily lead to large

gains for the speed of detection. Therefore, we choose an aggregated prediction threshold of 0.95, which achieves 45% precision and a median lead prediction time of two days before the first Symantec AV or WINE IPS signature dates. Figure 8b shows how our classifier detection lags behind first tweet appearances. The solid blue line shows the cumulative distribution function (CDF) for the number of days difference between the first tweet appearance for a CVE and the first Symantec AV or IPS attack signature. The green dashed line shows the CDF for the day difference between 45% precision classification and the signature creation. Negative day differences indicate that Twitter events occur before the creation of the attack signature. In approximately 20% of cases, early detection is impossible because the first tweet for a CVE occurs after an attack signature has been created. For the remaining vulnerabilities, Twitter data provides valuable insights into the likelihood of exploitation.

Figure 8c illustrates early detection for the case of Heartbleed (CVE-2014-1060). The green line indicates the first appearance of the vulnerability in ExploitDB, and the red line indicates the date when a Symantec attack signature was published. The dashed black line represents the earliest time when our online classifier is able to detect this vulnerability as a real-world exploit at 45% precision. Our Twitter-based classifier provides an "exploited" output 3 hours after the first tweet appears related to this CVE on April 7, 2014. Heartbleed exploit traffic was detected 21 hours after the vulnerability's public disclosure [41]. Heartbleed appeared in ExploitDB on the day after disclosure (April 8, 2014), and Symantec published the creation of an attack signature on April 9, 2014. Additionally, by accepting lower levels of precision, our Twitter-based classifiers can achieve even faster exploit detection. For example, with classifier precision set to approximately 25%, Heartbleed can be detected as an exploit within 10 minutes of its first appearance on Twitter.

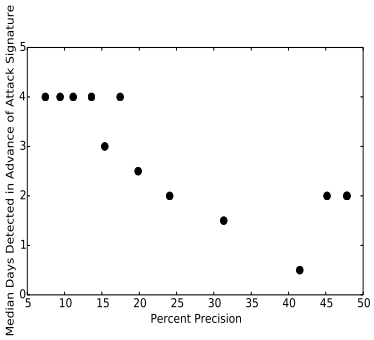
6 Attacks against the exploit detectors

The public nature of Twitter data necessitates considering classification problems not only in an ideal environment, but also in an environment where adversaries may seek to poison the classifiers. In causative adversarial machine learning (AML) attacks, the adversaries make efforts to have a direct influence by corrupting and altering the training data [28, 29, 31].

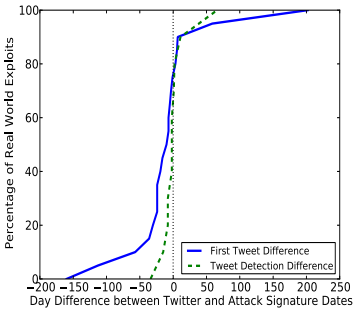
With Twitter data, learning the statistics of the training data is as simple as collecting tweets with either the REST or Streaming APIs. Features that are likely to be used in classification can then be extracted and evaluated using criteria such as correlation, entropy, or mutual information, when ground truth data is publicly available. In this regard, the most conservative assumption for security is that an adversary has complete knowledge of a Twitter-based classifier's training data as well as knowledge of the feature set.

When we assume an adversary works to create both false negatives and false positives (an availability AML security violation), practical implementation of a basic causative AML attack on Twitter data is relatively straightforward. Because of the popularity of spam on Twitter, websites such as `buyaccs.com` cheaply sell large volumes of fraudulent Twitter accounts. For example, on February 16, 2015 on buyaccs.com, the baseline price for 1000 AOL email-based Twitter accounts was $17 with approximately 15,000 accounts available for purchase. This makes it relatively cheap (less than $300 as a base cost) to conduct an attack in which a large number of users tweet fraudulent messages containing CVEs and keywords, which are likely to be used in a Twitter-based classifier as features. Such an attacker has two main limitations. The first limitation is that, while the attacker can add an extremely large number of tweets to the Twitter stream via a large number of different accounts, the attacker has no straightforward mechanism for removing legitimate, potentially informative tweets from the dataset. The second limitation is that additional costs must be incurred if an attacker's fraudulent accounts are to avoid identification. Cheap Twitter accounts purchased in bulk have low friend counts and low follower counts. A user profile-based preprocessing stage of analysis could easily eliminate such accounts from the dataset if an adversary attempts to attack a Twitter classification scheme in such a rudimentary manner. Therefore, to help make fraudulent accounts seem more legitimate and less readily detectable, an adversary must also establish realistic user statistics for these accounts.

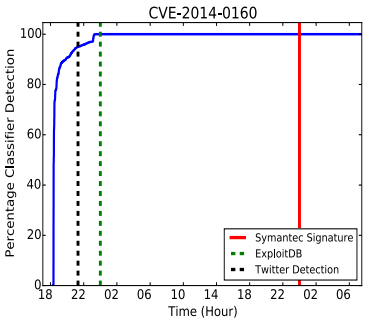
Here, we analyze the robustness of our Twitter-based classifiers when facing three distinct causative attack strategies. The first attack strategy is to launch a causative attack without any knowledge of the training data or ground truth. This *blabbering adversary* essentially amounts to injecting noise into the system. The second attack strategy corresponds to the *word-copycat adversary*, who does not create a sophisticated network between the fraudulent accounts and only manipulates word features and the total tweet count for each CVE. This attacker sends malicious tweets, so that the word statistics for tweets about non-exploited and exploited CVEs appear identical at a user-naive level of abstraction. The third, most powerful adversary we consider is the *full-copycat adversary*. This adversary manipulates the user statistics (friend, follower, and status counts) of a large number of fraudulent accounts as well as the text content of these CVE-related tweets to launch a more sophisticated Sybil attack. The only user statis-



(a) Tradeoff between classification speed and precision.



(b) CDFs for day differences between first tweet for a CVE, first classification, and attack signature creation date.



(c) Heartbleed detection timeline between April 7th and April 10th, 2014.

Figure 8: Early detection of real-world vulnerability exploits.

tics which we assume this full copycat adversary cannot arbitrarily manipulate are account verification status and account creation date, since modifying these features would require account hijacking. The goal of this adversary is for non-exploited and exploited CVEs to appear as statistically identical as possible on Twitter at a user-anonymized level of abstraction.

For all strategies, we assume that the attacker has purchased a large number of fraudulent accounts. Fewer than 1,000 out of the more than 32,000 Twitter users in our CVE tweet dataset send more than 20 CVE-related tweets in a year, and only 75 accounts send 200 or more CVE-related tweets. Therefore, if an attacker wishes to avoid tweet volume-based blacklisting, then each account cannot send a high number of CVE-related tweets. Consequently, if the attacker sets a volume threshold of 20-50 CVE tweets per account, then 15,000 purchased accounts would enable the attacker to send 300,000-750,000 adversarial tweets.

The blabbering adversary, even when sending 1 million fraudulent tweets, is not able to force the precision of our exploit detector below 50%. This suggests that Twitter-based classifiers can be relatively robust to this type of random noise-based attack (black circles in Fig. 9). When dealing with the word-copycat adversary (green squares in Fig. 9), performance asymptotically degrades to 30% precision. The full-copycat adversary can cause the precision to drop to approximately 20% by sending over 300,000 tweets from fraudulent accounts. The full-copycat adversary represents a practical upper bound for the precision loss that a realistic attacker can inflict on our system. Here, performance remains above baseline levels even for our strongest Sybil attacker due to our use of non-Twitter features to increase classifier robustness. Nevertheless, in order to recover performance, implementing a Twitter-based vulnerability classifier in a realistic setting is likely to require curation of whitelists and blacklists for informative and adversarial users. As shown in figure 6a, restricting the classifier to only consider the top 20% of users with the most relevant tweets about real-world exploits causes no performance degradation and fortifies the classifier against low tier adversarial threats.

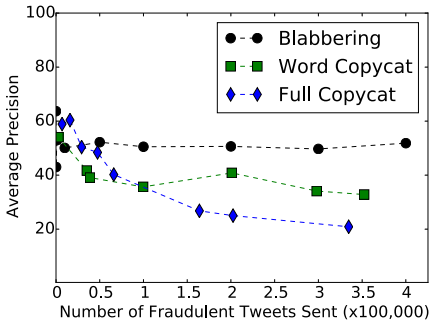


Figure 9: Average linear SVM precision when training and testing data are poisoned by the three types of adversaries from our threat model.

7 Related work

Previous work by Allodi et al. has highlighted multiple deficiencies in CVSS version 2 as a metric for predicting whether or not a vulnerability will be exploited in the wild [24], specifically because predicting the small fraction of vulnerabilities exploited in the wild is not one of the design goals of CVSS. By analyzing vulnerabilities in exploit kits, work by Allodi et al. has also established

that Symantec threat signatures are an effective source of information for determining which vulnerabilities are exploited in the real world, even if the coverage is not complete for all systems [23].

The closest work to our own is Bozorgi et al., who applied linear support vector machines to vulnerability and exploit metadata in order predict the development of proof-of-concept exploits [36]. However, the existence of a POC exploit does not necessarily mean that an exploit will be leveraged for attacks in the wild, and real world attacks occur in only of a small fraction of the cases for which a vulnerability has a proof of concept exploit. [25, 47]. In contrast, we aim to detect exploits that are active in the real world. Hence, our analysis expands on this prior work [36] by focusing classifier training on real world exploit data rather than POC exploit data and by targeting social media data as a key source of features for distinguishing real world exploits from vulnerabilities that are not exploited in the real world.

In prior analysis of Twitter data, success has been found in a wide variety of applications including earthquake detection [53], epidemiology [26, 37], and the stock market [34, 58]. In the security domain, much attention has been focused on detecting Twitter spam accounts [57] and detecting malicious uses of Twitter aimed at gaining political influence [30, 52, 55]. The goals of these works is distinct from our task of predicting whether or not vulnerabilities are exploited in the wild. Nevertheless, a practical implementation of our vulnerability classification methodology would require the detection of fraudulent tweets and spam accounts to prevent poisoning attacks.

8 Discussion

Security in Twitter analytics. Twitter data is publicly available, and new users are free to join and start sending messages. In consequence, we cannot obfuscate or hide the features we use in our machine learning system. Even if we had not disclosed the features we found most useful for our problem, an adversary can collect Twitter data, as well as the data sets we use for ground truth (which are also public), and determine the most information rich features within the training data in the the same way we do. Our exploit detector is an example of a *security system without secrets*, where the integrity of the system does not depend on the secrecy of its design or of the features it uses for learning. Instead, the security properties of our system derive from the fact that the adversary can inject new messages in the Twitter stream, but cannot remove any messages sent by the other users. Our threat model and our experimental results provide practical bounds for the damage the adversary can inflict on such a system. This damage can be reduced further by incorporating techniques for identifying adversarial Twitter accounts, for example by assigning a reputation score to each account [44, 48, 51].

Applications of early exploit detection. Our results suggest that, the information contained in security-related tweets is an important source for timely security-related information. Twitter-based classifiers can be employed to guide the prioritization of response actions after vulnerability disclosures, especially for organizations with strict policies for testing patches prior to enterprise-wide deployment, which makes patching a resource-intensive effort. Another potential application is modeling the risk associated with vulnerabilities, for example by combining the likelihood of real-world exploitation, produced by our system, with additional metrics for vulnerability assessment, such as the CVSS severity scores or the odds that the vulnerable software is exploitable given its deployment context (e.g. whether it is attached to a publicly-accessible network). Such models are key for the emerging area of cyber-insurance [33], and they would benefit from an evidence-based approach for estimating the likelihood of real-world exploitation.

Implications for information sharing efforts. Our results highlight the current challenges for the early detection of exploits, in particular the fact that the existing sources of information for exploits active in the wild do not cover all the platforms that are targeted by attackers. The discussions on security-related mailing lists, such as Bugtraq [10], Full Disclosure [4] and oss-security [8], focus on disclosing vulnerabilities and publishing exploits, rather than on reporting attacks in the wild. This makes it difficult for security researchers to assemble a high-quality ground truth for training supervised machine learning algorithms. At the same time, we illustrate the potential of this approach. In particular,our whitelist identifies 4,335 users who post information-rich messages about exploits. We also show that the classification performance can be improved significantly by utilizing a ground truth with better coverage. We therefore encourage the victims of attacks to share relevant technical information, perhaps through recent information-sharing platforms such as Facebook's ThreatExchange [42] or the Defense Industrial Base voluntary information sharing program [1].

9 Conclusions

We conduct a quantitative and qualitative exploration of information available on Twitter that provides early warnings for the existence of real-world exploits. Among the products for which we have reliable ground truth, we identify more vulnerabilities that are *exploited in*

the real-world than vulnerabilities for which *proof-of-concept exploits* are available publicly. We also identify a group of 4,335 users who post information-rich messages about real-world exploits. We review several unique challenges for the exploit detection problem, including the skewed nature of vulnerability datasets, the frequent scarcity of data available at initial disclosure times and the low coverage of real world exploits in the ground truth data sets that are publicly available. We characterize the threat of information leaks from the coordinated disclosure process, and we identify features that are useful for detecting exploits.

Based on these insights, we design and evaluate a detector for real-world exploits utilizing features extracted from Twitter data (e.g., specific words, number of retweets and replies, information about the users posting these messages). Our system has fewer false positives than a CVSS-based detector, boosting the detection precision by *one order of magnitude*, and can detect exploits a median of *2 days ahead of existing data sets*. We also introduce a threat model with three types of adversaries seeking to poison our exploit detector, and, through simulation, we present practical bounds for the damage they can inflict on a Twitter-based exploit detector.

Acknowledgments

We thank Tanvir Arafin for assistance with our early data collection efforts. We also thank the anonymous reviewers, Ben Ransford (our shepherd), Jimmy Lin, Jonathan Katz and Michelle Mazurek for feedback on this research. This research was supported by the Maryland Procurement Office, under contract H98230-14-C-0127 and by the National Science Foundation, which provided cluster resources under grant CNS-1405688 and graduate student support with GRF 2014161339.

References

[1] Cyber security / information assurance (CS/IA) program. http://dibnet.dod.mil/.

[2] Drupalgeddon vulnerability - cve-2014-3704. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-3704.

[3] Exploits database by offensive security. https:http://exploit-db.com/.

[4] Full disclosure mailing list. http://seclists.org/fulldisclosure/.

[5] Microsoft active protections program. https://technet.microsoft.com/en-us/security/dn467918.

[6] Microsoft's latest plea for cvd is as much propaganda as sincere. http://blog.osvdb.org/2015/01/12/microsofts-latest-plea-for-vcd-is-as-much-propaganda-as-sincere/.

[7] National vulnerability database (NVD). https://nvd.nist.gov/.

[8] Open source security. http://oss-security.openwall.org/wiki/mailing-lists/oss-security.

[9] Open sourced vulnerability database (OSVDB). http://www.osvdb.org/.

[10] Securityfocus bugtraq. http://www.securityfocus.com/archive/1.

[11] Shellshock vulnerability - cve-2014-6271. https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6271.

[12] Symantec a-z listing of threats & risks. http://www.symantec.com/security_response/landing/azlisting.jsp.

[13] Symantec attack signatures. http://www.symantec.com/security_response/attacksignatures/.

[14] Technet blogs - a call for better coordinated vulnerability disclosure. http://blogs.technet.com/b/msrc/archive/2015/01/11/a-call-for-better-coordinated-vulnerability-disclosure.aspx/.

[15] The twitter streaming api. https://dev.twitter.com/streaming/overview/.

[16] Vendors sure like to wave the "coordination" flag. http://blog.osvdb.org/2015/02/02/vendors-sure-like-to-wave-the-coordination-flag-revisiting-the-perfect-storm/.

[17] Windows elevation of privilege in user profile service - google security research. https://code.google.com/p/google-security-research/issues/detail?id=123.

[18] Guidelines for security vulnerability reporting and response. Tech. rep., Organization for Internet Safety, September 2004. http://www.symantec.com/security/OIS_Guidelines%20for%20responsible%20disclosure.pdf.

[19] Adding priority ratings to adobe security bulletins. http://blogs.adobe.com/security/2012/02/when-do-i-need-to-apply-this-update-adding-priority-ratings-to-adobe-security-bulletins-2.html, 2012.

[20] ACHREKAR, H., GANDHE, A., LAZARUS, R., YU, S.-H., AND LIU, B. Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on* (2011), IEEE, pp. 702–707.

[21] ALBERTS, B., AND RESEARCHER, S. A Bounds Check on the Microsoft Exploitability Index The Value of an Exploitability Index Exploitability. 1–7.

[22] ALLODI, L., AND MASSACCI, F. A Preliminary Analysis of Vulnerability Scores for Attacks in Wild. In *CCS BADGERS Workshop* (Raleigh, NC, Oct. 2012).

[23] ALLODI, L., AND MASSACCI, F. A preliminary analysis of vulnerability scores for attacks in wild. *Proceedings of the 2012 ACM Workshop on Building analysis datasets and gathering experience returns for security - BADGERS '12* (2012), 17.

[24] ALLODI, L., AND MASSACCI, F. Comparing vulnerability severity and exploits using case-control studies. *(Rank B)ACM Transactions on Embedded Computing Systems 9*, 4 (2013).

[25] ALLODI, L., SHIM, W., AND MASSACCI, F. Quantitative Assessment of Risk Reduction with Cybercrime Black Market Monitoring. *2013 IEEE Security and Privacy Workshops* (2013), 165–172.

[26] ARAMAKI, E., MASKAWA, S., AND MORITA, M. Twitter Catches The Flu : Detecting Influenza Epidemics using Twitter The University of Tokyo. In *Proceedings of the 2011 Conference on Emperical Methods in Natural Language Processing* (2011), pp. 1568–1576.

[27] ASUR, S., AND HUBERMAN, B. A. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (2010), vol. 1, IEEE, pp. 492–499.

[28] BARRENO, M., BARTLETT, P. L., CHI, F. J., JOSEPH, A. D., NELSON, B., RUBINSTEIN, B. I., SAINI, U., AND TYGAR, J. Open problems in the security of learning. In *Proceedings of the 1st ...* (2008), p. 19.

[29] BARRENO, M., NELSON, B., JOSEPH, A. D., AND TYGAR, J. D. The security of machine learning. *Machine Learning 81* (2010), 121–148.

[30] BENEVENUTO, F., MAGNO, G., RODRIGUES, T., AND ALMEIDA, V. Detecting spammers on twitter. *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS) 6* (2010), 12.

[31] BIGGIO, B., NELSON, B., AND LASKOV, P. Support Vector Machines Under Adversarial Label Noise. *ACML 20* (2011), 97–112.

[32] BILGE, L., AND DUMITRAȘ, T. Before we knew it: An empirical study of zero-day attacks in the real world. In *ACM Conference on Computer and Communications Security* (2012), T. Yu, G. Danezis, and V. D. Gligor, Eds., ACM, pp. 833–844.

[33] BÖHME, R., AND SCHWARTZ, G. Modeling Cyber-Insurance: Towards a Unifying Framework. In *WEIS* (2010).

[34] BOLLEN, J., MAO, H., AND ZENG, X. Twitter mood predicts the stock market. *Journal of Computational Science 2* (2011), 1–8.

[35] BOSER, B. E., GUYON, I. M., AND VAPNIK, V. N. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* (1992), pp. 144–152.

[36] BOZORGI, M., SAUL, L. K., SAVAGE, S., AND VOELKER, G. M. Beyond Heuristics : Learning to Classify Vulnerabilities and Predict Exploits. In *KDD '10 Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), p. 9.

[37] BRONIATOWSKI, D. A., PAUL, M. J., AND DREDZE, M. National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PLoS ONE 8* (2013).

[38] CHANG, C.-C., AND LIN, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology 2* (2011), 27:1—-27:27.

[39] CORTES, C., CORTES, C., VAPNIK, V., AND VAPNIK, V. Support Vector Networks. *Machine Learning 20* (1995), 273˜–˜297.

[40] DUMITRAȘ, T., AND SHOU, D. Toward a Standard Benchmark for Computer Security Research: The {Worldwide Intelligence Network Environment (WINE)}. In *EuroSys BADGERS Workshop* (Salzburg, Austria, Apr. 2011).

[41] DURUMERIC, Z., KASTEN, J., ADRIAN, D., HALDERMAN, J. A., BAILEY, M., LI, F., WEAVER, N., AMANN, J., BEEKMAN, J., PAYER, M., AND PAXSON, V. The matter of Heartbleed. In *Proceedings of the Internet Measurement Conference* (Vancouver, Canada, Nov. 2014).

[42] FACEBOOK. ThreatExchange. `https://threatexchange.fb.com/`.

[43] GUYON, I., GUYON, I., BOSER, B., BOSER, B., VAPNIK, V., AND VAPNIK, V. Automatic Capacity Tuning of Very Large VC-Dimension Classifiers. In *Advances in Neural Information Processing Systems* (1993), vol. 5, pp. 147–155.

[44] HORNG, D., CHAU, P., NACHENBERG, C., WILHELM, J., WRIGHT, A., AND FALOUTSOS, C. Polonium : Tera-Scale Graph Mining and Inference for Malware Detection. *Siam International Conference on Data Mining (Sdm)* (2011), 131–142.

[45] LAZER, D. M., KENNEDY, R., KING, G., AND VESPIGNANI, A. The parable of Google Flu: traps in big data analysis.

[46] MITRE. Common vulnerabilities and exposures (CVE). urlhttp://cve.mitre.org/.

[47] NAYAK, K., MARINO, D., EFSTATHOPOULOS, P., AND DUMITRAȘ, T. Some Vulnerabilities Are Different Than Others: Studying Vulnerabilities and Attack Surfaces in the Wild. In *Proceedings of the 17th International Symposium on Research in Attacks, Intrusions and Defenses* (Gothenburg, Sweeden, Sept. 2014).

[48] PANDIT, S., CHAU, D., WANG, S., AND FALOUTSOS, C. Netprobe: a fast and scalable system for fraud detection in online auction networks. *Proceedings of the 16th ... 42* (2007), 210, 201.

[49] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[50] QUINN, S., SCARFONE, K., BARRETT, M., AND JOHNSON, C. Guide to Adopting and Using the Security Content Automation Protocol {(SCAP)} Version 1.0. NIST Special Publication 800-117, July 2010.

[51] RAJAB, M. A., BALLARD, L., LUTZ, N., MAVROMMATIS, P., AND PROVOS, N. CAMP: Content-agnostic malware protection. In *Network and Distributed System Security (NDSS) Symposium* (San Diego, CA, Feb 2013).

[52] RATKIEWICZ, J., CONOVER, M. D., MEISS, M., GONC, B., FLAMMINI, A., AND MENCZER, F. Detecting and Tracking Political Abuse in Social Media. *Artificial Intelligence* (2011), 297–304.

[53] SAKAKI, T., OKAZAKI, M., AND MATSUO, Y. Earthquake shakes Twitter users: real-time event detection by social sensors. In *WWW '10: Proceedings of the 19th international conference on World wide web* (2010), p. 851.

[54] SCARFONE, K., AND MELL, P. An analysis of CVSS version 2 vulnerability scoring. In *2009 3rd International Symposium on Empirical Software Engineering and Measurement, ESEM 2009* (2009), pp. 516–525.

[55] THOMAS, K., GRIER, C., AND PAXSON, V. Adapting Social Spam Infrastructure for Political Censorship. In *LEET* (2011).

[56] THOMAS, K., LI, F., GRIER, C., AND PAXSON, V. Consequences of Connectivity: Characterizing Account Hijacking on Twitter. ACM Press, pp. 489–500.

[57] WANG, A. H. Don't follow me: Spam detection in Twitter. *2010 International Conference on Security and Cryptography (SECRYPT)* (2010), 1–10.

[58] WOLFRAM, M. S. A. Modelling the Stock Market using Twitter. *iccsinformaticsedacuk Master of* (2010), 74.