Microsoft

Microsoft Digital Defense Report 2024

The foundations and new frontiers of cybersecurity

A Microsoft Threat Intelligence report

MICROSOFT EXHIBIT 1079
Microsoft v. Qomplx
IPR2026-00184

In this report

Overview

About this report 02
Introduction by Tom Burt 03
Our unique vantage point 05
Cybersecurity at Microsoft: the CISO's perspective 07

1 The evolving cyber threat landscape

Key developments 09
Introduction 10
Threat actors and motivations 11
Nation-state threats 12
- Nation-state threat activity by the numbers 12
- Blurring lines between nation-state threat actors and cybercriminals 17
- The many faces of hybrid war 18
- Deterring the most advanced threats 22
- Election interference 24

Ransomware 27
- Landscape and trends 27
- How cybercriminals are tampering with security products 28
- Octo Tempest: a case study and a cautionary tale 29
- Disrupting ransomware threat actors 30

Fraud 31
- Landscape and trends 31
- Phishing 34
- Impersonation 36

Identity and social engineering 39
- Insights on identity attacks and trends 39
- Identity attacks in perspective 41
- Security to the max: the optimal mindset for security professionals 42
- Social engineering "next generation" 44
- Stormy skies: the rise of cloud identity compromise 47

DDoS attacks 50
- DDoS: Stealthier threats emerge 50
- Attack landscape 50
- A new threat: Application loop attacks 50

2 Centering our organizations on security

Key developments 54
Introduction 55
Secure Future Initiative 56
Strategic approaches to cybersecurity: "Managing your own house" 57
- Data security 57
- Hierarchy of cybersecurity needs 60
- Threat-informed defense 61
- Optimizing governance and accountability 63
- Security incident decisions: Dispatches from the field 64
- Resilience maturity 66

Supporting the ecosystem 67
- The passkey journey: a story of collaboration across the industry 67
- Critical environments 69

Collective action 77
- The digital transformation of defense and a call for partnership 77
- How Microsoft helps support democratic elections 79

3 Early insights: AI's impact on cybersecurity

Key developments 83
Introduction 84
- Understanding how generative AI systems work 85
- Two key insights 86

Emerging threat landscape 87
- The generative AI threat landscape 87
- Sophisticated AI-enabled human targeting 89
- Emerging techniques in AI-enabled attacks 90
- Nation-state threat actors using AI for influence operations 91

AI for defense 94
- Harnessing AI to detect cyberattacks 95
- AI's early impact on the security operations center 96
- Seven areas of efficiencies in Microsoft security operations 97
- Using generative AI to understand cyberattacks and create tailored mitigations 99

How governments and industries are advancing global AI security 101
- Government approaches to AI security 101
- Collaborative policy initiatives for AI security 104
- International standards for AI security 105
- Staying a step ahead of threat actors in the age of AI 106

Appendix

References 108
Contributing teams 110

Cyber Point of View stories

Japan 16
Australia 23
Israel 30
Canada 49
India 52
Sweden 59
Latin America 65
France 68
Africa 73
United Kingdom 81
Albania 103

About this report

Report scope

The data, insights, and events in this report represent July 2023 through June 2024 (Microsoft fiscal year 2024), unless otherwise noted.

Please note that due to rounding, the percentages in some charts may not total 100%.

Relevant discussion from the 2023 edition of the Microsoft Digital Defense Report is referenced in this report. You can access the 2023 report in the archive section at aka.ms/MDDR.

Report viewing and navigating

There are links in the headers and table of contents for easy navigation throughout the report.

For easier viewing and navigating through the report on certain browsers, we suggest using Adobe Reader, which is available for free on the Adobe website.

Our commitment to preserving privacy

Any and all data included in this report is presented in alignment to our privacy principles. Microsoft is committed to its focus on preserving customers' control over their data and their ability to make informed choices that protect their privacy.

We advocate for strong global privacy and data protection laws requiring companies, including ours, to only collect and use personal data in responsible, accountable ways.

Threat actor terminology used in this report

- **Nation-state threat attacks/operations:** Malicious cyberattacks that originate from a particular country and are an attempt to further that country's interests. These attacks are often fueled by geopolitical competition and a desire to gain an advantage over other nations. Common objectives include stealing intellectual property for economic benefit or supporting traditional espionage.
- **Cybercriminal activity:** Cybercriminals are typically motivated by financial gain. They may use similar tools and tactics as nation-state threat actors, such as bespoke malware, password spray infrastructure, and phishing or social engineering campaigns. However, their primary goal is to profit from their activities, rather than to further a nation's geopolitical objectives.
- **Cyber operations:** An overarching term referring to all computer network operations, from computer network defense to computer network attacks, and to computer network exploitation.
- **Influence operations (IO):** The coordinated, integrated, and synchronized application of national diplomatic, informational, military, economic, and other capabilities in peacetime, crisis, conflict, and post conflict to foster attitudes, behaviors, or decisions by foreign target audiences that further nation-state interests and objectives.
- **Cyber-enabled influence operations:** Operations which combine offensive computer network operations with messaging and amplification in a coordinated and manipulative fashion to shift perceptions, behaviors, or decisions by target audiences to further a group or a nation's interests and objectives.

Key information

Throughout this document look out for features offering insights and detail from Microsoft experts.

Look out for highlighted text like this and the Actionable Insights sections:

Actionable Insights

Our commitment to developing technology responsibly

As we design, build, and release AI products, six values—transparency, accountability, fairness, inclusiveness, reliability and safety, and privacy and security—remain our foundation and guide our work every day.

Links

Microsoft Privacy Statement

Microsoft EU Data Boundary Overview | Microsoft Trust Center

Empowering responsible AI practices | Microsoft AI

Responsible AI Transparency Report | May 2024

Complex, challenging, and increasingly dangerous

The new cyber threat landscape: an introduction by Tom Burt

"We all can, and must, do better, hardening our digital domains to protect our networks, data, and people at all levels."

In the last year, the cyber threat landscape continued to become more dangerous and complex.

The malign actors of the world are becoming better resourced and better prepared, with increasingly sophisticated tactics, techniques, and tools that challenge even the world's best cybersecurity defenders.

Because these actors conduct both targeted and opportunistic attacks, the threat they present is universal, meaning organizations, users, and devices are at risk anywhere, anytime. Even Microsoft has been the victim of well-orchestrated attacks by determined and well-resourced adversaries, and our customers face more than 600 million cybercriminal and nation-state attacks every day, ranging from ransomware to phishing to identity attacks.

These cyberattacks are continuing at a breathtaking scale, and as they increasingly put human health at risk, the stakes for stopping them couldn't be higher. In the US alone this fiscal year, 389 healthcare institutions were successfully hit by ransomware, resulting in network closures, systems offline, critical medical operations delayed, and appointments rescheduled. Worse, the increased risk of cyberattacks is no longer limited to civilian cybercriminals. Nation-states are becoming more aggressive in the cyber domain, with ever-growing levels of technical sophistication that reflect increased investment in resources and training. These state-sponsored hackers are not just stealing data, but launching ransomware, prepositioning backdoors for future destruction, sabotaging operations, and conducting influence campaigns.

We have to find a way to stem the tide of this malicious cyber activity. We all can, and must, do better, hardening our digital domains to protect our networks, data, and people at all levels. This challenge will not be accomplished solely by executing a well-known checklist of cyber hygiene measures but through a focus on and commitment to the foundations of cyber defense from the individual user level to the executive level.

However, improved defense will not be enough. The sheer volume of attacks must be reduced through effective deterrence, and while the industry must do more to deny the efforts of attackers via better cybersecurity, this needs to be paired with government action to impose consequences that further discourage the most harmful cyberattacks.

While in recent years a great deal of attention has been given to the development of international norms of conduct in cyberspace, those norms so far lack meaningful consequence for their violation, and nation-state attacks have been undeterred, increasing in volume and aggression. Cybercriminals similarly continue to attack with impunity, knowing that law enforcement is hampered by the challenges of investigation and prosecution of cross-border crime, and often operating from within apparent safe havens where government authorities turn a blind eye to the malicious activity.

While the immediate outlook is pessimistic, there are changes on the near horizon that provide cause for optimism. In this year's Microsoft Digital Defense Report, we dive deeper into the subject of AI in cybersecurity. We explore the associated emerging threats and defense strategies, as well as examine the responses of governments around the world to this rapidly evolving technology. And although we must anticipate the use of AI by attackers, advances in AI-powered cybersecurity should give defenders an asymmetric advantage in the near future.

This year we will also share how Microsoft is responding to the significant attacks on our corporate infrastructure. This includes details of our Secure Future Initiative and how we are orchestrating a company-wide initiative to make security our top corporate priority. We hope that these learnings will help others think through their own security posture and approach to cyber defense.

Microsoft is proud to deliver the Microsoft Digital Defense Report, now in its fifth edition, as part of our commitment to helping the world understand and mitigate cyber threats. We believe transparency and information-sharing are essential to the protection of the global cyber ecosystem. Communicating the insights that we derive from our unique vantage point is one of the many ways we work to make the cyber world a safer place.

As our CEO, Satya Nadella, has said: "This is a consequential time." We stand on the frontier of an AI-empowered world. It is up to us, however, to leverage AI most effectively. In the tug-of-war between attackers and defenders in which the attackers currently have an advantage, it will take conscientiousness and commitment by both the public and private sectors to ensure the defenders win.

Tom Burt
Corporate Vice President,
Customer Security and Trust

Our unique vantage point

The depth and breadth of Microsoft's presence in the digital ecosystem offers a unique perspective that we share in this report.

Our expansive, global vantage point gives us insight into key trends in cybersecurity that affect everyone from individuals to nations.

We process more than 78 trillion security signals per day, from billions of Windows endpoints, the cloud, and a broad spectrum of products and services. From these signals we gain visibility into attack activity, a unique understanding of emerging attack techniques, and deeper insights about the overall threat landscape.

This spectrum of security signals is further enhanced by the diversity of our customers and partners, including governments, enterprises large and small, consumers, and gamers.

Microsoft's commitment to supporting the cloud across infrastructure, platform, application, and multi-cloud scenarios complements the diversity of a large ecosystem of partners and suppliers which geometrically expands the richness of the data we use to understand the threat landscape.

Yet our understanding of the threat landscape is more than just data. It is informed by the expertise of our employees:

- Threat intelligence and geopolitical experts, tracking cybercriminal and nation-state threat actors.
- Security researchers, software architects, and engineers, responding to new threats and adding new security features for protection.
- Analysts, internal auditors, and risk specialists, maintaining operational compliance within a complex system of cybersecurity and privacy regulations.
- Incident responders, who "run to the fire" in support of customers.
- Security advisors, working with customers across the spectrum of cybersecurity.
- Investigators, analysts, and legal teams who work globally to disrupt borderless criminal networks, and align public policy objectives in support of digital international norms on cyberpeace.
- Microsoft executives, who are directly accountable for (and have their compensation tied to) the achievement of these security objectives.

Finally, the impact of AI is notable throughout our vantage point. Security researchers and threat hunters are seeing AI transform the threat landscape. However, Microsoft's recent investment in AI technologies reflects confidence in the benefits these tools can provide, including a perspective that exceeds human processing capacity.

Microsoft is proud of its commitment to cybersecurity and organizational resilience. As we celebrate our 50th year, we have gained valuable insights from past challenges. We are keen to share best practices that include maintaining and enhancing the right security culture, addressing technical debt associated with a longstanding corporate history, and investing in a secure future.

78 trillion
security signals per day inform our insights

34,000
full-time dedicated security engineers

15,000
partners with specialized security expertise

Our presence in the digital ecosystem positions us to observe key trends in cybersecurity. Microsoft's perspectives on cybersecurity are framed through 50 years of experience and insight.

Society | Microsoft Stakeholders | Microsoft Customers

Microsoft's unique vantage point

Microsoft serves billions of customers globally, allowing us to aggregate security data from a broad and diverse spectrum of companies, organizations, and consumers.

An extra 13 trillion security signals per day

2023: 65 trillion, 2024: 78 trillion

from the cloud, endpoints, software tools, and partner ecosystem, to understand and protect against digital threats and criminal cyberactivity.

1,500 unique threat groups tracked

Microsoft Threat Intelligence now tracks more than 1,500 unique threat groups—including more than 600 nation-state threat actor groups, 300 cybercrime groups, 200 influence operations groups, and hundreds of others.

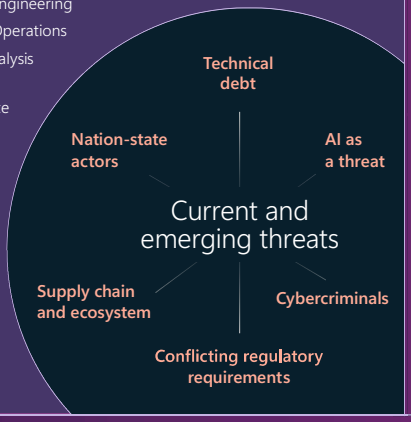
Microsoft's cybersecurity approach

Microsoft security investments

- AI Red Teams
- Defending Democracy
- Detection and Response
- Digital Crimes
- Digital Safety
- Incident Response
- National Security
- Physical Security
- Public Awareness and Education
- Responsible AI
- Security Engineering
- Security Operations
- Threat Analysis
- Threat Intelligence

34,000 dedicated security engineers

focused full-time on the largest cybersecurity engineering project in the history of digital technology.



Cybersecurity at Microsoft: the CISO's perspective

This edition of the Microsoft Digital Defense Report comes to you at a time when the cybersecurity threat landscape has intensified for every sector around the world. Microsoft, like many organizations, has become a primary target, and most notable is the dramatic increase in repeated, sophisticated, and brazen attacks by cybercriminals and nation-state attackers alike.

In January 2024 I took on the role of Microsoft Chief Information Security Officer (CISO). Immediately thereafter, we discovered we were under a massive cyberattack by the threat actor we refer to as Midnight Blizzard. The subsequent days are some I remember vividly. Every available resource across the company was utilized in our defense against this attack—a monumental effort that required speed, focus, and expertise. As I was directing our response, my priority became defending Microsoft and scaling our agility to face future nation-state attacks. A large portion of our third-party ecosystem was involved in this defense as well.

Given ever-changing geopolitical conditions, the world will face many such attacks in the future, and Microsoft must also adjust to face these threats. We have taken major steps over the past year in fortifying assets across the company to better prevent and defend against such threats. The cornerstone of our work to protect Microsoft, our partners, and customers is the Secure Future Initiative[1] (SFI), which dedicates the entire company to putting security above all other considerations.

As Satya Nadella, Microsoft's CEO, said in a company-wide announcement, "Security is a team sport, and accelerating SFI isn't just job number one for our security teams—it's everyone's top priority and our customers' greatest need." Everyone at Microsoft is committed to making our products and services secure by design, secure by default, and operationally secure.

Among the most significant mitigations and actions we have taken is a significantly expanded SFI to improve our defense posture. We made phishing-resistant multifactor authentication (MFA) mandatory across the company, and we increased the robustness of Microsoft's corporate network.

To protect Microsoft, our partners, and customers from future attacks, we dramatically grew our teams dedicated to monitoring of and responding to threats. And we reassigned roughly 34,000 full-time equivalent engineers to security initiatives. This is an important sampling of the many steps we have taken since the beginning of this year—with much more work in progress.

To increase the agility of Microsoft's response to this ever-changing threat environment, I instituted an Office of the CISO and have hired a number of Deputy CISOs. Our Deputy CISOs work with our major product groups and programs to drive greater depth and rigor in cybersecurity governance across the entire company and to direct SFI at the most pressing security risks. The Deputy CISOs take responsibility for risk ownership and accountability, determining the needed security architecture, and providing input to me on each business unit's progress toward our SFI goals. Based on the ongoing SFI work—and with input from the Deputy CISOs—I provide regular updates on existing risk and SFI performance to Microsoft's Senior Leadership Team and Board of Directors.

Every one of us at Microsoft shares a deep responsibility to do our part to keep the world safe and secure. As part of that commitment, we are collaborating closely with security experts, industry groups, and organizations like yours that face these threats every day. Please read on to learn more about the evolving threat landscape and how we are committed to making the world safer for everyone.

Igor Tsyganskiy
Chief Information Security Officer

Chapter 1

The evolving cyber threat landscape

How have trends and tactics changed?

Key developments 9
Introduction 10
Threat actors and motivations 11
Nation-state threats 12
Ransomware 27
Fraud 31
Identity and social engineering 39
Distributed denial of service (DDoS) attacks 50

Key developments
The evolving cyber threat landscape

As with any landscape, things change over time. In the world of cybersecurity, however, the pace of change has been astounding.

Observations over the past year have reaffirmed the convergence of nation-state and cybercriminal threat activity. Nation-state threat actors used cybercrime as a force multiplier, while financially motivated cybercriminals pursued levels of defense evasion and technical complexity once elusive outside of nation-state operations.

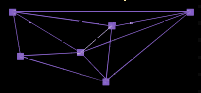
We have also seen rapid shifts in the tactics of hybrid war, wide-ranging attempts to interfere in democratic elections, and a surge in ransomware attacks and cyber-enabled financial fraud across the globe.

These trends underscore the ongoing necessity to enhance and implement robust deterrence and mitigation strategies to counter these threats effectively.

Blurred lines between nation-state threat actor activity and cybercrime

Nation-state threat actors are conducting operations for financial gain and enlisting the aid of cybercriminals and commodity malware to collect intelligence.

Find out more on p17.



The many faces of hybrid war

Threat actors serving Russia and Iran are leaning into cyber and influence operations as tools to advance political and military objectives in wartime.

Find out more on p18.

The need to impose deterrent consequences for cyber aggression

The pace of nation-state sponsored cyberattacks has escalated to the point that there is now effectively constant combat in cyberspace without any meaningful consequences to the attacker.

Find out more on p22.

600 million identity attacks per day

As multifactor authentication blocks most password-based attacks, threat actors are shifting their focus.

Find out more on p39.

Nation-state influence operations converge on elections

By the end of 2024, 2 billion people will have had the opportunity to vote in nationwide elections. Russia, Iran, and China all engaged in election influence efforts globally in 2024.

Find out more on p24.

2.75x increase in human-operated ransomware-linked encounters

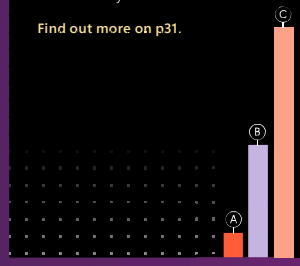
By disabling or tampering with defenses, attackers buy themselves time to install malicious tools, exfiltrate data for espionage or extortion, and potentially launch attacks like ransomware.

Find out more on p27.

Ingenuity and scalability of fraud tactics surging globally

Cyber fraud not only presents a theft risk, but it undermines the security, trust, and reputation of individuals, businesses, and organizations of all sizes and types, in every region and industry.

Find out more on p31.



Introduction: The evolving landscape of cybersecurity

"As we look to the future, the dawning of the age of AI means cybersecurity professionals will encounter both new opportunities and new challenges."

As we reflect on this past year, it is more apparent that the lines that once divided cybercrime, nation-state sponsored attacks, and influence operations have continued to blur. Cybercrime has continued to mature as a robust and elaborate ecosystem, with cybercriminal groups utilizing a full spectrum of tools and techniques, including those learned, borrowed, or stolen from nation-state actors. While these cybercriminals are evolving their tooling and targeting to evade defenders, many of their underlying techniques and behaviors remain unchanged due to their continued effectiveness. Meanwhile, nation-state actors remain committed to pursuing new levels of sophistication. This includes creating unique tooling, upskilling their capabilities, and targeting major technology providers—like Microsoft—and enterprise supply chains.

Defenders can proactively combat threats from both cybercriminal and nation-state actors by addressing them at the technique layer. This means implementing and enforcing policies and tooling, such as enhanced multifactor authentication (MFA) and attack surface reduction rules. At the same time, as the threat landscape evolves, securing identities, hardening endpoints, and protecting the cloud infrastructure has become more important than ever.

As Microsoft continues to take steps to protect ourselves and our customers through our Secure Future Initiative, we encourage all organizations to commit to the foundational security principles of secure by design, secure by default, and secure operations. By collectively working toward these fundamental security concepts, defenders can reduce the attack surface across the broader technology landscape.

At the same time, we have seen influence operations change and increase globally at an unprecedented scale as nation-states seek to sway public perception and sentiment, sow discord, and undermine trust in public institutions. In particular, governments have used geopolitical issues such as the Russia-Ukraine conflict and the Israel-Hamas war to spread divisive and misleading messages. At a time when the world is grappling with an overwhelming influx of information delivered through both formal and informal channels, the issue of combatting misinformation is becoming increasingly vital.

As we look to the future, the dawning of the age of AI means cybersecurity professionals will encounter both new opportunities and new challenges. Cybercriminal groups, nation-state threat actors, and other adversaries are exploring AI technologies to understand whether and how to leverage them in the course of operations. We as defenders must also explore and test these AI technologies, not only to understand how they can be used by adversaries, but how we can use them to strengthen our security, protection, and response.

Amy Hogan-Burney
Vice President and Deputy General Counsel
Customer Security and Trust,
Cybersecurity Policy & Protection Unit

John Lambert
Corporate Vice President, Security Fellow,
Microsoft Threat Intelligence Center

Threat actors and motivations

In this report, we discuss 30 different threat actors to provide examples of activity for a better understanding of attack targets, techniques, and motivations. Microsoft categorizes these actors using a weather-related naming system. For example, "Flood" refers to actors who engage in influence operations. The actors included in this year's report demonstrated significant activity and effectiveness from July 2023 through June 2024. In the chart below, we map some of the motivations tracked over the past five years, to show how these actors often have multiple motivations driving their operations. It's important to note that the threat landscape is vast, and the threat actors and motivations detailed here represent only a small portion of those tracked by Microsoft.

Nation-state actors

Cyber operators acting on behalf of or directed by a nation-state-aligned program, irrespective of whether for espionage, financial gain, or retribution.

Russia

Aqua Blizzard	E
Midnight Blizzard	E
Seashell Blizzard	Dd D Ei E I R
Secret Blizzard	E

China

Flax Typhoon	Dt E
Granite Typhoon	Dt E
Nylon Typhoon	E
Raspberry Typhoon	Dt E

North Korea

Citrine Sleet	C Dt
Jade Sleet	C Dt
Moonstone Sleet	E R
Sapphire Sleet	C Dt

Iran

Cotton Sandstorm	Dd Dt D Ei E I R
Mint Sandstorm	Ei E R

Influence Operations

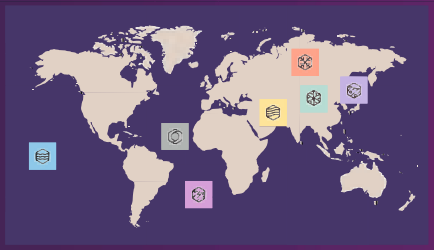
Information campaigns or groups employing communications online or offline in a manipulative fashion to shift perceptions, behaviors, or decisions by target audiences to further a group or a nation's interests and objectives.

Ruza Flood	Ei I
Sefid Flood	Ei E I
Taizi Flood	Ei I
Volga Flood	Ei I

Financially motivated

Cyber campaigns or groups directed by a criminal organization or person with motivations of financial gain and are not associated with high confidence to a known non-nation-state or commercial entity.

Octo Tempest	C Dt R



KEY TO MOTIVATIONS MAPPING

Cryptocurrency theft	C
Cybercrime services	CS
Data destruction	Dd
Data theft for profit	Dt
Disruption	D
Election influence	Ei
Espionage	E
Influence operations	I
Ransomware/Extortion	R

Groups in development

A temporary designation given to unknown, emerging, or developing threat activity. This designation allows Microsoft to track a group as a discrete set of information until we reach high confidence about the origin or identity of the actor behind the operation.

Storm–0501	Dt R
Storm–0539	Dt
Storm–0593	E
Storm–0784	D R
Storm–0842	Dd D Ei I R
Storm–0867	CS
Storm–1101	CS
Storm–1516	Ei I
Storm–1575	D
Storm–1679	I
Storm–2049	E

For more information on threat actor naming, please visit https://aka.ms/threatactors

Nation-state threats

Nation-state threat activity by the numbers

This past year, nation-state affiliated threat actors once again demonstrated that cyber operations—whether for espionage, destruction, or influence—play a persistent supporting role in broader geopolitical conflicts. In the wars in Europe and the Middle East, Russia and Iran centered their threat activity on their main adversaries in those fights, Ukraine, and Israel, respectively. Meanwhile, Beijing's long-term focus on controlling Taiwan drove a high level of targeting of Taiwan-based enterprises from Chinese threat actors, who also penetrated the countries around the South China Sea to collect insights into military exercises and national policy. What follows is a snapshot of the activity by-the-numbers.

The United States is consistently among the countries most impacted by the nation-state cyber threat activity that Microsoft observes, a reflection of the large US representation in our customer base and the role the United States plays in research and development and geopolitical events. Aside from the United States and the United Kingdom—which was the fifth most targeted nation this year—most of the nation-state affiliated cyber threat activity we observed was concentrated in sites of active military conflict or heightened regional tension: Israel, Ukraine, the United Arab Emirates, and Taiwan.

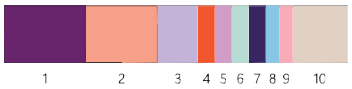
The Education and Research sector became the second most targeted by nation-state threat actors

In 2024, Education and Research became the second most targeted sector by nation-state threat actors.

In addition to offering intelligence such as research and policy discussions, education and research institutions are often used as testing grounds by threat actors before they pursue their actual targets.

For example, QR code phishing, a technique now used widely to compromise user accounts at scale and create an entry point for business email compromise (BEC) attacks discussed later in this chapter, became widely used in targeted attacks against this sector as early as August 2023.

Top 10 targeted sectors worldwide



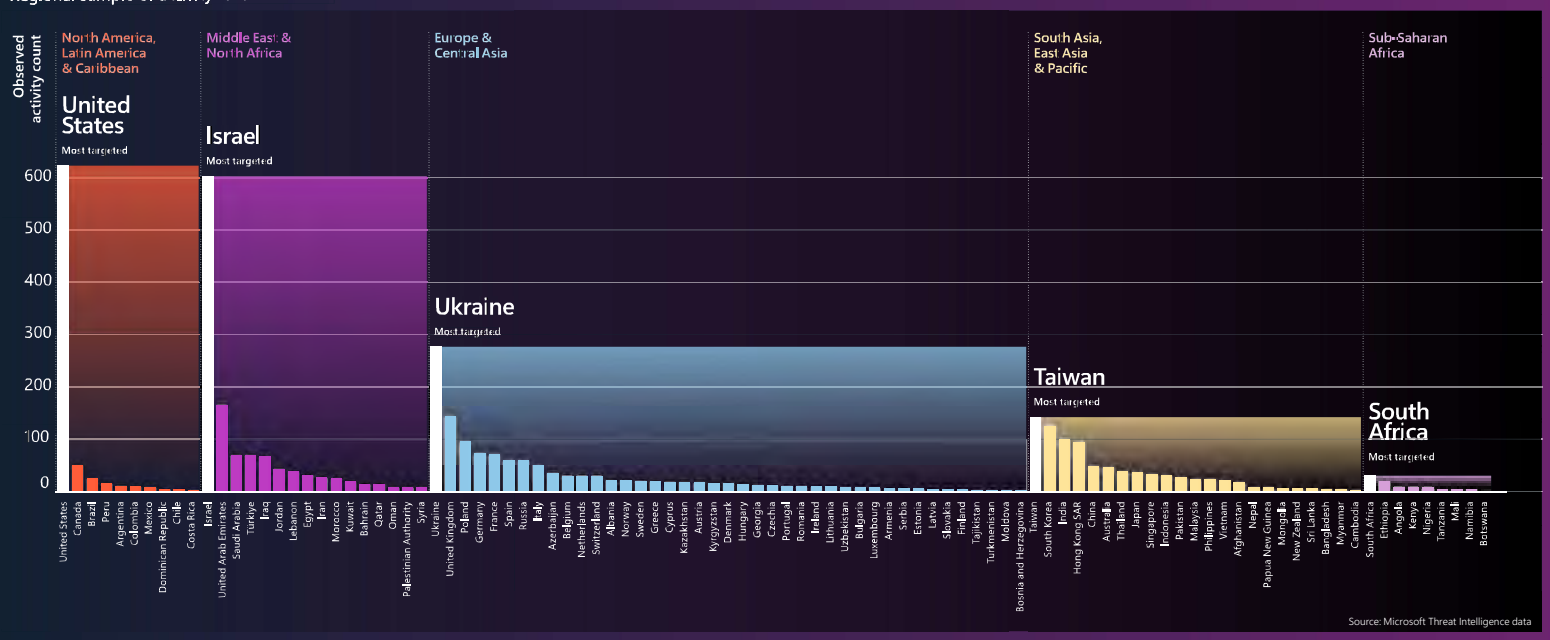
	Sector	Percentage
1	IT	24%
2	Education and Research	21%
3	Government	12%
4	Think tanks and NGOs	5%
5	Transportation	5%
6	Consumer Retail	5%
7	Finance	5%
8	Manufacturing	4%
9	Communications	4%
10	All others	16%

Threat actors from Russia, China, Iran, and North Korea pursued access to IT products and services, in part to conduct supply chain attacks against government and other sensitive organizations.

Source: Microsoft Threat Intelligence, nation-state notification data

Nation-state threat actor targeting

Regional sample of activity levels observed



Russia

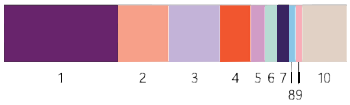
Nation state threat actor activity

Targeting by region

Sector	Percentage
1 Europe & Central Asia	68%
2 North America	20%
3 Middle East & North Africa	5%
4 East Asia & Pacific	3%
5 Latin America & Caribbean	3%
6 South Asia	1%
7 Sub-Saharan Africa	1%

Approximately 75% of targets were in Ukraine or a NATO member state, as Moscow seeks to collect intelligence on the West's policies on the war. Ukraine remains the country most targeted by Russian actors.

Most targeted sectors



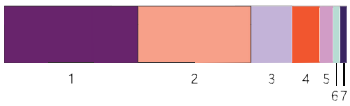
Sector	Percentage
1 Government	33%
2 IT	15%
3 Think tanks and NGOs	15%
4 Education and Research	9%
5 Inter-governmental organization	4%
6 Defense Industry	4%
7 Transportation	3%
8 Energy	2%
9 Media	2%
10 All others	13%

Russian actors focused their targeting against European and North American government agencies and think tanks, likely for intelligence collection related to the war in Ukraine. Actors like Midnight Blizzard also targeted the IT sector, suggesting it was in part planning supply-chain attacks to gain access to these companies' client's networks for follow-on operations.

China

Nation state threat actor activity

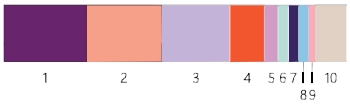
Targeting by region



Sector	Percentage
1 East Asia & Pacific	39%
2 North America	33%
3 Europe & Central Asia	12%
4 Latin America & Caribbean	8%
5 South Asia	4%
6 Middle East & North Africa	2%
7 Sub-Saharan Africa	2%

Chinese threat actors' targeting efforts remain similar to the last few years in terms of geographies targeted and intensity of targeting per location. While numerous threat actors target the United States across a wide variety of sectors, targeting in Taiwan is largely limited to one threat actor, Flax Typhoon.

Most targeted sectors



Sector	Percentage
1 IT	24%
2 Education and Research	22%
3 Government	20%
4 Think tanks and NGOs	10%
5 Manufacturing	4%
6 Defense Industry	3%
7 Communications	3%
8 Finance	3%
9 Transportation	2%
10 All others	9%

Most Chinese threat activity is for intelligence collection purposes and was especially prevalent in ASEAN countries around the South China Sea. Granite Typhoon and Raspberry Typhoon were the most active in the region, while Nylon Typhoon continued to target government and foreign affairs entities globally.

Source: Microsoft Threat Intelligence nation-state notification data

Iran

Nation-state threat actor activity

Targeting by region

Sector	Percentage
1 Middle East & North Africa	53%
2 North America	23%
3 Europe & Central Asia	12%
4 South Asia	6%
5 East Asia & Pacific	3%
6 Latin America & Caribbean	2%
7 Sub-Saharan Africa	1%

Iran placed significant focus on Israel, especially after the outbreak of the Israel-Hamas war. Iranian actors continued to target the US and Gulf countries, including the UAE and Bahrain, in part because of their normalization of ties with Israel and Tehran's perception that they are both enabling Israel's war efforts.

Most targeted sectors

Sector	Percentage
1 Education and Research	19%
2 IT	11%
3 Government	7%
4 Transportation	6%
5 Finance	4%
6 Communications	4%
7 Energy	3%
8 Commercial Facilities	3%
9 Manufacturing	3%
10 All others	42%

Iranian targeting focused on education, IT, and government as part of strategic intelligence collection. Iranian actors often target the IT sector to gain access to downstream customers, including those in government and the defense industrial base (DIB). "Other" includes media and think tanks or NGOs, which Iran often targets to gain insights into dissidents, activists, and persons who can impact policymaking.

North Korea

Nation-state threat actor activity

Targeting by region

Sector	Percentage
1 North America	54%
2 East Asia & Pacific	18%
3 Europe & Central Asia	18%
4 Latin America & Caribbean	3%
5 Middle East & North Africa	3%
6 South Asia	2%
7 Sub-Saharan Africa	2%

The United States remained the most heavily targeted country by North Korean threat actors, but the United Kingdom rose up the ranks this year to second place. The "Other" category comprised 44 other countries targeted by North Korean threat actors.

Most targeted sectors

Sector	Percentage
1 IT	44%
2 Education and Research	21%
3 Manufacturing	6%
4 Consumer Retail	5%
5 Finance	5%
6 Think tanks and NGOs	3%
7 Communications	2%
8 Government	2%
9 Health	2%
10 All others	10%

North Korean threat actors targeted the IT sector the most, particularly to conduct increasingly sophisticated software supply chain attacks. They also continued to heavily target experts in the education sector for intelligence collection. The "Other" category comprised seven other sectors.

Source: Microsoft Threat Intelligence nation-state attribution data

Cyber Point of View: Japan

Japan is surrounded by three nation states backing threat actors who conduct extensive cyberattacks: Russia, China, and North Korea.

In recent years, Japanese entities from large organizations to small companies downstream in the supply chain, have experienced large-scale cyberattacks. Against this backdrop, its government revised its National Security Strategy[1] in December 2022 to identify cybersecurity as a national security matter for the first time. The new National Security Strategy also introduced Active Cyber Defense (ACD), a government initiative to preemptively counter significant cyberattack risks that could raise national security concerns.

Japan's new Defense Whitepaper outlines its cybersecurity measures

The 2024 edition of Japan's Defense Whitepaper[1] outlines comprehensive new measures to enhance the cybersecurity of The Japan Self-Defense Forces (SDF), including the development of the newly established Cyber Command, the migration of the JSDF's IT systems to the cloud, the implementation of advanced security architecture, and improved cybersecurity for Japan's defense industry. The document also stresses the importance of international cooperation with like-minded countries and companies.

Since adopting this new strategy, the government has been aggressive in bolstering its cybersecurity posture to protect the government, businesses, and civil society. Notable initiatives include:

- Directing the JSDF to establish a new Cyber Command with 20,000 personnel by 2027.
- Elevating the Cabinet's cybersecurity center (NISC) into a new government agency with more legal and regulatory authority on cybersecurity.

Additionally, to strengthen the security of IT systems and supply chain management, the government began operating an advanced certification system for cloud (ISMAP)[2] in 2022 for services and other items to be employed in government information systems. ISMAP is expected to expand its scope to the systems of critical infrastructure operators in the future.

Blurring lines between nation-state threat actors and cybercriminals

This year, state-affiliated threat actors increasingly used criminal tools and tactics—and even criminals themselves—to advance their interests, blurring the lines between nation-state backed malign activity and cybercriminal activity.

Microsoft observed nation-state threat actors conduct operations for financial gain, enlist cybercriminals to collect intelligence on the Ukrainian military, and make use of the same infostealers, command and control frameworks, and other tools favored by the cybercriminal community.

$3 billion

in cryptocurrency stolen by North Korean hackers since 2017

North Korean threat actors have long straddled this blurry line, conducting financially motivated operations to secure funding for state coffers and priority initiatives. The UN estimates North Korean hackers have stolen over $3 billion US in cryptocurrency since 2017, with heists totaling between $600 million and $1 billion US in 2023 alone.[5] These stolen funds reportedly finance over half of North Korea's nuclear and missile programs.[6]

Since 2023, Microsoft has identified three major North Korean threat groups—Jade Sleet, Sapphire Sleet, and Citrine Sleet—that have been particularly active in targeting cryptocurrency organizations. Moreover, North Korea may also be getting into the ransomware game. Moonstone Sleet, a new North Korean actor identified in May 2024, developed a custom ransomware variant called FakePenny which it deployed at organizations in aerospace and defense after exfiltrating data from the impacted networks. This behavior suggests the actor had objectives for both intelligence gathering and monetization of its access.

Beyond North Korea, Microsoft observed Iranian nation-state threat actors seeking financial gain from some of their offensive cyber operations. This marks a change from previous behavior, whereby ransomware attacks that were designed to appear financially motivated were actually destructive attacks.[7] For example, a cyber-enabled influence operation run by an Islamic Revolutionary Guard Corps (IRGC) group we track as Cotton Sandstorm (also known as Emennet Pasargad) marketed stolen Israeli dating website data through two of its cyber personas between September 2023 and February 2024. The personas also offered to remove specific individual profiles from their data repository for a fee.

Meanwhile, Russian threat actors have integrated evermore commodity malware in their operations and appear to have outsourced some cyberespionage operations to criminal groups. In June 2024, Storm-2049 (UAC-0184) used Xworm and Remcos RAT--commodity malware associated with criminal activity--to compromise at least 50 Ukrainian military devices. There was no obvious cybercriminal use for this compromise, suggesting the group was operating in support of Russian government objectives.

Between June-July 2023, Microsoft observed Federal Security Service (FSB)-attributed Aqua Blizzard appear to "hand-off" access to 34 compromised Ukrainian devices to the cybercriminal group Storm-0593 (also known as Invisimole). The hand-off occurred when Aqua Blizzard invoked a Powershell script that downloaded software from a Storm-0593-controlled server. Storm-0593 then established command and control infrastructure and deployed Cobalt Strike beacons on most of the devices for follow-on activity. This beacon was configured with the domain dashcloudew.uk, which Microsoft assesses Storm-0593 registered and used in a previous spear-phishing campaign against Ukrainian military machines last year, suggesting a pattern by Storm-0593 of supporting state intelligence collection objectives.

The many faces of hybrid war

The ongoing conflicts in Ukraine and the Middle East illustrate how some countries are using both cyber approaches and influence campaigns to further their goals. These activities extend beyond the geographical boundaries of the conflict zones, demonstrating the globalized nature of hybrid warfare.

How Iran is using cyber-enabled influence operations to degrade Israel

Following the outbreak of the Israel-Hamas war, Iran surged its cyber, influence, and cyber-enabled influence operations against Israel. From October 7, 2023, to July 2024, nearly half of the Iranian operations Microsoft observed targeted Israeli companies.

Iranian groups also expanded their cyber-enabled influence operations beyond Israel, with a focus on undermining international political, military, and economic support for Israel's military operations.

In November 2023, IRGC groups ran cyber-enabled influence operations targeting US water controllers made in Israel and Bahrain in retaliation for Bahrain's normalizing of ties with Israel.[1]

Microsoft Threat Intelligence assesses that an IRGC unit known as Shahid Kaveh Group, which we track as Storm-0784, was responsible for defacing a water controller in Pennsylvania under the guise of a cyber persona called "CyberAvengers," leaving a message that Israeli-made systems are legal targets.

Throughout the conflict, Iranian threat actors have used cyber personas to broadcast and amplify their destructive attacks against Israeli enterprises, trying to project power and aggrandize the impact of their cyber operations. Within two days of Hamas' attack on Israel, Iran stood up several new influence operations. The influence actor Sefid Flood launched the online personas "Tears of War" and "Hamsa1948."

The former impersonated Israeli activists critical of Israeli Prime Minister Benjamin Netanyahu's handling of the hostage situation while the latter tried to convince Arab-Israelis to violently oppose Israeli authorities and protest in support of Gazans. Microsoft Threat Intelligence assesses that Storm-0842, an Iranian Ministry of Intelligence and Security (MOIS) unit, launched another cyber persona, "KarMa," the day after the war broke out,[1] posing as Israelis seeking to remove Netanyahu from office.

Iranian threat actors also began impersonating partners after the war started. Microsoft assesses Cotton Sandstorm used the name and logo of Hamas's military wing, the al-Qassam Brigades, to spread false messaging about the hostages in Gaza and send Israelis threatening messages. Another Telegram channel that we assess was run by the Iranian Ministry of Intelligence and Security (MOIS), which also used the al-Qassam Brigades logo and threatened Israeli military personnel and leaked their personal data. It remains unclear whether Iran acted with Hamas's consent.

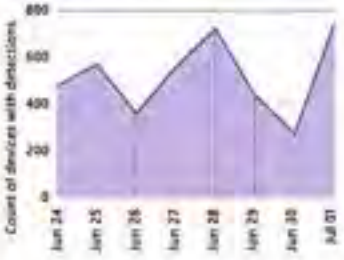
Russia's wide-reaching tactics for spying on Ukraine's military and its allies

Russian threat actors have focused on accessing and stealing intelligence from Ukrainian warfighters and the international partners that supply them weapons. The techniques employed have the potential to cause unintended damage by posing risk to computer networks globally.

Since June 2023, threat actors associated with Russian military intelligence (GRU) and the FSB have used at least two undisciplined approaches to gain access to Ukrainian military and military-adjacent devices:

1. **USB-delivered worms:** Aqua Blizzard—a Russian Federal Security Service (FSB)-affiliated actor that has targeted Ukraine-based entities since 2013—accessed 500-750 Ukraine-based devices daily through the USB-delivery of a Windows ShortCut file and a heavily obfuscated Powershell or VBScript. The scripts establish command and control that facilitates theft of specified file types. Since wormable malware and malicious USBs are hard to contain and can traverse to devices outside the scope of Aqua Blizzard's operations, there is increased risk that USBs and malware will make their way onto networks outside of Ukraine and onto partner military systems.

Daily count of Aqua Blizzard malware detections



Source: Microsoft Threat Intelligence

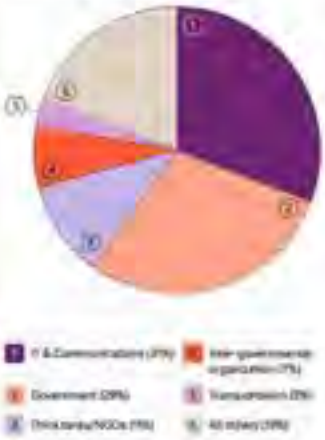
2. **Amadey Bot and torrents:** FSB-affiliated Secret Blizzard and GRU-affiliated Seashell Blizzard gain access to as many devices as possible before pursuing devices of interest. Secret Blizzard has done this by commandeering third-party infections, like the multipurpose Amadey Bots,[10] to download a custom reconnaissance tool that helps operators decide whether to deploy their first-stage backdoor. Seashell Blizzard offers malicious, pirated versions of Microsoft software on torrents, often promoting them on Ukrainian file sharing websites to gain initial footholds in networks.

Midnight Blizzard threatens IT supply chain

Russian threat actors are casting wide nets to gain insights into Western organizations involved in policy, military, and humanitarian support to Ukraine. Midnight Blizzard attempted to gain access to IT firms in part for widespread, indiscriminate access to systems. Historically, this actor exploits the IT software and services supply chains to target downstream customers in government and other policy organizations in North America and Europe.[11]

Microsoft has been transparent about Midnight Blizzard's efforts against our networks, and we were not the only IT sector targets. Midnight Blizzard's history of supply chain compromises and continued pursuit of IT organizations suggests widespread compromise remains a major risk to providers worldwide.

Midnight Blizzard's most targeted sectors



Source: Microsoft Threat Intelligence

Operational technology (OT) systems are at risk in hybrid warfare

Critical infrastructure is a key target of physical strikes and cyberattacks in modern hybrid conflicts. Since late 2023, Microsoft has observed an increase in reports of attacks on internet-exposed, poorly secured OT devices that control real-world critical processes. As discussed in greater detail in previous editions of this report, this is particularly concerning given these systems often have inadequate security practices, including being left unpatched, using default passwords, or even no passwords at all.

Internet-exposed OT equipment in water and wastewater systems (WWS) in the United States were targeted in multiple attacks from October 2023 through June 2024 by different nation-backed actors, including IRGC-affiliated CyberAv3ngers (tracked at Microsoft as Storm-0784) and pro-Russian hacktivists.[9] CyberAv3ngers and the pro-Russia Cyber Army of Russia group, conduct, claim, or amplify attacks likely intended to intimidate targeted nations into capitulating or ceasing support for Israel and Ukraine, respectively.

Links

Onyx Sleet uses array of malware to gather intelligence for North Korea | Jul 2024

Exploitation of Unitronics PLCs used in Water and Wastewater Systems | CISA | Nov 2023

Distribution of internet-exposed Unitronics controllers communicating over PCOM protocol as of June 2024

Programmable logic controllers are human-machine interfaces that automate and control physical processes and can be found in every industrial environments. As these devices are used in industrial environments and critical infrastructure, their current level of exposure leaves many critical infrastructure processes open to attack.

Source: Microsoft Defender Attack Surface Management, Microsoft Threat Intelligence

Chinese threat actors target military and IT entities in the South China Sea

China-based cyber actors Raspberry Typhoon, Flax Typhoon, and Granite Typhoon have intensively targeted entities associated with IT, military, and government interests around the South China Sea.

The activity has particularly targeted countries within the Association of Southeast Asian Nations (ASEAN). Raspberry Typhoon has been extremely active, successfully infiltrating military and executive entities in Indonesia and Malaysian maritime systems in the lead-up to a rare naval exercise involving Indonesia, China, and the United States in June 2023. Similarly, Flax Typhoon focused on entities linked to joint US-Philippines military exercises. Since August 2023, Flax Typhoon has expanded its targets to include IT and government organizations in the Philippines, Hong Kong, India, and the United States.

Since July 2023, Granite Typhoon has compromised telecommunication networks in Indonesia, Malaysia, the Philippines, Cambodia, and Taiwan. This group's activities highlight a sustained pattern of strategic cyber engagements by Chinese state-affiliated actors aimed at gathering intelligence and potentially disrupting military activities in strategically important areas like the South China Sea.

Deterring the most advanced threats

As highlighted throughout this report, cyberattacks are on the rise and the pace of nation-state sponsored attacks has escalated to the point there is now effectively constant combat in cyberspace.

With more than 600 million attacks per day targeting Microsoft customers alone, there must be countervailing pressure to reduce the overall number of attacks online. Deterring this malign activity will require a robust combination of technological and geopolitical solutions. This deterrence can be achieved in two ways – by denial of intrusions or imposing consequences. While companies like Microsoft can help "deny" successful cyberattacks via innovation and further improvements in cybersecurity, enforcing international rules with deterrent consequences must fall on governments.

Microsoft therefore urges governments to consider the following actions to improve adherence to international law and online norms by strengthening digital diplomacy, sharpening public attributions, and imposing meaningful consequences for cyber aggression.

I. **Strengthen international norms and diplomacy.** Deterrence requires clear expectations around acceptable and unacceptable behavior. To that end, Microsoft encourages governments to embrace:

- **New norms:** The United Nations (UN) and other forums should recognize cloud services and the information and communications technology (ICT) supply chain as critical infrastructure that is off-limits to targeting. Moreover, states should be expected to fulfill their due diligence obligations to address malicious activity originating from within their territories.
- **Multistakeholder inclusion:** States should embrace more inclusive diplomatic processes that ensure participation of critical non-governmental stakeholders in discussions on peace and security online, including leading voices from the tech industry and civil society.
- **Bilateral agreements:** In addition to working through multilateral forums, governments should explore potential bilateral agreements as a means to set expectations and curb dangerous cyber operations.

II. **Sharpen government attributions of malicious activity.** Public attribution contributes to deterrence by calling out internationally unacceptable behavior and serving as a necessary precursor for imposing further consequences. The following are ways in which governments might strengthen the impact of public attribution statements:

- **Uniformity:** A single agency should handle public attribution statements with a standard format detailing the incident, responsible parties, impact, evidence, rule violations, consequences, and any preventive measures.
- **Contextualization:** Public attributions should also include any broader insights into the threat actor's activities to support more comprehensive accountability.
- **Collaboration:** Coalition attributions by multiple governments help substantiate and build confidence in findings. Governments should also partner with the tech industry for further validation and with civil society groups to provide further context around impact and harms of cyberattacks.

III. **Impose deterrent consequences.** The escalating volume of nation-state sponsored cyberattacks necessitates more decisive governmental action that stems the growth. Possible response strategies include:

- **Enhanced countermeasures:** Beyond public attribution, states conducting illegal cyber activities should expect firm countermeasures in response, this includes targeted sanctions among other options.
- **Collective countermeasures:** Governments should embrace as lawful collective countermeasures, multiple states imposing countermeasures in response to illegal cyber operations targeting any one of them.
- **Clarify red lines:** In line with the UN Charter's prohibition on "threat or use of force", it should be explicitly stated that state-sponsored cyber intrusions that could be used to damage or interrupt critical civilian services constitute an unlawful threat of force and allow for more significant consequences in response.

A more robust deterrent framework will help to promote stability, protect critical infrastructure, and avoid some of the most harmful cyberattacks. To support this, governments should deepen partnerships across stakeholder groups to identify the essential critical infrastructure. Given the growing significance of this technology, this should also include essential AI infrastructure and the intellectual property behind the development of new AI models that might otherwise be attractive targets for rival governments.

Cyber Point of View: Australia

The power of public/private partnerships

In October 2023, Microsoft and the Australian Signals Directorate (ASD) announced the Microsoft-ASD Cyber Shield (MACS) initiative. This unique public-private partnership was created to enhance cybersecurity collaboration between the two organizations to protect the Australian Government, businesses, and citizens.

The MACS partnership shows how closer public and private sector collaboration can act as a force multiplier in the fight against cybercrime and aggression. In January 2024 for example, the Australian Government announced it had identified and issued sanctions against the perpetrators of a 2022 ransomware attack against Medibank, Australia's largest medical insurance company using evidence provided by the Microsoft Threat Intelligence Center (MSTIC).

Joint engineering between Microsoft and ASD also produced a world-first free-to-download connector for Microsoft Sentinel customers to participate in the ASD's country-level Cyber Threat Intelligence Sharing (CTIS) platform. This lowered the barrier to entry for Sentinel customers participating in CTIS and strengthened the platform by enabling more organizations to participate, increasing the cyber resiliency of the country.

Links

Microsoft announces A$5 billion investment to help Australia seize the AI era | Oct 2023

Working with the Australian Signals Directorate to hunt threat actors | Jan 2024

Microsoft, ASD Join Forces: Uniting Sentinel and CTIS for Enhanced Resilience | Mar 2024

Election interference

The goal of some nation-state-backed threat actor groups is to influence and undermine the results of democratic elections. These efforts to manipulate electoral outcomes underscore the need for continued vigilance and collaboration, enhanced defensive measures, and content authenticity indicators such as content provenance.[13] Defending elections against influence campaigns—as well as opportunistic cybercriminal efforts—demands a collective commitment from industry, media, and governments alike.

Nation-state threat actors and elections

By the end of 2024, approximately two billion people will have had the opportunity to vote in nationwide elections. The widespread accessibility of generative AI tools coupled with significant geopolitical events has created a ripe environment for nation-state influence operations aimed at high-stakes contests. Russia, Iran, and China all engaged in election influence efforts in 2024, with Russia implementing the most wide-reaching, persistent campaigns and Iran coming into the cycle later.

Continuing its well-known influence efforts in democratic processes, Russian influence actors deployed a spectrum of covert and semi-covert operations aimed at undermining trust in democratic institutions across Europe and in the United States, with the goal of eroding support for Ukraine.

Iran and China, meanwhile, escalated their influence capabilities and objectives throughout 2024. Iranian influence actors increasingly tried to influence elections in the Middle East and in the US, to include Israel's February 2024 municipal elections and the US 2024 presidential election. Cyber personas run by Iranian influence actors sought to highlight vulnerabilities in Israel's government and elections infrastructure as well as Prime Minister Netanyahu's failure to secure voting during the election to elicit a sense of insecurity among Israelis. Weakness and vulnerability are common themes of Iranian influence operations, and Iran will likely continue to use them in upcoming elections.

Russia, Iran, and China influence efforts converge on US election

Russian influence actors launched operations aimed at the 2024 presidential election at a slower pace than in previous election cycles. Nevertheless, Russian influence actors Ruza Flood (aka Doppelganger), Volga Flood, and Storm-1516 demonstrated the ability to create dynamic and creative content aimed at American audiences.

Ruza Flood's US election-themed websites, using names like "50 States of Lies" and "Election Watch", spread anti-Ukraine, anti-US propaganda across social media platforms. Meanwhile, videos by Russian influence actor Storm-1516, such as a staged video depicting the burning of an effigy of Donald Trump, received coverage from several major international media outlets.[14] Russia also continued to leverage agents-of-influence—for example, resurrecting Russian agent Andrei Derkach's NABU Leaks campaign, which was sanctioned by the US in 2021 for malign influence in the 2020 election.

In May of 2024, Iran began preparations for influence operations ahead of the US elections in two ways: conducting cyber intrusions into political accounts potentially for hack-and-leaks and launching a stream of polarizing content on covert news sites.

In mid-June, Mint Sandstorm sent a spear-phishing email to a high-ranking official of a presidential campaign from a compromised email account of a former senior advisor. Days earlier, the same actor, which we assess is connected to the IRGC intelligence unit, also unsuccessfully targeted an account of a former presidential candidate.

Iran also likely ran a network of websites masquerading as news outlets that actively engaged US voter groups on opposing ends of the political spectrum with polarizing messaging on issues such as the US presidential candidates, LGBTQ rights, and the Israel-Hamas conflict. Microsoft found evidence indicating the sites are using AI-enabled services to plagiarize content from US publications. Examination of source code and indicators in the articles suggest the sites' operators are using search engine optimization (SEO) plugins and generative AI-based tools to create article titles and keywords and to automatically rephrase stolen content in a way that drives traffic to their sites while obfuscating the content's original source.

Meanwhile, China's use of covert social media networks to sow discord ahead of the presidential election suggests the Chinese Communist Party (CCP) was emboldened by its 2022 midterm elections influence campaign, the first time it was observed attempting to interfere in a US election.

In late April, CCP-linked influence actor Taiji Flood (previously Storm-1376 and commonly referred to as "Spamouflage") launched an influence campaign leveraging the surge of Israel-Palestine-related protests on US college campuses. Some of Taiji Flood's personas on Telegram implied that they themselves were students or parents of students involved in the protests, and injected left-leaning messages into right-wing groups. They likely did so to sow conflict about the protests, or perhaps they misunderstood which audiences would be most receptive to their message.

The convergence and parallel nature of nation-state operations throughout 2024 underscores just how persistent adversarial states are in their attempts to exert influence over US electors and outcomes. Left unchecked, this poses a critical challenge to US national security and democratic resilience.

Links

Combatting the deceptive use of AI in election (microsoft.com)

Iran Targeting 2024 US Election - Microsoft On the Issues | Aug 2024

Russian US election targets support for Ukraine after slow start | Apr 2024

Expanding our Content Integrity tools | Microsoft On the Issues | Apr 2024

Content Credentials

Microsoft's efforts to enhance the security of Indian elections | Jun 2024

Addressing the deepfake challenge ahead of the European elections | May 2024

Elections create another opportunity for impersonation threats

Microsoft observed a surge in election-related homoglyph domains delivering phishing and malware payloads. We believe these domains are examples of cybercriminal activity driven by profit and reconnaissance by nation-state threat actors in pursuit of their own political objectives. Homoglyph domains are fraudulent domains that exploit the similarities of alphanumeric characters to create deceptive domains to impersonate legitimate organizations.

Threat actors use these malicious domains to deceive victims, often in combination with credential phishing and account compromise.

During an election cycle, there is significant focus on domain infrastructure to host campaign content and mail domains to communicate with supporters and voters. This increase in domains creates opportunities for cybercriminals and nation-state actors, who may use impersonation for political or criminal reasons.

Using data from previous attacks, the Microsoft Digital Crimes Unit has set up monitoring for domains related to elections around the world in an effort to detect impersonations. Our objective is to ensure Microsoft is not hosting malicious infrastructure and inform customers who might be victims of such impersonation threats. At present, we are monitoring over 10,000 homoglyph domains.

We note, however, that homoglyph domains are often registered by legitimate companies—either defensively to prevent abuse or for profit with the goal of eventually selling the domain.

Target domain	Homoglyph domain	Technique	Payload delivered
crd.org	crd.com	org to com	Phish
crd.org	crd.com	org to com	Malware
gop.com	gop.com	domain o for g	Phish
gop.com	gops.com	domain with s	Phish
gop.com	go.com	drop terminating letter	Phish
rnc.org	rnc.com	org to com	Phish
rnc.org	rnc.com	org to com	Malware
dnc.org	dnc.com	org to com	Phish
dnc.org	dn.org	drop terminating letter	Phish
dccc.org	dccc.com	org to com	Phish
nrcc.org	nrcc.com	org to com	Phish
gnsa.com	gns.com	drop terminating letter	Phish
mymgp.com	myng.com	drop terminating letter	Phish
ngpweb.com	ngpwe.com	drop terminating letter	Phish
wawd.com	waw.com	drop terminating letter	Phish
wawd.com	waw.com	drop terminating letter	Malware

Source: Microsoft Threat Intelligence

Examples of homoglyph techniques

Original	Replacement
w	vv
0	o
.org	.info
.org	.com
.gov	.org
.com	.org
.uk	.co.uk
.com	.cam
m	rn
g	q
l	ll
l	ii
l	ii
l	ll
Domain address structure	Add or remove an "s" at the end of a string

Actionable insights

1. The US Cybersecurity and Infrastructure Security Agency (CISA) and Federal Bureau of Investigation (FBI) recommend that all election offices adopt a .gov domain to help mitigate impersonation and cybersecurity risks. This is because .gov domains are only available to US-based government organizations and publicly controlled entities, which helps the public recognize official government sites and emails and avoid phishing attempts and websites that impersonate government officials.
2. Use defensive registrations of obvious homoglyphs of your organization's domains to prevent them being used in a cyberattack.

Ransomware

Landscape and trends

Ransomware remains one of the most serious cybersecurity concerns. And for valid reasons.

Among our customers, Microsoft observed a 2.75x increase year over year in human-operated ransomware-linked encounters (defined as having at least one device targeted for a ransomware attack in a network).

Meanwhile, the percentage of attacks reaching actual encryption phase has decreased over the past two years by threefold. Automatic attack disruption contributed to this positive trend in decreasing successful attacks. In more than 90% of cases where attacks progressed to ransom stage, the attacker had leveraged unmanaged devices in the network, either to gain initial access or to remotely encrypt assets at the impact stage.

3x

threefold decrease in ransom attacks reaching encryption stage over the past two years

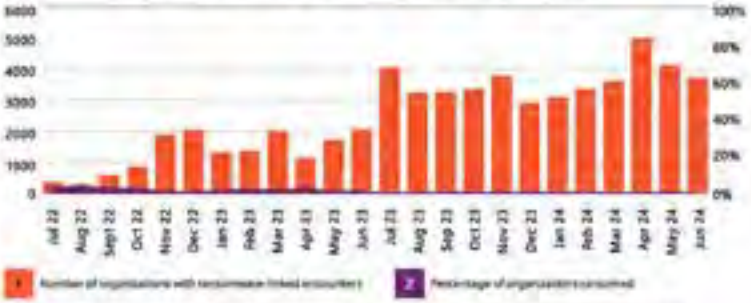
Top human-operated ransomware groups

Sector	Percentage
1 Akira	17%
2 Lockbit	15%
3 Play	7%
4 Blackcat	6%
5 Bazta	6%

The top five ransomware families accounted for 51% of attacks. These families continue to use longstanding techniques, showing their effectiveness even against rising cybersecurity awareness globally.

Source: Microsoft Defender for Endpoint

Organizations with ransom-linked encounters continues to increase while the percentage of those ransomed is decreasing (July 2022–June 2024)



Although organizations with ransom-linked encounters continues to increase, the percentage that are ultimately ransomed (reaching encryption stage) decreased more than threefold over the past two years.

Source: Microsoft Defender for Endpoint

The most prevalent initial access techniques continue to be social engineering—specifically email phishing, SMS phishing, and voice phishing—identity compromise, and exploiting vulnerabilities in public facing applications or unpatched operating systems. Attackers continue to take advantage of newly identified common vulnerabilities and exposures (CVE) with Common Vulnerability Scoring System (CVSS) scores above 8. Once the attacker is in the network, they tamper with security products or install remote monitoring and management tools (RMMs) to disable or evade detections and persist in the network.

We observed remote encryption in 70% of successful attacks, with 92% originating from unmanaged devices in the network, underscoring the need for organizations to enroll devices into management, or exclude unmanaged devices from the network.

How cybercriminals are tampering with security products

After compromising an organization, threat actors usually begin by tampering with its security solutions.

By disabling or tampering with defenses, attackers buy themselves time to install malicious tools, exfiltrate data for espionage or extortion, and potentially launch attacks like ransomware.

Microsoft consistently observes a prolific number of attacks involving antivirus tampering. In May 2024, Microsoft Defender XDR detected over 176,000 incidents involving tampering with security settings, impacting more than 5,600 organizations. On average, during that time frame, organizations that encountered tampering activity saw over 11 attempts.

In May 2024, we detected over 176,000 incidents involving tampering with security settings.

After gaining a foothold in a network, attackers conduct reconnaissance to determine security tools in place or they might test security measures by dropping tools or payloads like commodity malware. If detected and blocked, actors may instead tamper with the security products they encounter. Attackers generally seek to gain access to privileged accounts within a compromised environment so they can use elevated privileges to configure any policy settings, including security setting modification.

Microsoft has observed various techniques to disable or otherwise tamper with security policies, including Windows Registry modifications; malicious tooling such as NSudo (Defeat Defender), Defender Control, Configure Defender, and ToggleDefender; custom malicious PowerShell or batch scripts and commands; and driver tampering.

Actionable insights

1. Some endpoint detection and response (EDR) solutions provide tamper protection features that can help prevent attackers from disabling security settings.
2. Organizations can configure the Disable Local Admin Merge setting to limit the ability to make local administration changes antivirus policy settings.
3. Alerts that detect tampering tools and activity might precede the delivery of additional malware or the launch of malicious commands and should respond accordingly. As a result, these notifications should be actioned immediately.

Links

Protect security settings with tamper protection | Microsoft Learn | May 2024

Configure local overrides for Microsoft Defender Antivirus | Microsoft Learn | Jul 2024

Octo Tempest: a case study and a cautionary tale

A notable development in the evolution of ransomware attacks since last year's report is the increase in hybrid attacks targeting both on-premises and cloud assets.

At a time when sophisticated threat actors are continuing to add new tactics, techniques, and procedures (TTPs) to their already wide-ranging playbooks, the threat actor Octo Tempest (aka Scattered Spider) offers a good example of this evolution and growth.

Octo Tempest is a financially motivated cybercriminal group known for wide-ranging campaigns that feature adversary-in-the-middle (AiTM) techniques, social engineering, and SIM swapping capabilities. First observed in 2022, it targeted mobile telecommunications and business process outsourcing organizations to initiate phone number ports (SIM swaps). By mid-2023, Octo Tempest had become an affiliate of ALPHV/BlackCat, a human-operated ransomware-as-a-service operation, and began deploying ALPHV/BlackCat ransomware payloads to victims. By the second quarter of 2024, Octo Tempest added Qilin and Ransomhub to their ransomware payloads.

Tactics, techniques, and procedures used by Octo Tempest

Initial access	Discovery	Credential access, lateral movement	Defense evasion, execution	Persistence	Actions on objective
Social engineering	Enumerating internal documentation	Identifying high-value assets	Leveraging EDR and management tooling	Installing a trusted backdoor	Staging and exfiltrating stolen data
Manipulating and impersonation	Continuing environmental reconnaissance	Accessing enterprise environments via VPN	Circumventing Conditional Access	Manipulating existing accounts	Deploying BlackCat ransomware
		Collecting additional credentials		Establishing access to resources	

Octo Tempest leverages a diverse array of TTPs to navigate complex hybrid environments, exfiltrate sensitive data, and encrypt data and leverage tradecraft that many organizations don't have in their typical threat models, such as SMS phishing, SIM swapping, and advanced social engineering techniques.

Source: Microsoft Threat Intelligence

Octo Tempest uses extensive social engineering techniques, including researching an organization to identify targets and then impersonating employees or members on phone calls to trick technical administrators into performing password resets or resetting multifactor authentication (MFA) methods. The group also uses SIM swapping to gain access to an employee's phone number and then initiate a self-service password reset of the user's account.

Octo Tempest uses its initial access to carry out broad searches across the network to identify documents related to network architecture and other sensitive intelligence, then explores the environment to enumerate assets and resources across cloud environments.

The group uses device management technologies to push additional malicious tooling, disable/evade security products, or create new virtual machines inside the organization's cloud. In addition to asset encryption, the group targets data exfiltration using Azure Data Factory and automated pipelines to extract data to its Secure File Transfer Protocol (SFTP) servers.

Disrupting ransomware threat actors

The past year has proved once again that defeating ransomware threats requires a layered and multi-tiered approach.

One of those tiers needs to focus on disrupting actors responsible for this activity in the real world. Microsoft teams have served as leaders in bringing together experts from industry and law enforcement to share threat intelligence and evidence about ransomware actors, their infrastructure, identities, and even finances.

In May 2024, the US Federal Bureau of Investigation (FBI) Cyber Division identified Octo Tempest as its third highest priority behind China and Russia nation-state threat actors. During the period covered in the scope of this report, Microsoft contributed intelligence and evidence essential to the arrest of multiple Octo Tempest members and other law enforcement disruptions of ransomware actors. We believe our contribution to these public-private partnerships helps collectively erode the technical capabilities and infrastructure of the group, ultimately leading to its dismantlement.

Microsoft is aggressively pursuing our ability to share information as authorized by law and policy to combat the most significant threats to our customers and our business.

> The US Federal Bureau of Investigation Cyber Division identified Octo Tempest as its third highest priority behind China and Russia.

While our specific role may differ in every operation, Microsoft's threat intelligence and law enforcement liaison capabilities are, and will continue to be, brought to bear against the most significant threats we face from criminal and nation-state threat actors.

Links

- Ransomware operators exploit vulnerability for mass encryption | Jul 2024
- Download Ransomware Incident Response Playbook Template | May 2023
- Defend against ransomware with Microsoft Security | Nov 2023
- What is Ransomware? | Microsoft Security
- Octo Tempest crosses boundaries to facilitate extortion, encryption, and destruction | Oct 2023

Cyber Point of View: Israel

Combatting ransomware collectively

In 2024 the Israel National Cyber Directorate, Cyber Security Council of the United Arab Emirates, and Microsoft Israel joined forces to create a collaborative threat intelligence platform, "Crystal Ball," for use by the International Counter Ransomware Initiative (CRI), a new 60-country coalition. The domains of the platform are Attribution, Deterrence, and Culture, which address cybersecurity collaboration between nations.

The Crystal Ball Platform is designed for modern work with embedded security, automation, and AI. The platform also considers data residency and geographic regions to meet the regulatory standards of the CRI partners. As of June 2024, more than 10 countries are using and sharing intelligence on the platform, with the goal to onboard the remaining CRI members by the end of 2024.

Fraud

Landscape and trends

Incidents of fraud and abuse are increasing globally in both volume and sophistication. Fraud is a form of cybercrime and it undermines the security, trust, and reputation of individuals, businesses, and organizations of all sizes and types, in every region and industry.

From nation-state actors to cybergangs to lone fraudsters, malicious actors exploit vulnerabilities in services, programs, online properties, promotions, and systems to obtain fraudulent access. They use gained resources for cyberattacks, financial crimes, or reselling assets. The World Economic Forum[11] reports scammers stole over $1 trillion US globally from victims in 2023. This means companies lost an average of 1.5% in profits due to fraud,[12] while consumers faced a staggering $8.8 billion US in losses—up 30% from 2022.

In an era where digital transformation accelerates almost every type of business operations, the ingenuity and scalability of fraud tactics continues to challenge resilience around the world. Organizations face a barrage of scams, such as payment and quick response (QR) code fraud, business email compromise (BEC), AiTM, video phishing, and investment scam techniques such as "pig butchering."

At the same time, the fight against impersonation is getting significantly more difficult due to the increasing ease of access to deep-fake technology, which enables cybercriminals to create highly convincing forgeries of not only the voices of business leaders but even video.

The shift to cloud-based computing is proving a double-edged sword. While cloud computing provides scalability, elasticity, cost savings, and enhanced computational capabilities that drive innovation, it grants these same advantages to malicious actors, amplifying their potential for misconduct. Microsoft has observed fraudsters using cloud services to launch attacks, steal data, impersonate users, launder money, and evade detection. These activities are used in various types of fraud such as account takeover, domain typo-squatting, payment fraud, and other types of cloud impersonation.

As discussed on the following pages, Microsoft collaborates with law enforcement, industry partners, and customers to actively combat these illegal activities, to protect and uphold the rights of our customers.

The ever-growing threat of cyber-enabled financial fraud

Cyber-enabled financial fraud covers a range of fraudulent activities facilitated by the internet, including investment scams, BEC, and tech support scams.

According to the FBI, losses due to investment scams have surpassed all other online fraud types, accounting for more than $4.5 billion US in losses in 2023 alone.[24]

Teams at Microsoft, LinkedIn, and Skype are advancing efforts to proactively detect such criminal activities, and Microsoft suspended upwards of 64 million abusive service accounts in 2023. We are also working with industry and law enforcement partners to disrupt these actors in the real world. In addition, we are currently working with law enforcement partners to improve intelligence exchange on cyber threats to dismantle criminal operations.

For example, in May 2024, Microsoft worked with the Indian Cybercrime Control Center to shut down over 1,000 Skype accounts involved in harassment, blackmail, extortion, and fake "digital arrests" by fraudsters impersonating police and other officials.[25]

The Microsoft Digital Crimes Unit combats financially motivated cybercrime using pioneering legal strategies, state-of-the-art technology, and collaboration to counteract threats to consumers.

Microsoft also works to undermine cybercriminals by proactively dismantling their operational infrastructure, disrupting their financial motivation, and partnering with organizations like the US National Cyber Forensics and Training Alliance and the Japan Cybercrime Control Center to enhance sharing of actionable intelligence. As the world recognizes the persistent threat of cyber-enabled crime, we are seeing more public and private partners joining forces to disrupt criminal infrastructure, hold them to account, and support victims of cybercrime.

Authorities conducted searches at seven locations across India, intercepting live cybercriminal operations and gathering substantial evidence, including computer hard disks, mobile phones, and laptops, along with details of financial transactions, call recordings, and transcripts. A total of 43 individuals have been arrested in this operation, with many others still under investigation. The CBI, in coordination with the FBI and international law enforcement agencies, continues to trace the network's operations and financial activities to identify and apprehend additional suspects in this ongoing investigation.

Our collaboration with law enforcement to combat cyber-enabled fraud has resulted in 30+ call center raids, 100+ arrests, and increasingly severe prison sentences worldwide.[26]

In October 2023, Microsoft and Amazon joined forces to combat a cybercriminal network conducting tech support fraud against more than 2,000 customer organizations globally. Following a joint investigation, we provided India's Central Bureau of Investigation (CBI) a criminal referral identifying multiple companies, call centers, and individuals that directly contributed to "Operation Chakra II," a law enforcement operation of more than 75 criminal raids across India to dismantle organized cyber-enabled financial crimes.[27]

Despite efforts by law enforcement and partners in the public and private sector, the complexity, speed, impact and severity of cybercrime is escalating.

Cybercriminals are leveraging the growing cybercrime-as-a-service (CaaS) ecosystem as well as AI technologies to launch phishing and social engineering attacks at scale. Simultaneously, they are increasingly evading security measures like multifactor authentication (MFA) to conduct targeted attacks. As a result, the battle against cyber-enabled financial fraud requires a multi-faceted response. Enhancing cooperation and strengthening detection and prevention measures are key areas of focus. Public awareness, vigilance, and the facilitation of fraud reporting are also vital components in preventing these crimes and mitigating their impact.

Novel trends and nightmare scenarios in the world of e-commerce

Even as card-present payment security improves through mobile wallet, Eurapay, Mastercard, and Visa (EMV) chip, and near-field communication (NFC) technologies, fraudsters remain attracted to the e-commerce or card-not-present (CNP) space, in which payment cards are not physically present for the transactions.

By 2026, the annual losses attributed to e-commerce payment fraud globally are expected to surpass $90 billion US,[28] with merchants and financial institutions bearing much of that economic impact.

Card security has seen numerous advances, including MFA, tokenization,[29] and the expansion of address verification services. However, issues with interoperability and incomplete implementation of planned improvements prevent these technologies from being universally adopted.

In the past year, Microsoft conducted over 1.5 billion risk evaluations for potential payment fraud and rejected $1.58 billion US in fraudulent transaction requests. We've observed a rise in sophisticated fraud tactics targeting online transaction vulnerabilities, including web interface breaches, phishing, spoofing, and synthetic identity generation to steal credentials and payment instrument information.

Traditional methods, such as exploiting large-scale data breaches, remain prevalent, enabling fraudsters to bypass identity validations and access extensive personal data. The growth of the CaaS economy also simplifies the execution of complex fraud schemes by providing ready access to stolen data and fraudulent tools. Concurrently, we're observing a shift away from older hacking techniques in favor of methods like phishing and spoofing to compromise credentials and gain access to payment instruments.

Generative AI accelerates the creation of fake identity elements, such as high-quality images, deepfakes, and voice impersonations, making it easier to deceive merchants and individuals. This falsified identity information can either disguise the fraudster's true identity or impersonate a victim to fool a merchant, or impersonate a trusted contact to fool a victim. Such deception can trick a merchant's risk engine during transactions or, if initially detected, can persuade customer support to override the rejection. Consequently, this enables fraudsters to unlawfully obtain goods or services using stolen payment methods.

In addition to the above general trends, Microsoft has observed the following specific methods used in e-commerce payment fraud:

- Enumeration techniques pose significant risks in e-commerce as regulatory compliance does not mandate the protection of all digits in the 16-digit card schema, allowing some digits to be guessed. Fraudsters use public payment schemes and automated methods to deduce authentication details like Card Verification Value (CVV) codes or expiration dates. Once they generate valid payment credentials, these can be sold on the dark web.
- Biometric spoofing and the creation of synthetic identities using generative AI are increasing threats. AI-generated deepfakes can bypass biometric security in many mobile payment methods relying on biometrics technologies native to hardware and operating systems. Additionally, fraudsters use AI to craft realistic synthetic identities to manipulate merchant customer support functions.

Actionable insights

1. Incorporate AI and machine learning (ML) models into existing policies and rules to detect unusual transaction patterns and flag potentially fraudulent activities in real-time.
2. When using voice as a factor for authentication, be sure to incorporate additional factors due to the rise of AI audio models capable of reproducing individual voices.
3. Apply risk-based containment strategies using trend product access and customer behavior monitoring to manage malicious use of AI and fake identities.
4. Deploy robust authentication measures to verify payment credentials and use tokenization to eliminate the need to store full card numbers.
5. Collaborate with industry partners using secure technologies like confidential computing and clean room environments to enhance data sharing and fraud prevention while protecting privacy.
6. Enhance authentication with phish-resistant FIDO.

Phishing

Phishing remains a perennial cybersecurity threat. According to TrendMicro, phishing attacks increased by 58% in 2023, with an estimated financial impact of $3.5 billion US in 2024.[40]

Threat actors continue to use longstanding and new TTPs to access targets, but a growing concern this year is the misuse of legitimate web services and tools for phishing deployment.

Software-as-a-Service (SaaS)-based email, developer tools, captcha services, cloud storage, click tracking, marketing platforms, customer survey platforms, lesser known email clients, and backup and mass emailing tools have all been weaponized for a range of malicious activities. One of the key advantages of using these services is that they can evade detection systems because they are less likely to be preemptively blocked due to their established levels of trust and legitimate usage. Additionally, many phishing campaigns combine the use of multiple legitimate services simultaneously, complicating the detection process for both human analysts and automated systems.

775 million

email messages contained malware (July 2023-June 2024)

Source: Microsoft Defender for Office 365

Top email phishing types

Source: Microsoft Defender Threat Expert notifications

QR code phishing

QR codes have become much more prevalent since the pandemic and therefore more trusted by users. By their nature, QR codes obscure the destination from the user. Around mid-September 2023, Microsoft analysts observed a significant increase in phishing attempts using these codes, which presents a unique challenge for security providers as they appear as an image during mail flow and are unreadable until rendered.

In QR code phishing, threat actors send phishing messages containing a code encoded with a URL. The message prompts the recipient to scan the QR code with their device, redirecting them to a fake sign-in page where they are prompted to input their credentials. This page may include A/TM capabilities that circumvent some forms of MFA.

Of note, Microsoft Defender for Office 365 image detection technology significantly disrupted QR code phishing attacks, causing a 94% decrease in phishing emails using this attack technique between October 2023-March 2024. Threat actors continually adapt their attacks in response to effective detection and blocking, returning to older, more well-known tactics or spending time and effort to innovate new techniques. Disruption of effective techniques such as QR code phishing is an important part of network and data protection. While QR codes that are plain and square with a black and white barcode persist in phishing attacks, since more effective detection and blocking has successfully reduced their efficacy, threat actors have resorted to experimenting with different QR code visualizations. For example, QR codes comprised of a blue barcode on a red background became commonplace in phishing attacks until that variation, too, was rendered ineffective.

By their nature, QR codes obscure the destination from the user, which creates a challenge for security teams.

Other attempts to confuse image processing include placing a black border around the QR code or putting the QR code in an attachment. Although effective detection and blocking measures greatly reduced the volume of QR code phishing attacks from the millions of messages observed in 2023, actors continued to test new techniques and innovate old ones throughout 2024 to find ways to evade protections and impact organizations.

During April and May 2024, there was a surge in phishing attempts targeting Microsoft Teams users. The attackers created a new tenant and registered a domain specifically for this attack. These attacks often included QR codes within the messages, which, upon scanning, directed users to A/TM phishing sites. Common themes in these phishing campaigns involved impersonating Office, Microsoft or Security services.

Links

To scan or not to scan: The shady side of QR codes – Microsoft 365 | May 2023

Expanding our Content Integrity tools ahead elections – Microsoft On the Issues | Apr 2024

Created Credentials

Build trust with content credentials in Microsoft Designer | Learn at Microsoft Create | Dec 2023

Overview – C2PA

Business email compromise (BEC)

BEC attacks remain a prevalent threat, with inbox rule manipulation the favored method.

Inbox rule manipulation: A new variation has emerged involving manipulation through API/App usage. Instead of using the usual "New-InboxRule" or "Set-InboxRule" commands, the attackers now use "UpdateInboxRules". This allows them to redirect emails with keywords related to credentials or financial matters to less monitored folders like Spam, Conversation History, or Deleted Items, hiding their fraudulent activity from the user's immediate view.

BEC lateral phishing: After compromising an account, attackers aim to move laterally within the organization, targeting multiple users to either gain access to high-privilege accounts or trick users into paying fake invoices. This is achieved by sending phishing emails to other users within the organization.

Conversation hijacking: The attacker compromises the sender's email account and injects themself into an existing email thread using a similar-looking account, keeping the sender's display name unchanged. The hijacked account domain is usually newly created for financially motivated scams to lure users.

MFA tampering post AITM attack: After compromising a user account, the attacker attempts to add an additional device for MFA, such as a phone number to approve two-factor authorization or registering a new device with an authenticator, to maintain ongoing access.

Other noteworthy post-compromise behaviors observed

Legitimate applications abuse: We observed three new legitimate tools being abused by adversaries for mailbox exfiltration and BEC.

- PerfectData Software: An application integrated with Microsoft 365/Azure to provide a mailbox and backup services. Threat actors used it to secretly access and steal mailboxes from compromised users.
- Newsletter Software Supermailer: Legitimate software used to create and send personalized bulk emails and newsletters. Adversaries exploited it to conduct lateral phishing attacks from compromised user accounts.
- eMClient: A desktop email client for Windows and macOS. Adversaries used this to exfiltrate mailboxes from compromised users.

Low and slow BEC: Attackers discreetly read a small number of emails (between two to five) daily, and sparingly accessed OneDrive/SharePoint files, all in an effort to evade detection. These low-profile attacks challenged detection systems, which could identify them only by correlating with unusual sign-in activities.

Targeted BEC: Personalized phishing campaigns are crafted using local languages, targeting IT, finance, and legal departments with specific topics such as "software updates" or "tax submissions." This tailored strategy significantly boosts compromise success rates.

Top post-compromise BEC behaviors

Source: Defender Threat Signal Conditioning

Actionable insights

1. Even if a tool seems familiar, don't assume it is safe. In the past year, Microsoft identified a new trend where adversaries were misusing three legitimate tools for malicious activities.
2. Attacks are more sophisticated. Another notable trend is the personalization of phishing campaigns and outbound communications using local languages.
3. QR code phishing is on the rise but effective detection and blocking measures can greatly reduce the volume of attacks.
4. Shadow IT, or pieces of hardware or software that users install without the approval of the IT department, are a threat to the organizations and make them vulnerable to phishing and post-compromise activities. IT teams should periodically scan the infrastructure to detect unauthorized software or hardware and take remedial actions.

Impersonation

Impersonation is a key method used by fraud actors to gain trust from their victims. While it takes many forms, impersonation of legitimate companies is a risk both to customers and to the reputation of the business. As we look across the world of corporate impersonation, Microsoft has seen fraud actors increase sophistication and speed across the board.

Deepfakes

Most AI-generated synthetic media, also known as "deepfakes," target either communities (for example, false news reports) or individuals (scams).

As with other impersonation-based threats, deepfakes exist on a spectrum of sophistication. At its most basic, for example, a threat actor could use shallow fakes in email and text messages designed to convince workers that a superior or colleague needs them to take an action. However, the significant rise in sophistication on the horizon will produce major changes, including in identity verification.

54%

of phishing campaigns targeting consumers impersonated online software and service brands

Sectors impersonated in consumer phish

1. Software and services (54%)
2. Financial (15%)
3. Retail (12%)
4. Media and Entertainment (9%)
5. Freight and Logistics (6%)
6. All others (3%)

A majority of phishing campaigns targeting consumers this year impersonated online software and service brands. This comes as no surprise given the value to attackers in compromising and exploiting consumer accounts on platforms that can span social media, cloud storage, email, e-commerce, and more. Phishing will become less prevalent and less profitable for attackers as more consumers adopt strong MFA and passwordless technology.

Source: Outlook.com consumer phish reports. These include consumer emails received by our customers and reported by Microsoft as phishing.

According to Gartner, by 2026, 30% of enterprises expect to no longer consider facial biometric identity verification and authentication solutions to be reliable in isolation.

As deepfakes become more common in the business environment, organizations will have to implement countermeasures, such as requiring additional verification for transactions. At the same time, Microsoft is exploring provenance solutions to help increase transparency of online digital content.[9]

Corporate impersonation

Domain spoofing involves various classic techniques that enable impersonation of corporate entities. One such technique is the look-alike domains we identified earlier: homoglyph domains. The imposters rely on users not noticing the slight variation the characters used in the domain name, such as using zero instead of the letter "O" in a domain name. By tricking users into clicking links or visiting these fraudulent websites, the threat actors can deceive them into sharing sensitive information, potentially resulting in financial or identity theft.

Fraudsters have doubled down on various forms of impersonating domains including homoglyphs, sub-domain squatting, and plausible alternate domain registration. Sub-domain squatting involves setting up a sub-domain in a cloud service under a trustworthy name to carry out email-based attacks, such as using "contoso.onmicrosoft.com" as a sub-domain.

On the other hand, plausible alternate domain registration involves adding words or changing the top-level domain (TLD) to trick users. For instance, registering a domain like "contoso.store".

Microsoft has begun piloting Entra Verified ID as an element of advanced identity proofing, requiring users to share government-issued ID in situations where the authenticity of their identity is in question. Initial results indicate this control is having promising results, effectively preventing the majority of corporate impersonation attempts. As we continue to train the models and build for scale, we expect to further improve these results.

Actionable Insights

1. Educate employees to be on the lookout for impersonated domains.
2. When enabling accounts and services, use of verified identity models and AI detection can significantly reduce the risk of allowing fraudulent access.
3. Educate employees about how to check content integrity.

Links

Microsoft Content Integrity website

The dire state of techscam

Techscam refers to fraudulent activities or scams that make your computer vulnerable to additional malicious activities. Microsoft has identified that most techscams originate from malicious advertisement platforms. There are a variety of techscam types, each with its own unique modus operandus:

- **Microsoft Support techscam/McAfee/Apple techscam:** These scams impersonate legitimate support services from industry leading tech companies to deceive users into providing sensitive information or making payments for non-existent issues.
- **Cryptocurrency/fake shopping:** Scammers use malicious advertisements to promote fraudulent cryptocurrency schemes or fake shopping deals, luring users into financial traps.
- **Malicious browser extension scam:** These scams trick users into installing browser extensions that can manipulate search results, display intrusive ads, or steal personal data.
- **Malicious browser notification scam:** Users are misled into allowing browser notifications from malicious sites, which then bombard them with misleading alerts or phishing attempts.

> In the realm of cloud services, we saw a significant uptick in techscam traffic, with a daily frequency surging from 7,000 in 2023 to 100,000 in 2024, an over twelvefold increase.

The current landscape of techscam is alarming, with SmartScreen traffic statistics from 2022 to 2024 indicating that over 90% of malicious traffic in the Edge browser is attributed to techscam activities.

Among techscam frameworks, investment and cryptocurrency scams and technical support scams have incurred the highest financial losses globally. Overall, techscams have 10 times the financial impact of phishing.

The transient nature of malicious hosts on cloud servers—such as Azure, DigitalOcean, and CloudFront—poses a significant challenge to detection and neutralization. Cloud servers provide an easy and cost-effective way to create host pages. Moreover, over 70% of malicious entities are active for less than two hours, meaning they may be gone before they're even detected. This rapid turnover rate underscores the need for more agile and effective cybersecurity measures.

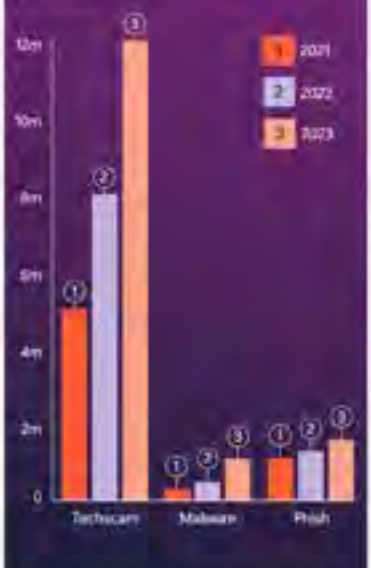
Microsoft researchers are building a new AI detection model to detect scams with a local small language model, protecting the first victim that sees a scam and reporting scams to SmartScreen to protect other users.

Microsoft researchers are also developing client-side signals to analyze the visual and structural elements of techscam pages. This new capability has significant detection potential, especially since techscam incidents are often short-lived. This feature enables us to identify threats more quickly, dramatically increasing the efficiency of our techscam detection abilities.

Actionable insights

1. Preemptively block known malicious domains by creating a blocklist based on the domain architectures—such as IP, Whois, and PDNS (protective domain name system)—and redirector chains' information in telemetry logs commonly used in techscam operations.
2. Perform continuous updates on this dynamic use of blocklists to stay ahead of scammers' evolving tactics.

Daily malicious traffic volume
(millions)



The daily volume of techscam traffic has escalated dramatically, skyrocketing since 2021, a stark contrast compared to malware and phishing over the same period.
Source: SmartScreen log data

Account takeovers (ATOs)

In the business-to-consumer world there has been a steady pace of account hacking, identity theft, and payment instrument abuse in context of ATOs.

Threat actors prey on the least protected accounts and the most vulnerable individuals, so consistently raising the bar on account protection is important. While the best practice for consumers is to have and keep current MFA while being vigilant about account monitoring, a whole industry has developed to help mitigate the impacts of successful compromises.

In recent years, industry reports have indicated an increasing threat in the business-to-business world and the arena of managed online services. Because compromise of the latter may impact the downstream customers of client businesses, this adds another layer of complexity and risk.

According to one study, 29% of internet users have now experienced ATO, up from 22% in 2021.[1] Business account takeovers rose from 19% to 21% in the same period.

According to Microsoft's trend monitoring, many Microsoft accounts that were taken over last year did not follow basic account security best practices, like using MFA. Despite the emergence of more advanced hacking techniques, most ATOs still happen through simple methods like password spraying, phishing, keylogging, and using passwords from previous attacks found on the web.

Recently, Microsoft has experienced a surge in attacks on "generic," "group," or "multiuser" accounts. These accounts typically have outdated passwords, and the original user may no longer be monitoring them. Generic accounts are commonly used by administrators for ease of maintenance but are harder to secure. Implementing effective multi-factor authentication is challenging, and identifying compromises through pattern recognition becomes difficult when multiple users share access. Additionally, account recovery for service providers is problematic, making it challenging to restore hacked accounts to their rightful owners.

Actionable insights

1. Keeping account security up to the latest standards is critical to protecting yourself from ATO risk.
2. Hackers look for weak security and seasonal opportunities when account owners are not focused on monitoring, to scale their attacks.
3. Consider moving your user base to an authentication app to enable an easy upgrade to new standards for MFA as they release.

Links

Threat actor leverages Quick Assist in social engineering attacks leading to ransomware | Microsoft Security Blog | May 2024

Identity and social engineering

Insights on identity attacks and trends

As organizations move to the cloud and adopt SaaS applications, identities are becoming increasingly crucial for accessing resources.

Cybercriminals exploit legitimate and authorized identities to steal confidential data, and access credentials in various ways like phishing, malware, data breaches, brute-force/password spray attacks, and prior compromises.

As in past years, password-based attacks on users constitute most identity-related attacks, supported by massive infrastructure that threat actors have dedicated to combing the digital world for passwords.

7,000

password attacks blocked per second over the past year

Microsoft Entra data shows that of more than 600 million identity attacks per day, more than 99% are password-based. Advances such as default security configurations and Conditional Access policies have helped more organizations embrace multifactor authentication (MFA), increasing adoption to 41% among Microsoft enterprise customers.[1]

However, as MFA blocks most password-based attacks, threat actors are shifting their focus, moving up the cyberattack chain in three ways:

1. Attacking infrastructure
2. Bypassing authentication
3. Exploiting applications

Attacks on identity infrastructure in the spotlight

Infrastructure attacks have become popular with sophisticated actors, both nation-state and criminal.

They can be difficult to detect without careful configuration monitoring, AI-driven threat detection, and log analysis. Once a threat actor infiltrates an organization's infrastructure, they make changes to maintain persistence and remain unnoticed.

For example, they may steal credentials to impersonate a non-human identity, elevate its permissions for a few seconds to create new credentials used to access and steal data, then return the application's identity to its previous state.

Actionable insights

1. Employ advanced monitoring and threat detection that uses AI to recognize outlier patterns.
2. Carefully monitor access and configuration changes to identity infrastructure.
3. Enhance monitoring for devices and networks on which identity infrastructure depends.

Threat actors are bypassing MFA, using innovative AiTM phishing attacks and token theft

As we highlighted last year, as organizations strengthen their authentication protocols with MFA, threat actors are pivoting to AiTM phishing attacks and to token theft. Token theft occurs after a user successfully authenticates and receives a valid token. The attacker then steals the token from the victim's device, from compromised routers or proxies, or from application or network logs. Although token theft results in far fewer identity compromises than password attacks, our detections indicate incidents have grown to an estimated 39,000 per day. Moreover, over the last year we've seen a 146% rise in AiTM phishing attacks, which occur when attackers trick users into clicking a link and completing MFA on the attacker's behalf.

Actionable insights

1. Retire passwords in favor of phishing-resistant, passwordless authentication methods such as passkeys.
2. Require all users to run on their devices as standard users and not as administrators.
3. Only allow access from managed and compliant devices.
4. Mitigate AiTM and token theft attacks with policies that require interactive strong authentication when anomalies are detected.
5. Use access policies to require token protection and prevent access from untrusted environments.
6. To reduce time to mitigation and increase detection capability, adopt applications that support continuous access evaluation.

Exploiting applications to access high-value resources

Threat actors are taking advantage of abandoned, unmonitored, and overprivileged cloud-based applications with insecure credentials so they can access high-value resources.

Most organizations carry substantial security debt in such applications. For example, developers may enable broad permissions and check credentials into code to facilitate application development and testing but then fail to correct these issues before the application ships.

In the past year, Microsoft found only 2.6% of workload identity permissions were used and 57% of workload identities were completely inactive.

Between January and June 2024, we detected over 1.5 million credentials (such as passwords or certificates) discoverable by attackers in locations such as source code repositories. In fact, 18% of code repositories we examined in the past year exposed such secrets.

These statistics underscore the importance of secure development practices, which include preventing secrets in code, securing test environments, minimizing permissions for applications, and retiring unused applications and tenants. Just as using phishing-resistant credentials greatly reduces the risk of identity compromise, using managed service identities eliminates the risk associated with managing service credentials in code.

Actionable insights

1. Use managed service identities instead of developer shared secrets.
2. Govern permissions to ensure identities, including workload identities, have only the privileges they need.
3. Secure test environments and retire unused applications and tenants.

Identity attacks in perspective

Password-based attacks continue to dominate, but can be thwarted by using strong authentication methods.

Less than 1% combined

MFA attacks

SIM swapping
MFA fatigue
AiTM

End-run MFA protection by intercepting security codes using stolen phone numbers, barraging users with MFA notifications until they approve, and capturing first and second factor credentials using fake replicas of legitimate websites.

Post-authentication attacks

Token theft
Consent phishing

Infiltrate a user's account after they authenticate by stealing a legitimate token created on their device and moving it to a device under the attacker's control, by searching source code repositories for Open Authorization (OAuth) tokens and other non-human credentials, or by tricking the authenticated user into granting permissions to malicious apps.

Infrastructure compromise

Often silently executed by professional groups or nation-state-backed threat actors with sophisticated operations, making them very hard to detect. Threat actors may compromise an on-premises federation server and copy its private signing key to forge tokens, compromise a privileged cloud user and add new federation contracts, or compromise a non-human workload identity and create new credentials with elevated privileges.

Source: Microsoft Threat Intelligence

Security to the max: the optimal mindset for security professionals

Outdated security programs leave configurations insecure. For example, virtual private networks (VPN) typically grant remote users access to the entire corporate environment instead of limiting access to specific applications. Password-only authentication configurations, exacerbated by archaic expiry and complexity policies, result in more than 99% of identity compromises. And many identity attacks rely on protocols such as IMAP, POP, and basic authentications that were once necessary but have long been replaced by modern protocols such as Open Authorization (OAuth) and OpenID Connect (OIDC).

Password spray attacks are authentication attacks that employ a large list of usernames and pair them with common passwords in an attempt to "guess" the correct combination for as many users as possible.

Technical debt makes maintaining a secure environment challenging

Security incidents are more prevalent in older versions of domain controller operating systems (chart data represents January-March 2024). Although Microsoft officially ended mainstream support of Windows Servers 2016 and 2019, many customers who experienced security incidents were still running domain controllers on these out-of-support versions.

Source: Microsoft Incident Response

"Secure by default" settings reduce identity compromises

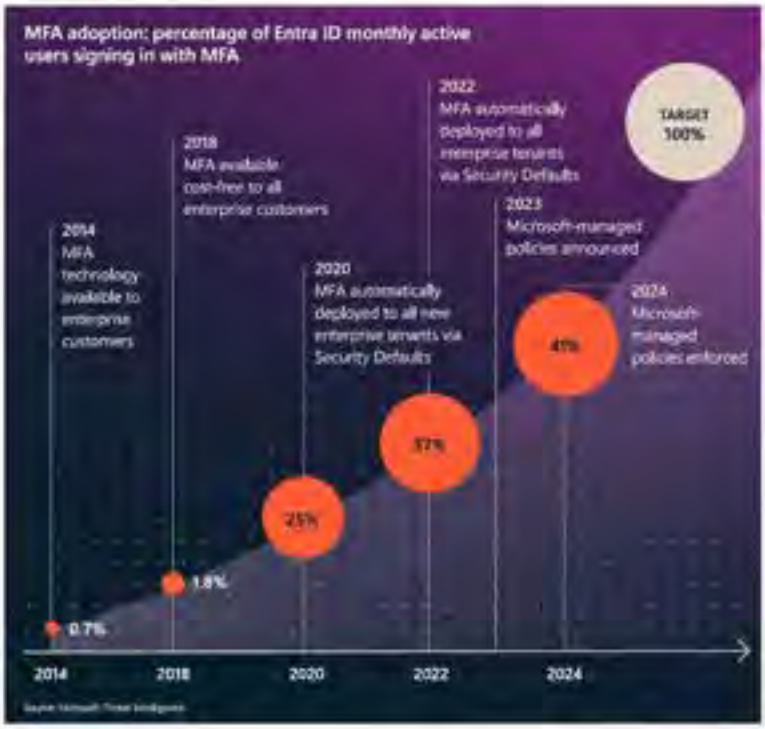
Although modern MFA techniques reduce the risk of identity compromise by 99.2%, many organizations have been slow to adopt them. So, in January 2020, Microsoft introduced "security defaults" that turn on MFA while turning off basic and legacy authentication for new tenants and those with simple environments. The impact is clear: tenants that use security defaults experience 80% fewer compromises than tenants that don't.

In November 2023, we started deploying Microsoft-managed Conditional Access policies for existing tenants with more complicated environments. We then started enforcing three MFA-related policies in March 2024. Even among these more sophisticated tenants with these policies, the compromise rate has decreased by 39%.

This "secure by default" approach represents a critical mindset shift for security professionals: instead of "dialing up" security to where it's "safe," we must start at the maximum level of security possible, then dial back as necessary. It's both harder and less effective to start at zero security and assemble it layer by layer than to start at 100 percent security and then customize the configuration to specific business needs.

Actionable insights

1. Enable MFA in all your tenants.
2. Enable phishing-resistant MFA for your admins.
3. For all new tenants, start with the strongest security settings possible.
4. Test pre-configured security settings, such as security defaults or managed Conditional Access policies, in report-only mode to understand their potential impact before going live.



Social engineering "next generation"

Regardless of the technique, social engineering remains a constant threat that ultimately cannot be fully mitigated via technology. Training and education, both at the helpdesk and user level, is central to preventing successful social engineering attacks.

Teams and Skype phishing

In recent years, there has been a significant rise in novel phishing techniques like the use of QR codes discussed earlier. Another of these techniques is the use of collaboration platforms such as Teams or Skype to phish users. Microsoft has observed threat actors using a previously compromised tenant to create a new onmicrosoft.com tenant with a tech support theme. These tenants are then abused to send malicious files, links, and requests for users to provide credentials or obtain MFA approvals.

Beginning in July 2023, threat actors began sending Teams users attachments hosted on an actor-controlled SharePoint tenant. In these cases, the target tenant had been configured to allow messages from users outside of the organization. In April 2024, Midnight Blizzard masqueraded as Microsoft Security on Teams, providing targets a link for AiTM credential harvesting. That month and the next, attackers impersonated Microsoft Office, Microsoft Security, or Microsoft in general to phish

> Regardless of the technique, social engineering remains a constant threat that ultimately cannot be fully mitigated via technology.

Teams users, often with QR codes that directed users to AiTM sites. In some cases, attackers had already obtained credentials through password spray attacks and sent a two-factor authentication request via Teams, prompting the target to provide the number in the authentication app. A sense of urgency was the key in these attacks since the cookies were valid only for a few minutes.

From April onwards, there was an uptick in social engineering attacks through Teams. Attackers posed as Help Desk or IT support staff and persuaded users to establish remote monitoring and management (RMM) connections with the attacker's system, leading to ransomware incidents. In all these Teams-related instances, the attacks were conducted through one-on-one conversations, originating from newly created attacker-owned tenants. SharePoint was frequently utilized by the attackers as the preferred platform for hosting malicious files.

SIM swapping

The growing acceptance of MFA has forced threat actors to impersonate users as a workaround. As a result, Microsoft has observed SIM swapping gaining popularity in recent years, led by Octo Tempest.

In SIM swapping, the threat actor contacts a mobile carrier and gets a target victim's SIM card moved to their own device. To do this, the actor must first collect personal information about the target to answer the security questions necessary to gain access to the target's account.

Once the actor has control of the victim's SIM, they can receive MFA codes and one-time passcodes. Operational security on the part of individuals is crucial in preventing this kind of attack. Individuals should monitor their online footprint to see what information is publicly available about them that a threat actor could use to impersonate them.

Helpdesk social engineering

Microsoft has observed an uptick in threat actors contacting helpdesks, impersonating a user to obtain a password reset or register a new MFA device.

In the last year, more than half of all Microsoft Incident Response engagements attributed to Octo Tempest were able to be tracked back to helpdesk social engineering. Helpdesks have begun to counter this by requiring further levels of verification such as video calls, but as noted earlier in this chapter, the rise of deepfakes will enable a threat actor to impersonate the voice, image, and video of a victim, making even this identity verification avenue more difficult.

Threat actors such as Octo Tempest have also been observed communicating directly with senior executives and other individuals involved in an investigation as part of their extortion campaign or in an effort to gain access to credentials. In cases where extortion is part of the attack, threat actors may also use text messages to pressure victims into paying.

How easy is it to carry out different types of social engineering attacks?

EASY	MEDIUM	HARD
Email and Teams phishing	Phone and video calls	SIM swapping
QR code phishing	Helpdesk social engineering	

Links

Midnight Blizard conducts targeted social engineering over Microsoft Teams | Aug 2023

Malware distributor Storm-0324 facilitates ransomware access | Sep 2023

Octo Tempest crosses boundaries to facilitate extortion, encryption, and destruction | Oct 2023

Actionable insights

1. Provide helpdesk staff training on identifying social engineering attempts.
2. Ensure helpdesk systems have adequate activity logs.
3. Move toward passwordless authentication; MFA alone is not enough.
4. Adopt phishing-resistant MFA for administrators.
5. Evaluate and strengthen helpdesk password reset procedures.
6. Employ alerting on any changes to administrators.
7. Engage in regular tabletop exercises.
8. Vet key suppliers relating to SIM cards and helpdesk services.

AiTM credential phishing

Credential phishing attacks with AiTM capabilities are continually observed by Microsoft through daily, high-volume email campaigns sent from phishing-as-a-service (PhaaS) platforms.

In 2024 to date, the top five kits by email volume were: Caffeine, Tycoon, Greatness, NakedPages, and Dadsec. Each of these PhaaS services represents tens to hundreds of millions of phishing messages observed each month.

While the top phishing services are largely the same in 2024 as 2023, there have been some changes. For example, in November and December 2023, the Dadsec service disappeared from our tracking. In January 2024, the creator and operator of the Dadsec PhaaS platform, Storm-1575, resurfaced and announced a rebrand of the service as "Rockstar2FA." Operations of the new service continued mostly as before, with intermittent updates to phishing attachments, messages, or infrastructure to evade detection efforts and a new communications channel for clients.

In May 2024, Storm-1101, the actor behind the NakedPages PhaaS platform, announced they would be permanently winding down their operations. The actor claimed they had provided the NakedPages source code to some individuals who had worked as support for the NakedPages service. At least one of these individuals has since started their own phishing service based on that source code.

In January 2024, Caffeine was rebranded to ONNX. Communication channels for the kit's operations were changed and the service began allowing customers to use their own domains in April, making it harder to track activity related to the kit. While Caffeine/ONNX was the most prominent AiTM phishing service by volume of phishing messages observed through the first half of 2024, it was supplanted by Tycoon in May. In June, Caffeine owner and operator Storm-0867's identity was revealed in a blogpost from DarkAtlas,[19] resulting in an abrupt cessation of operations.

The use of HTML attachments to deliver URLs or phishing pages to recipients continued to be a popular tactic among phishing actors in the last year. The attached HTML file may contain a URL that sends the recipient to a phish landing page or it may contain code that reaches out to an actor-controlled server to download a phishing page and present it to the recipient upon opening the file. HTML files may also be contained within ZIP files, Microsoft Office files, additional—sometimes multiple—email files attached to the initial email, or other file types to evade detection.

PDF attachments also continued to be a popular vector for phishing. Usually, the PDF contained a URL leading to a phish landing page, likely through a multi-step process including a redirection URL through a legitimate or abused service and/or a captcha check. Occasionally, the link went straight to a phishing domain. Like the HTML vector, a PDF file may be included within multiple layers of other email files, ZIP files, other filetypes or may be hosted on a legitimate filesharing service accessed through a URL provided in the original email.

Stormy skies: the rise of cloud identity compromise

With more organizations moving to hybrid or cloud-only models, it is becoming increasingly important to secure both cloud resources and cloud identities.

Identity is a central piece in a functional organization's cloud environment and represents a critical target for attackers. An attacker who manipulates identity can also manipulate any resource or process that identity is trusted to access, including email, other cloud services, or the on-premises environment.

In the past, cloud identity compromise was thought to be reserved for only a handful of advanced, perhaps exclusively state-sponsored, actors. However, financially motivated actors like Octo Tempest, Storm-0539, and Storm-0501 have recently shown sophisticated competency in the cloud across a large variety of industry verticals, indicating that more and more threat actors will be able to use this technique.

In Octo Tempest cloud attacks identified by Microsoft Threat Intelligence, this prolific threat actor targeted federated identity providers using tools like AADInternals to federate domains. The actor then used the newly federated identity provider to sign in as a valid user. Similarly, in March 2024, ransomware threat actor Storm-0501 attacked Azure environments using the AADInternals tool to federate attacker-owned domains within compromised tenants, using the newly federated identity provider to sign into additional valid accounts.

Storm-0539, which primarily targets retail organizations for gift card fraud and theft, also has a deep understanding of cloud environments that it exploits to conduct reconnaissance on organizations' gift card issuance processes and employee access. The actor then conducts phishing attacks to hijack cloud accounts. After gaining access to a user's session and token, Storm-0539 registers its own malicious devices so that multifactor authentication (MFA) prompts associated with a compromised victim account go to the threat actor's device. Registering a device lets them wholly compromise an identity and persist in the cloud environment.

Links

Inside the growing risk of gift card fraud | Security Insider | May 2024

Cloud identity compromise

Lifecycle stages for a cloud abuse attack starting clockwise from the top
Source: Microsoft Threat Intelligence

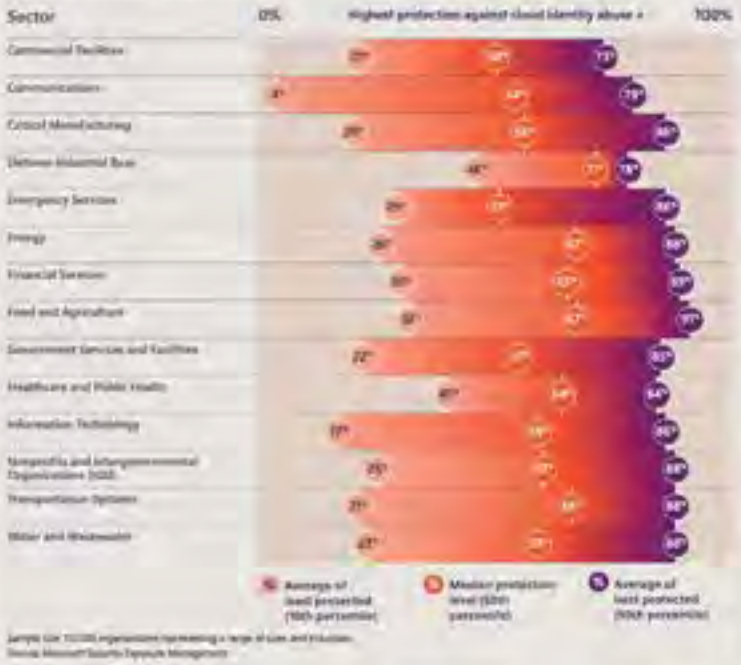
The chart on this page shows average preparedness against cloud identity abuse by sector, in three groupings: best (90th percentile), median (50th percentile), and least (10th percentile) protected. A tenant will have a score of 100% if all the known protections for cloud identity abuse, as mapped by Microsoft researchers, have been applied to their full estate. The communications sector shows a broad range in levels of protection, with the 10th percentile group at just 5%. By contrast, the Defense Industrial Base sector's least prepared group is closer to 50% of protections in place.

Links

Five steps to securing your identity infrastructure | Microsoft Learn | Oct 2022

Security Exposure Management | Microsoft Security, Compliance, and Identity Blog

Cloud identity abuse preparedness



Sample size: 10,000 organizations representing a range of sizes and industries.
Source: Microsoft Security Exposure Management

Actionable insights

1. Centralize your organization's identity management into a single platform.
2. Require all users to enroll in MFA. MFA can inherently stop an attack before it even begins by preventing an attacker who has managed to compromise a user's credentials from accessing network resources. Require phishing-resistant authentication for all developers and all users in administrative roles.
3. Block legacy authentication. Apps using their own legacy methods to authenticate and access company data pose another risk for organizations. The alternative, modern authentication, reduces security risk by supporting MFA and Conditional Access.
4. Understand your cloud environment's "trust chain." With the rise of SaaS applications, guest accounts, and delegated privileges, an organization might fail to properly determine who has access to what within their cloud tenant. Attackers use this ambiguity to identify identities that already have access to resources of interest.
5. Create a custom activity policy to get alerts about suspicious usage patterns.

Cyber Point of View: Canada

How Canada is boosting security by investing in innovation and partnerships

Housed within the Royal Canadian Mounted Police, the National Cybercrime Coordination Centre (NC3) performs a key role when it comes to Canada's global contribution to reduce cybercrime.

The NC3 has not only established operational links with domestic policing but allowed Canada to join international efforts in joint-sequenced operations that attack the enabling pillars of global cybercriminality. These efforts have included Canadian law enforcement working alongside law enforcement from 10 countries as part of "Operation Cronos" to systematically disrupt and discredit LockBit's business model and expose affiliates associated with this notorious ransomware gang.

NC3 has also been active in warning businesses and organizations about impending cyberattacks. It warns 300–400 victims per year and facilitates the deployment of decryption tools where available so victims can regain access to their data or systems. The NC3 also enhances the capabilities of Canada's law enforcement agencies through technical, intelligence and case coordination support.

Microsoft further integrates the Cyber Centre's Aventail feed into its own threat intelligence ecosystem contributing to improved cyber threat detections for customers worldwide. Microsoft's security engineering teams extensively utilize the Cyber Centre's assembly line malware detection and analysis tool within our cloud based cyber defense systems.

DDoS: Stealthier threats emerge

Distributed denial of service (DDoS) attacks are cyberattacks that aim to disrupt or disable a website or online service by overwhelming it with traffic from multiple sources. DDoS attacks can cause significant losses for businesses such as downtime, lost revenue, damaged reputation, and increased costs.

Attack landscape

Beginning in mid-March, we observed a rise in network DDoS attacks, reaching approximately 4,500 attacks per day in June. Additionally, there was a significant surge in attacks targeting medium size applications. Application layer attacks are more stealthy, sophisticated, and difficult to mitigate than network-level attacks.

These attacks, which are in the range of 100,000 to 1 million packets-per-second, are aimed directly at specific web applications, revealing the relentless nature of attackers trying to evade volumetric DDoS protection tactics. Without adequate protection, these applications would experience availability issues.

The increased focus of DDoS attacks on the application layer rather than the more traditional network layers has created a greater risk of impact on business availability, such as access to online banking services or the ability to check-in for airline flights.

A new threat: Application loop attacks

A new type of cyberattack is targeting the very protocols that form the backbone of our internet communication. Dubbed the "loop attack," this vulnerability reveals a critical weakness in application-layer protocols that rely on the User Datagram Protocol (UDP). According to the Helmholtz Center for Information Security (CISPA), these attacks could potentially affect 300,000 application servers worldwide. The loop attack does not discriminate in its choice of targets. Protocols that many consider the lifeblood of the internet—such as TFTP, DNS, and NTP—are at risk, along with legacy protocols like Echo, Chargen, and QOTD. The vulnerability triggers an endless loop of error messages between servers, leading to a severe degradation of service and network quality.

Unlike the more commonly known reflected UDP-based floods, loop attacks may not amplify the traffic volume with each spoofed packet. However, they can still cause significant disruption by ensnaring multiple servers in a never-ending communication loop. This is initiated by a single, well-crafted packet, and once the loop starts, there's no stopping it, and the network flood that ensues can threaten not just the application servers but also the underlying network infrastructure.

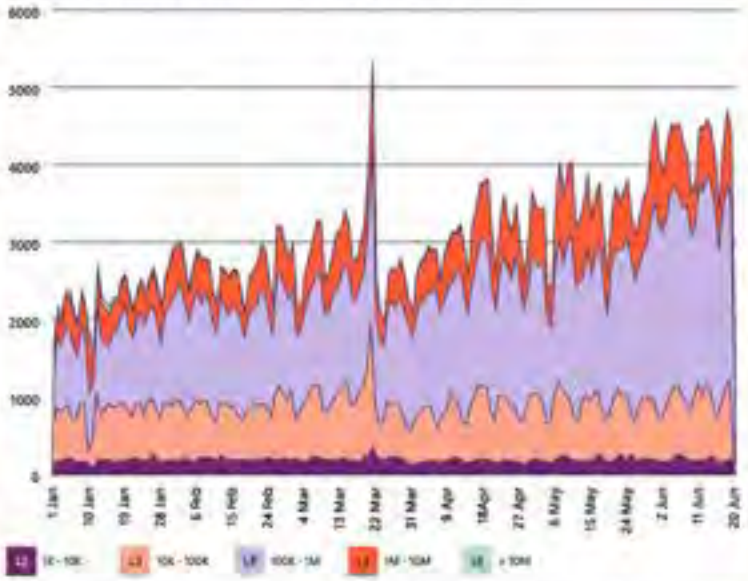
The loop attack is a stark reminder of the vulnerabilities that exist within our network protocols. It highlights the need for continuous vigilance and the development of robust security measures to protect against such sophisticated threats.

Application layer attacks are more stealthy, sophisticated, and difficult to mitigate than network-level attacks.

4X

We mitigated 1.25 million DDoS attacks in the second half of the year, representing a fourfold increase compared to last year.

Number of network DDoS attacks (January-June 2024)



The number of DDoS attacks mitigated continues to increase, with a notable surge layer 4 (L4, application layer) attacks. Application layer attacks are more stealthy, sophisticated, and difficult to mitigate than network-level attacks. Layers on the key are in "packets per second (pps)".

Source: Microsoft Global DDoS Mitigation Campaigns

Actionable insights

1. Where possible, minimize exposure of your applications over the public internet to minimize the attack surface area for DDoS attacks.
2. For applications that are exposed over the public internet, follow a defense-in-depth strategy and ensure the applications have network layer DDoS protection in place. Specific to web applications, it's important to protect them with a web application firewall that provides comprehensive application layer protection.
3. Integrate DDoS simulations in the software development lifecycle and as a regular part of security operations to ensure the applications and workloads have the right level of protection and scale.

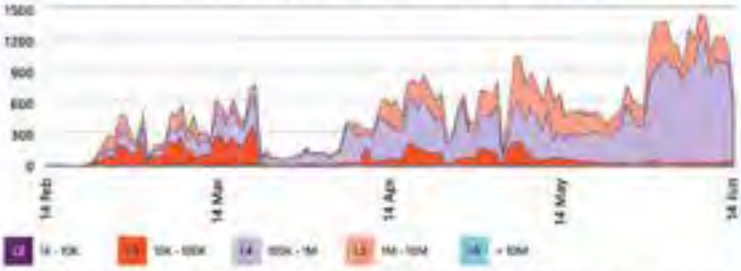
Cyber Point of View: India

DDoS attacks on the rise in India

India was one of the countries most affected by DDoS in 2024, continuing the trend from last year. In 2023, it ranked third in the number of DDoS attacks per customer organization in the APAC region and ninth in the world, with the finance, technology, and government sectors the most targeted.

The number of DDoS attacks per customer in India has more than doubled since 2020.[8] Most of the DDoS attacks in the APAC region from February to June 2024 targeted India, especially the gaming sector. Online gaming is prone to DDoS attacks, and it is a growing sector in India. The mid-size throughput attacks reached ~1,000 attacks per day on India's gaming sector alone, accounting for ~20% of all attacks. The attack volume per customer during that time also increased from 1.4 Gbps to 2.4 Gbps.

Daily number of attacks targeting the APAC region (February-June 2024)



Layer 4 (L4) attacks were the most prevalent DDoS attack type in the APAC region, as well as globally. Layers in the key are in "packets per second (pps)".
Source: Microsoft Global DDoS Mitigation Operations

Separately, DNS query floods are the most common type of application-level DDoS attacks in India. Hacktivists, who use cyberattacks to express their political, social, or ideological views, are a major source of DDoS attacks, and cloud-based resources are increasingly used by both attackers and defenders. Microsoft found a spike in DDoS activity in India in June of 2024. This is unsurprising given there has historically been an increase in cyberattacks during election periods, and India's national elections occurred from April to June.

DDoS attacks in India January–June 2024:

- Maximum number of vectors seen in a single attack: 9.
- Maximum attack throughput: 61 Gbps and 41.2 Mpps.
- Top attack vectors: TCP ACK flood, SSDP amplication, DNS amplification.

Common tactics, techniques, and procedures of DDoS attacks in India

- Botnets to generate and amplify DDoS traffic.
- Living-off-the-land techniques (where malware uses only resources already available in the operating system) to evade detection and mitigation.
- Proxy services to obfuscate the source of DDoS traffic.
- Encryption to bypass security controls.

Actionable insights

1. Implement a DDoS protection solution, securing the network and application infrastructure, hardening the DNS infrastructure, and preparing an incident response plan.
2. Implement security measures such as firewalls, load balancers, and routers to secure the network and application infrastructure.
3. Implement security measures such as DNSSEC and DNS filtering to harden the DNS infrastructure.

Chapter 2

Centering our organizations on security

What is the path forward to improve resilience?

Key developments	54
Introduction	55
Secure Future Initiative	56
Strategic approaches to cybersecurity	57
Supporting the ecosystem	67
Collective action	77

Key developments
Centering our organizations on security

In this chapter we emphasize the responsibility of everyone for keeping their own house in order, emphasizing robust accountability alongside a fundamental mastery of cybersecurity essentials. More than just compliance checklists, we advocate for a threat-informed strategy that enhances resilience across the cyber landscape.

We also extend our focus beyond organizational security to incorporate the broader ecosystem, particularly in critical environments and electoral processes. The chapter concludes with a call for collective action, urging stronger collaborations between industry and government to bolster our collective security.

The Secure Future Initiative (SFI)

Taking proactive steps to keep security deficits from re-accumulating, we share what we are doing, how customers can benefit, and how they can better protect themselves.

Find out more on p55.

Security stories from critical infrastructure frontlines

Helping to support the ecosystem through transparency of datacenter application security findings.

Find out more on p60.

Taking a threat-informed approach to defense

80% of organizations have attack paths that expose critical assets.

Find out more starting on p61.

Best practices for robust cybersecurity governance and accountability

Everyone in the organization, including Board members, must have basic literacy of cybersecurity threats, a sense of personal responsibility for security, and clarity on their role.

Find out more on p63.

Hierarchical pyramid of cybersecurity needs

It starts with the basic need to protect identities, against ransomware, supply chain attacks, and other threats that bypass traditional security measures.

Find out more on p66.

Generative AI is fueling the need for data security policy implementation

The use of generative AI applications can pose serious risk to organizations that haven't implemented sufficient data governance controls. On the other hand, generative AI can be used to kick-start a strategy and approach to understanding their data perimeter.

Find out more on p67.

Collective action through deeper partnerships between industry and governments

Hybrid warfare, cyberattacks, and foreign influence operations pose grave risks to society.

Find out more on p72.

Supporting democratic elections

During this unprecedented period of critical elections worldwide, we are working to safeguard institutions from malicious schemes that aim to disrupt or influence electoral processes.

Find out more on p79.

Introduction: Tackling technical debt and shadow IT for a secure future

Threat actors prey on unaddressed technical debt, outdated security controls, and shadow IT.

If there is a weak point in your system, threat actors are going to find it. You may be using the latest security tools to fortify your core environment, but if you still have old infrastructure, unpatched systems, outdated configurations, and apps granted too many permissions by departments you aren't even aware of, then you may be unwittingly leaving security holes for threat actors to exploit.

Leaving these issues unaddressed is like installing a vault with an impenetrable lock, then forgetting the vent that leads to the roof. Or, forgetting about the drain to the sewer or the unfortified wall adjacent to the parking lot. The burglar won't be discouraged by the lock—they'll just find one of the alternate pathways you have left for them.

When it comes to digital security, it doesn't matter how locked-down your user policies are: if an adversary can gain entry via a long-forgotten avenue, they will.

Whether it is a test app from an untracked satellite tenant that doesn't enforce multifactor authentication (MFA), devices infected with malware, or legacy authentication protocols, security teams can't act on resources they simply don't know about. These include:

- Unsanctioned, unmonitored or abandoned tenants built ad hoc for development, testing, or demos.
- Applications and workload identities with no known owner or governance.
- Developer secrets checked into public code repositories.
- Storage repositories with inadequate access controls.

Clearing out technical debt

As part of our Secure Future Initiative (SFI), Microsoft embarked on rigorous "spring cleaning" to strengthen our environment and cloud services against threats. We removed millions of unused and non-compliant applications and tenants from our environment, refreshed hundreds of thousands of credentials (including security certificates), and segmented and isolated our network.

SFI is a multiyear initiative to evolve the way we design, build, test, and operate our products and services so we can achieve the highest possible standards for security.

We are taking proactive steps to keep security deficits from re-accumulating:

- Maintaining a comprehensive inventory of all production software and hardware assets.
- Enforcing a standard approach for creating secure test tenants with zero trust principles, automatically deleting them after use to avoid legacy infrastructure buildup.
- Increasing isolation of development and test environments to prevent lateral movement into production.
- Enforcing the use of standard libraries and advanced code security checks for all apps and services.
- Automatically scanning all internal productivity systems to remove passwords, secrets, and keys that attackers could exploit.
- Improving logging capabilities to detect, investigate, and mitigate vulnerabilities faster, and share insights with customers sooner.

In all cases, we're creating "paved paths" for engineers, so that the easiest way to do something is the also right way. We continuously apply lessons from security incidents to improve our methods. In response to rising phishing and social engineering attacks, for example, we're issuing phishing-resistant credentials like passkeys to all employees. We also introduced video-based user verification for lost credentials and automated processes for deploying security keys and storing secrets. Our platforms operate at the highest industry standards, and we're building systems to maintain these levels as standards evolve.

The Secure Future Initiative is not a destination, but an ongoing commitment to a security-first culture that proactively identifies and openly discusses risks, issues, and blockers; quickly learns and iterates; and standardizes tools, dashboards, practices, and principles across all engineering teams. As with every feature and experience we ship, we'll share with customers what we do, how they benefit, and how they can better protect themselves.

Joy Chik
President, Identity and Network Access

Putting security above all else

The Microsoft Secure Future Initiative (SFI) is a multiyear initiative to evolve the way we design, build, test, and operate our products and services, to achieve the highest possible standards for security.

It's our long-term commitment to protect both the company and our customers in the ever-evolving threat landscape.

730k

SFI non-compliant apps eliminated

5.75 million

Inactive tenants eliminated, drastically reducing the potential cyberattack surface.

Links

Secure Future Initiative | Microsoft

"If you're faced with the tradeoff between security and another priority, your answer is clear: Do security. In some cases, this will mean prioritizing security above other things we do, such as releasing new features or providing ongoing support for legacy systems."

Satya Nadella
Microsoft CEO,
May 3, 2024

Strategic approaches to cybersecurity: "Managing your own house"

Data security

Accountability is increasingly central to the world of data security. From security strategies to new policies governing generative AI, organizations must start taking responsibility for what is going on under their own digital roof.

Key components of an effective data security strategy

In our experience, the most successful data security implementation strategies consider the following: visibility, risk detection, classification, labeling, data protection, and data leakage prevention across your multi-cloud and hybrid digital estate.

It is no longer enough to focus solely on the data; it's just as important to understand how that data moves within the organization, how users, customers or partners interact with it, and what level of risk is acceptable for the organization.

Data doesn't move on its own. It's moved by people. Because different people require different levels of access, a comprehensive data security policy must be dynamic, considering both data and user context. This lets organizations balance protection and productivity, allowing low-risk users to continue working as usual while restricting the actions of users with elevated risk.

As data types proliferate, sources get more complex, and generative AI technology gains traction, data security is inevitably becoming a pressing concern. A 2023 Microsoft study found that over 40% of enterprise (>500 employees) organizations' annual cybersecurity budget on average is now allocated to data security.

Securing organizational data has also become a multifaceted task, leading to the adoption of multiple, hard-to-manage tools. This kind of fragmented approach creates more noise from duplicated alerts, making it harder to identify and investigate actual incidents. Organizations using over 15 tools experienced nearly three times more data security incidents than organizations using fewer tools. This is why it is so important to invest in integrated, automated data security solutions to achieve the best outcomes.

How generative AI is fueling the need for data security policy implementation

Microsoft's AI products, such as Copilot, are designed to use only information you already have access to. When other generative AI apps are deployed on ungoverned data estates it can result in data oversharing or leakage as users may end up accessing sensitive data. It is difficult to protect data from AI-related security risks given many organizations don't actually know where—or even what—their sensitive data is.

Studies show 83% of organizations experience multiple data breaches over time, so getting ahead of the risks is critical. Data environments must be prepared for AI, which requires inventorying data stores, identifying sensitive data, then labeling and protecting it to ground the data and prevent its unintended exposure to AI apps.

Applying data loss prevention policies for inputs and outputs from AI apps helps to prevent both overexposure and leakage for new AI generated data, while automating data classification and labeling vastly reduces the risk of data exposure. In summary, data loss prevention policies can apply to data that AI models consume and generate.

An integrated approach to data security

- Classify and label sensitive data, and prevent its unauthorized use across apps, services, and devices.
- Understand the user intent and context around the use of sensitive data to identify the most critical risks.
- Assign high-risk users to appropriate DLP, data lifecycle, and Conditional Access policies.

Links

Microsoft Insights and best practices for securing data | Microsoft Security Blog | Oct 2023

Empowering employee self-service with guardrails: How we're using sensitivity labeling | Apr 2024

How to use prompts in Microsoft Copilot for Security | Microsoft Security Blog | Feb 2024

Microsoft Copilot for Security in Microsoft Purview | Microsoft Learn | Sep 2024

GitHub - Azure/Copilot-For-Security

Data security (continued)

Harnessing generative AI to define your data perimeter

AI can be a powerful tool for data exploration, and customers and partners worldwide are increasingly using it to improve their data management practices. For example, data security teams are using AI to refine their data loss prevention policies, classification labeling practices, and encryption usage.

While innovative AI applications like Copilot offer exciting possibilities, it is crucial for organizations to first understand their data perimeters. By doing so, they can implement effective governance controls and data loss prevention policies to prevent overexposure and loss. Proactive measures must also be taken to safeguard infrastructure, devices, and containers against data-targeted attacks.

> The rise of generative AI applications poses a serious threat to organizations that haven't implemented sufficient data governance controls.

To expedite this process in the realm of cloud computing, generative AI applications can be used to suggest improvements to data loss prevention policies. Our latest findings indicate that these applications can provide a quick and strategic approach, particularly for engaging and convincing policy users and creators who may be initially resistant. Moreover, when there is a need to discover the data estate, the computational capacity of generative AI can handle vast amounts of data, allowing for efficient governance.

Compliance managers have been able to discern the sensitive, unprotected data from other data, and gain valuable time to approach the next steps. An example of an interaction with a generative AI assistant would be: "Show me the unprotected documents with sensitive information types in this SharePoint site I can access" and the AI assistant would generate a successful eDiscovery query.

Readiness levels: Protecting and governing data while benefitting from generative AI

1 Prepare data

Prepare your data for generative AI. Focus on labeling data, implementing controls, and educating users about data protection.

2 Limited implementation

Limited implementation of generative AI. Restrict access to sites that may contain sensitive files. Leverage tools that provide visibility into how users are using AI, which can inform stronger protection controls.

3 Used to enhance productivity

Generative AI used to enhance productivity. Optimize data governance and loss prevention, use advanced capabilities for risk management and compliance.

4 Driving force for innovation

Generative AI used as a driving force for innovation. Expand adoption throughout the organization, continuously financial user behaviors and accountability, and extend data governance to cover all environments.

Links

Whitepaper: Prepare your data for secure AI adoption | Microsoft Security Blog | Jul 2024

Data security and compliance protections for Microsoft Copilot | Microsoft Learn | Aug 2024

Zero Trust principles for Microsoft Copilot for Microsoft 365 | Microsoft Learn | Apr 2024

Cyber Point of View: Sweden

Using the cloud to protect against ransom attacks

In November 2023 the Church of Sweden, a religious institution with over 5.4 million members and 3,400 churches, was targeted by Russian ransomware-as-a-service operator Blackcat.

Despite detecting the threat and acting quickly, data was exfiltrated and a significant number of systems encrypted before the threat was isolated. It took over two months to recover, impacting the Church's ability to raise funds during the critical Christmas period and to perform some services, such as funerals.

Since the Church's cloud services remained in-service, the Church was able to maintain its internal and external communication through M365, which was a key factor for maintaining effective crisis communication throughout the crisis.

As has become common in recent years, not only were the Church's systems encrypted and data exfiltrated, but the Church then faced an extortion threat if it did not pay to prevent that data from being published. Ultimately, a second threat actor, Lockbit, published 2.3 million files after the Church refused to pay.

The key learnings of this event were:

- It is crucial to have advanced detection capabilities for identifying and mitigating data exfiltration with 24/7 active monitoring. Without this, the impact would have been much greater.
- The time window for patching critical vulnerabilities has narrowed, from 14–30 days five years ago to a mere 24–72 hours today. This is in part because software vulnerabilities have become more prevalent as initial access vectors.
- It is important to have ongoing business continuity planning that includes cyber threats in order to minimize the disruption and inconvenience caused by such attacks.

Hierarchy of cybersecurity needs

Drawing inspiration from Maslow's hierarchy of needs, this graphic illustrates a prioritization of cybersecurity, starting with the most basic need: protecting identities. AI has a role at each tier, underscoring its potential to enhance security measures. Cultivating a robust security culture within the organization, helps ensure the technological defenses and human practices evolve in concert to mitigate threats effectively.

AUTOMATE SECURITY OPERATIONS
Automating security operations is the holistic approach to building on perspectives and insights across all layers in the pyramid.

IMPACT...
Automating processes at scale creates new opportunities for insights as well as relief for stressed humans.

DETECT AND REMEDIATE THREATS
Monitoring your ecosystem to identify anomalous activity and contain threats.

IMPACT...
The ability to identify and respond quickly can limit lateral movement, contain damage to assets and deny persistence.

SECURE DIGITAL ASSETS
Digital assets, whether code, traditional data stores, and now generative AI models are all key components of modern workloads.

IMPACT...
Modern workloads deliver the value add to end users who increasingly rely on their integrity and availability.

PROTECT ENDPOINTS
Protected endpoints include the multiple dimensions of devices in use today – from PCs and mobile devices, to network and operational technology (OT), and servers in datacenters.

IMPACT...
Effective endpoint protection can limit the repercussions of unauthorized access.

PROTECT IDENTITIES
"Attackers don't break in, they log in." Credentials for both individuals and machines are the perimeter of the modern attack surface.

IMPACT...
Strong identity security can greatly reduce risk exposure—particularly for ransomware attacks.

Threat-informed defense

Thinking differently to address threats

Most organizations rate bugs according to severity and how difficult they are to mitigate before assigning a team to fix them within a set compliance window. However, what happens is clashing prioritizations and silos with no knowledge of an adversary's attack path. Hence the saying: "Defenders think in lists and attackers think in graphs".

Organizations have complex operating environments that require defenders to see across various vendors in order to discover attack paths. Instead, they should look to understand their critical assets and crucially how they are, or could be, connected. The resulting view of an organization's posture is key to understanding the risk exposure to cyber threats. By adopting an attacker's perspective, the prioritization of mitigation efforts is enhanced.

Pre-breach attack path analysis

Traditionally, organizations have leaned on all sorts of different security tools to manage threat exposure across their estate. This messy patchwork of approaches however, can lead to exposure visibility gaps and efficiency challenges.

This makes it imperative for security leaders to reach a unified and comprehensive view of their estate and to both continuously and smartly prioritize exposure reduction efforts. Prioritization should seek to understand threats and attacker perspective, identify "crown jewels" of interest to the attacker, and both identify and mitigate any paths that lead to them.

Three key components are required for threat-informed defense: single pane of glass, critical asset protection, and attack path management.

Single pane of glass

Organizations should consolidate threat exposure insights across their estate into a single view covering cloud assets, on-prem devices, data, identities, applications, network, and the Internet of Things (IOT). This should then be used to manage top threats such as ransomware and business email compromise, as well as exposure to threat campaigns and actors.

80%

of organizations have attack paths that expose critical assets

Critical asset management

It is imperative to thoroughly map an estate's "crown jewels." This can include critical servers, highly privileged identities, sensitive data, or other assets. Microsoft data indicates that an average <1% of organizational assets are of high interest to attackers.

Attack path management

Organizations should identify the most likely attack paths leading to critical assets and continuously mitigate them. An attack path calculation incorporates things such as asset inventories, vulnerability/weakness data, and external attack surfaces to construct a possible attack chain leading to a critical asset.

Links

Introducing Security Exposure Management - Microsoft Community Hub | Mar 2024

Identifying and Protecting the Crown Jewels of your Cloud | Aug 2024

Exposure insights and secure score in Microsoft Security Exposure Management | Aug 2024

One graph of everything - Microsoft Security Exposure Management Graph | May 2024

Optimizing governance and accountability

When viewed as a business risk, cybersecurity is everyone's responsibility. As senior managers wrestle with how to set up their organization's governance, the need for a more responsible approach to operating models quickly emerges, particularly when it comes to defining security responsibilities for roles that sit outside the scope of the security team. These tasks are further complicated by how dynamic and rapidly changing the threats that cyber challenges post to technology platforms and transforming business models.

To manage this, organizations should adopt comprehensive, clear, and adaptable operating models. Their culture and governance structure must make it clear that security is everyone's role, providing clear guidelines, and building in flexibility to accommodate changes in the threat landscape. Cybersecurity incidents are like forest fires: they can start anywhere and spread anywhere within minutes. Organizations should focus their security culture and governance efforts on accountability, teamwork, and shared responsibility.

Accountability always starts at the top, with organizational leaders who not only understand their responsibility for security outcomes but ensure that security risk management is embedded across their business in an organization-wide and collaborative way.

Leaders must establish a system of accountability, prioritization, and aligned incentives that is executed and monitored across the organization. They must delegate risk accountability, mitigation implementation responsibility, and associated budgets/costs to leaders, managers, and individual contributors, as accountability alone cannot create a healthy culture. Unfunded mandates or the belief that "security is the security team's job" will result in avoidable security weaknesses, increased burnout of security resources, and, in time, greater organizational cost.

Leaders must support mechanisms that incorporate security into business unit KPIs/scorecards, inclusion of security in enterprise business discussions among executives and boards of directors, and security education for all roles. They must also evaluate the systems of financial incentives in place, including those at the senior level to ensure good security behavior is prioritized and rewarded. They must also publicly promote the idea that security is important, and demonstrate that everyone is expected to collaborate to solve security problems.

Establishing cross-team processes and goals is crucial, as many people have never even worked with other teams on security issues and very often don't know who to work with in the first place. Shared goals must be reflected in performance metrics for executives, teams, and individuals throughout the organization.

Key elements should include:

Avoiding blame.

Unless there is a clear violation of professionalism or failure of due diligence, avoid assigning blame for security incidents. Blame invariably increases risk by poisoning a culture with fear and undermining the collaboration required for an effective coordinated defense.

Making sure learnings or issues don't slip through the cracks.

Security issues are often discovered through critical thinking, fresh perspectives, and unexpected sources. Establish ways to capture inputs and feedback regardless of where they are found ("See something, say something" adage).

Sharing responsibility.

Organizational leaders must normalize security as part of everyone's job. Everyone in the organization, including board members, must have basic literacy of cybersecurity threats, a sense of personal responsibility for security, and clarity on their role in security.

Requiring cross-team training and learning.

Fostering understanding, empathy, and repeatable cross-team processes enables teams to work well together on security. This typically requires ongoing focus, reinforcement, and practice to overcome past habits.

Tips to build security literacy:

Make it personal and human.

Build training tailored to your organization and roles (so people can apply it easily and quickly). Relate stories about how attacks could happen in real life, and teach people to use safe cybersecurity practices at home and at work.

Make it clear.

Always ensure people understand why it's important, what they need to do, and how to do it through policy and education.

Make it engaging and fun.

Use gamification, interactive hands-on components, positive reinforcement, and/or public recognition to keep people engaged.

Make it easy.

Most people prefer to take the easy path in all things, so ensure the security behaviors you want to drive are simple and straightforward.

> Organizations should focus their security culture and governance efforts on accountability, teamwork, and shared responsibility.

Security incident decisions: Dispatches from the field

Security incident decisions

Preparation | Communication | Execution

Microsoft's Incident Response (IR) team are the First Responders of the cybersecurity space.

Similar to how police, fire or, paramedics are called to the scene of an accident, the IR team must quickly assess the situation, devise a plan, and take immediate action. And just like how society has come to depend on the lessons First Responders have gleaned over decades of service, the real-world experience of our IR teams can be used as a template for organizations to better prepare for cyber incidents.

In particular, we highlight three categories: preparation, communication, and execution.

Ever wondered why First Responders are able to control an emergency so quickly and confidently?

1. Preparation
2. Established playbooks

Too often, IR teams find that their customers don't have a reporting plan in place. This means it takes precious time to understand the needs of each individual stakeholder and establish the necessary line of reporting.

Preparation

It is crucial to have a well-prepared response plan in place before an incident, as scrambling for key information during an incident can be chaotic. This means identifying key decision makers, business-critical applications and services, roles and responsibilities, and response and recovery processes well in advance. Not having this information at hand leads to longer recovery times and higher impacts on the business.

Communication

Without accurate information and established communication lines, key data may be lost or poorly relayed during a security incident. A company should therefore tailor its communication based on its audience, for example: company executives, regulatory bodies, employees, and the public.

Each group requires different levels of technicality. Executives need brief, high-level summaries that focus on the impact on business operations and steps being taken to resolve the issue. Regulatory bodies require detailed reports that comply with legal and industry-specific regulations. Communication with the public is equally important and should provide transparent and reassuring messages that protect customers' trust and address their concerns without revealing sensitive details.

A similar approach should be taken for internal communication. Establishing a single source of truth for employees—an internal communications channel or newsletter—is important for reducing misinformation. In situations where established communication channels could be compromised, having an alternative communication channel is crucial.

Communication does not end once the incident is resolved. Ongoing updates on the progress, root cause analysis, and preventative measures are vital. Providing regular post-incident communication that includes lessons learned and actions taken to prevent future incidents demonstrates transparency, builds trust, and shows the organization's commitment to security and continuous improvement.

Execution

In terms of technical preparation, execution encompasses all aspects for which a company can plan.

Established playbooks not only consist of procedural plans to contain, recover, or remediate risks, but also include actionable steps to address these tasks. Common examples include: containment and recovery of identity systems which may require a mass password reset.

Other containment actions need to be taken into account. What measures, technologies, tools, or practices should be followed in the event of a compromise on multiple client systems? Although a company may have excellent security tools in place, their effectiveness is diminished if the team responsible for managing them lacks proper training.

Additionally, recovery actions (such as recovering data rendered inaccessible by the threat actor) can be beneficial. Organizations do not always have the chance to practice these actions and this lack of familiarity with recovery practices can significantly impede attempts to restore the environment in a timely manner during incidents. A mature playbook process can be compared to fire drills at a work site. Many companies regularly test their preparedness for a fire emergency, but only a limited number do the same with cybersecurity. This reinforces the importance of preparation, including creating tabletop exercises and conducting drills to validate their effectiveness.

The following are the most common challenges we encountered during IR engagements:

Reporting lines are not clearly defined

- Reporting lines are needed to make the right decisions and keep everyone informed of current ongoing tasks, investigation and recovery progress, and business impact.

Roles and responsibilities are not clearly defined

- Having unclear definitions of roles and responsibilities can hinder the effective and timely response to security incidents. The speed of recovery necessitates prompt decision-making.
- Business decision makers have a crucial role in defining business-critical services, making investment decisions, and setting company strategy.
- Technical decision makers and operations personnel are also crucial during incidents as they possess deep knowledge of dependencies.
- To navigate the legal aspects and maintain compliance, a legal team (internal and/or external) is critical especially when incidents involve sensitive data, such as personally identifiable information (PII).

Lack of preparedness simulation exercises

- To prepare for incidents, nothing is more effective than conducting mock or tabletop exercises.
- These exercises equip individuals with the skills and knowledge to handle real incidents and provide valuable insights about areas in need of improvement.

Links

Creating a proactive incident response path | Microsoft Security Blog | June 2024

Cyber Point of View: Latin America

Tough lessons for board members about cybersecurity

Microsoft recently launched an initiative that brought together unions like the Instituto de Directores de Chile, iDirectores, icare, and Women Corporate Directors and board members from over 150 companies in Mexico, Chile, Colombia, and Peru.

The initiative simulated a cyberattack on a retail company during the peak of Cyber Monday, to serve as both a crisis management test and an opportunity for learning and collaboration. It was executed without alerting the participants beforehand and began with a phishing email sent to an employee, and the board members' responses provided valuable insights into organizational preparedness and areas for improvement.

Following the simulation, Microsoft facilitated dialogue and knowledge exchange among the participants, underscoring the power of collaboration in addressing cybersecurity challenges. This exercise showed how experiential learning and collective engagement can enhance board members' understanding of cybersecurity and strengthen an organization's resilience.

Resilience maturity

Within the dynamic realm of cybersecurity, the IR team regularly confronts a wide spectrum of customer challenges. Drawing from this rich experience, we have found that an organization's resilience maturity can be determined based on four pillars: Operational, Tactical, Readiness, and Strategic. Maturity in each of these pillars is categorized as either Basic, Moderate, or Advanced.

Operational

For day-to-day IT operations, good preparation and maturity can ensure that an organization has good visibility of its estate, documented reliable playbooks, and rapid response capabilities based on automation.

- Deploy an endpoint detection and response solution on all desktops and servers, with a dedicated security operations team whose primary role is monitoring and actioning alerts.
- Automation within an existing SIEM / SOAR solution.
- Test, tune, and actively manage custom playbooks and adjust them to specific needs.
- Implement a multi-tier security operations center (SOC), where common alerts are automatically triaged. Establish a feedback loop to improve playbooks and adjust environment hardening.

Tactical

Prepare for initial response to an incident to respond logically and efficiently.

- Maintain detailed and practiced IR plans with clear actions to be taken in the event of an incident.
- Provide IR and threat-hunting teams with a clearly defined scope for proactive security hardening duties. Enforce strong phishing resistant MFA for all user accounts.
- Establish a ready, out-of-band communication channel in case there is a severe infrastructure compromise. This helps to ensure timely and secure communication with dedicated update meetings to keep all participants informed.
- Implement firewall and endpoint containment capabilities. Set up mass password reset capabilities and automatic attack disruption mechanisms.

Readiness

Prepare for a cybersecurity incident.

- Require employees to complete training and demonstrate understanding of material before granting continued access to company resources and data.
- Conduct continuous access reviews for company resources and data.
- Implement service level agreements for recovery time and recovery point objectives.
- Maintain up-to-date infrastructure diagrams and documentation across entire environment.
- Compare changes against an existing desired state before implementing them. Regularly update documentation to reflect these changes. Maintain detailed asset management, including location, owner, and automatic device discovery, and with device compliance.
- Enforce strict compliance policies to ensure only compliant devices have access to company resources.
- Conduct tabletop exercises on a regular basis. Track and implement feedback and conclusions.

Strategic

Take steps to improve overall security posture in the longer term.

- Actively managed software and technology, with planned migration and modernization projects to keep infrastructure up to date. Proactively implement new technologies.
- Conduct proactive and automatic vulnerability scans on a scheduled basis, for impact analysis on newly published vulnerabilities. Track and follow up on mitigations. Ensure scheduled maintenance windows.
- Clearly define access and authorization strategy to implement zero trust principles. Clearly define and enforce just-in-time (JIT) and just-enough administration (JEA).
- Use passwordless authentication for all identities, privileged or otherwise.
- Implement zero trust strategy with a clear desired future state, including continuous evaluation, improvement and defined timelines.

Supporting the ecosystem

The passkey journey: a story of collaboration across the industry

Passkeys perform a simple function. They offer users faster, easier, and most critically, more secure sign-ins to websites and apps across their devices than password-based methods.[iii]

Instead of vulnerable secrets or potentially identifiable personal information, a passkey uses a private key stored safely on the user's device. It only works on the website or app for which the user created it, and if that same user unlocks it with their biometrics or PIN. This means passkey users can't be tricked into signing in to a malicious look-alike website, and are unusable unless the user is present and consenting. These are some qualities that make passkeys a "phishing-resistant" form of authentication.

Passkeys are not only more secure than passwords, but also are easier to use and manage. Signing in requires a simple unlock gesture: looking into the device camera, pressing a finger to a fingerprint reader, or entering a PIN. Neither biometric information nor the local PIN ever leaves the device and are never shared with a site or service. Some passkeys sync between devices, meaning users can recover them if they lose or upgrade their device. Others are bound to the device. And last but not least, passkeys are much more convenient for users as people no longer have to worry about creating, remembering, resetting, or losing passwords. Passkeys can be stored in a variety of industry solutions including Windows Hello, hardware security keys, mobile devices, and third-party credential managers.

Industry-wide efforts to eliminate passwords in favor of phishing-resistant authentication are gaining traction. Passkeys represent the most significant collaborative effort thus far. Adoption has accelerated after operating system providers and password managers made it easy to issue passkeys and bind them to hardware. Members of the FIDO Alliance[iv] and the World Wide Web Consortium (W3C) worked together on the standards. They include vendors who create browsers, operating systems, and hardware security keys, as well as banks, hardware platform providers, major retailers, and government bodies. All major operating systems, browsers, and mobile platforms now support passkeys.

Instead of vulnerable secrets or potentially identifiable personal information, a passkey uses a private key stored safely on the user's device.

Platform support for creating and managing passkeys is the first step towards mass adoption. According to the FIDO Alliance, more than 140 major websites had added support for passkey sign-in as of June 30, 2024, including Amazon, Best Buy, CVS Specialty, eBay, Home Depot, Instacart, Lowe's, PayPal, PlayStation, Shopify, Sirius XM, Stripe, Target, Uber, WhatsApp, X, and Yahoo, plus services from Apple, Google, and Microsoft. Public sector support for the FIDO2 standard is gaining momentum and national agencies in at least six countries as well as some US state and local governments are now using FIDO2 technology.

As industry support for passkeys grows, general awareness will increase as a natural consequence. Makers of operating systems, platforms, and credential managers, as well as relying parties such as providers of consumer-facing websites, are working with the standards bodies to make the passkey experience even easier and more secure. In the meantime, the message for anyone concerned about cyber security is clear: passkeys are better than passwords and most forms of legacy MFA.

Actionable insights

1. Consumers: look for the passkey logo to identify websites or services that support passkeys. Create and use passkeys wherever possible.
2. Security professionals: give vendors feedback to help shape the future of passkeys. Ask whether they support passkeys and explore whether their implementation supports your use cases.
3. Software developers: visit the FIDO Alliance website for resources on how to add passkey support to your website, app, or service.

Links

Public preview: Expanding passkey support in Microsoft Entra ID | Microsoft Community Hub | May 2024

Cyber Point of View: France

Enhancing France's cybersecurity workforce

France is grappling with a deficit of 60,000 cybersecurity experts, a part of the broader European shortfall of 300,000 professionals. In the face of this acute shortage, the local Microsoft cybersecurity team developed the "Cybersecurity. My Future Job!" pedagogical kit.

The kit is part of a wider cybersecurity skills training plan launched in 2022 and was created using existing local content from ANSSI, CNIL, and Cybermalveillance.gouv.fr, all of which are French organizations dedicated to cybersecurity and privacy.

It is designed to help young people (15 to 21 years old) better understand cybersecurity issues and discover associated professions. It features independent modules that can be completed in a one-hour session or extended to a half-day workshop and two guides: one for participants and one for facilitators.

The kit is available in French and English and can easily be translated into different languages and tailored to specific countries. Its flexible and adaptive nature, coupled with its open-source availability on GitHub, allows for continuous updates and customizations. This makes an excellent resource to attract interest in the cybersecurity field. Since its pilot in May 2022, the kit has been used not only with students, but also in corporate training settings as well.

Links

Cyber Kit | Jun 2024

Microsoft lance son Plan Compétences Cybersécurité pour former 10 000 nouveaux professionnels en France en 3 ans | May 2022

Critical environments

Our previous Microsoft Digital Defense Reports have shown that while IT hardware and software security has strengthened, the security of IoT and operational technology (OT) devices has not kept pace. In this section we offer security stories from the critical infrastructure frontline.

At Microsoft, we manage a large and growing estate of OT devices essential to the operations of our global datacenters. The following section of the report details our experiences managing this infrastructure in two important aspects: building and operating a program for reviewing the security of third-party OT applications, and managing the unique process of updating fleets of OT devices to address security vulnerabilities.

Security stories from the frontline of OT

Threat actors are now exploiting OT devices to do everything from accessing critical and operational networks, to enabling lateral movement, establishing a foothold in a supply chain, or disrupting the target's OT operations.

A three-step action plan: insights from testing OT applications

Our global datacenters rely heavily on OT equipment, such as sensors and actuators. These devices are pivotal for managing critical processes including power management and cooling systems within our infrastructure. Recognizing the need to secure these foundational technologies, Microsoft has spearheaded a crucial initiative over the past four years aimed at fortifying the security of third-party OT applications which are integral to not only our operations but also the wider ecosystem.

Our targeted security program reviewed these third-party OT applications to identify and address potential vulnerabilities, to help ensure their robustness and reliability. The initiative not only bolstered the security posture of our datacenters but significantly contributed to enhancing security standards across the OT industry.

We shared the findings from our reviews with the respective vendors of the products evaluated, creating a collaborative environment for knowledge sharing and mutual improvement in OT application security. Raising the security bar in this way, the products are being made more secure for the whole industry. For instance, this collaborative approach has led to significant security enhancements in products such as power monitoring systems, which now integrate more securely with Windows Server Active Directory, marking a substantial improvement from non-directory based accounts.

Based on this work we've identified three core actions that, if taken by the operations technology industry, would significantly improve the security of systems across the industry:

1. Adopt modern authentication for users and devices.
2. Enable centralized device configuration management and secure apps and devices by default.
3. Implement a Secure Development Lifecycle (SDLC) program for product development that is certified by independent security experts.

Types of OT systems in datacenters

The OT infrastructure systems in datacenters are critical for maintaining operational integrity and safety, focusing on ensuring optimal environmental conditions and monitoring essential operational parameters:

- **Industrial control systems (ICS):** Also referred to sometimes as OT, these systems monitor hardware to ensure everything runs at optimal levels. They include sensors and devices for managing power and environmental conditions within datacenters.
- **Building automation systems:** These systems are focused on cooling systems, HVAC, water chillers, and other mechanisms for producing cold air. They are considered active systems with moving parts like fans, water chillers, and pumps.
- **Electrical power monitoring systems:** These systems monitor electrical power aspects such as frequency, voltage, and wattage. They are passive systems connected to power meters and circuits, focusing on monitoring and situational awareness of the health of all electrical systems in orchestration of power flowing to customer facing servers.
- **Battery monitoring systems:** These continuously assess the health and performance of batteries under different load conditions to ensure datacenter availability by preventing battery backup failures.

Building on this, the goal of any OT application security review program is to:

- **Identify and mitigate security vulnerabilities.** Identify security vulnerabilities within third-party OT applications, which are critical for the operation of datacenters.
- **Ensure operational integrity.** Review and secure OT applications that manage critical infrastructure, such as the above-mentioned systems. This is vital for maintaining the availability and reliability of services.
- **Offer compliance and risk management.** Conducting security reviews helps in compliance with internal and external security standards and regulations. It also plays a significant role in risk management by proactively identifying and addressing security risks.

Inherent risks of vulnerabilities in OT equipment

- Health and safety: The exploitation of vulnerabilities in OT software can lead to significant health and safety risks. For example, if the cooling systems in datacenters are compromised, it could lead to overheating, posing a risk to both the equipment and individuals within the facility.
- Service disruption: Vulnerabilities can lead to disruptions in datacenter operations, affecting the availability of services to customers.
- Data breach and loss: Security weaknesses could enable unauthorized access, leading to data breaches or loss of sensitive information.
- Reputational damage: Incidents resulting from unaddressed vulnerabilities can damage the service provider's reputation, affecting customer trust and business continuity.
- Compliance violations: Failure to secure OT equipment can result in violations of regulatory requirements, leading to legal and financial consequences.

Emerging challenges and trends

Looking ahead, we are seeing a number of trends that will increasingly impact OT security:

1. ICS/OT solution providers, like all solution providers, aim to integrate and upgrade their existing solutions with modern cloud and AI/ML-based solutions for industrial control processes. While these advancements are exciting, they also challenge the effectiveness of existing security controls and network isolation techniques for critical processes.
2. Wireless networking is now prevalent in consumer and business technology products, and is increasingly appearing in OT products as well. These wireless capabilities must evolve to meet the needs of industrial control environments before they can be further adopted.
3. ICS/OT attack frameworks and toolkits that support OT devices and protocols used for critical industrial processes are being developed and used by malicious actors.
4. Automated and AI enabled attack techniques create a sophisticated global attacker workforce that never sleeps and is always looking for vulnerabilities in security defenses.
5. Securing ICS/OT systems is challenging because change is purposefully avoided to ensure the process always works and can have decades-long lifecycles in production. These systems risk becoming collateral damage even when not directly targeted by attackers.

As of July 2024, we had identified and shared over 300 vulnerabilities in third-party OT applications. The initiative contributed to significant improvements in security across the OT industry.

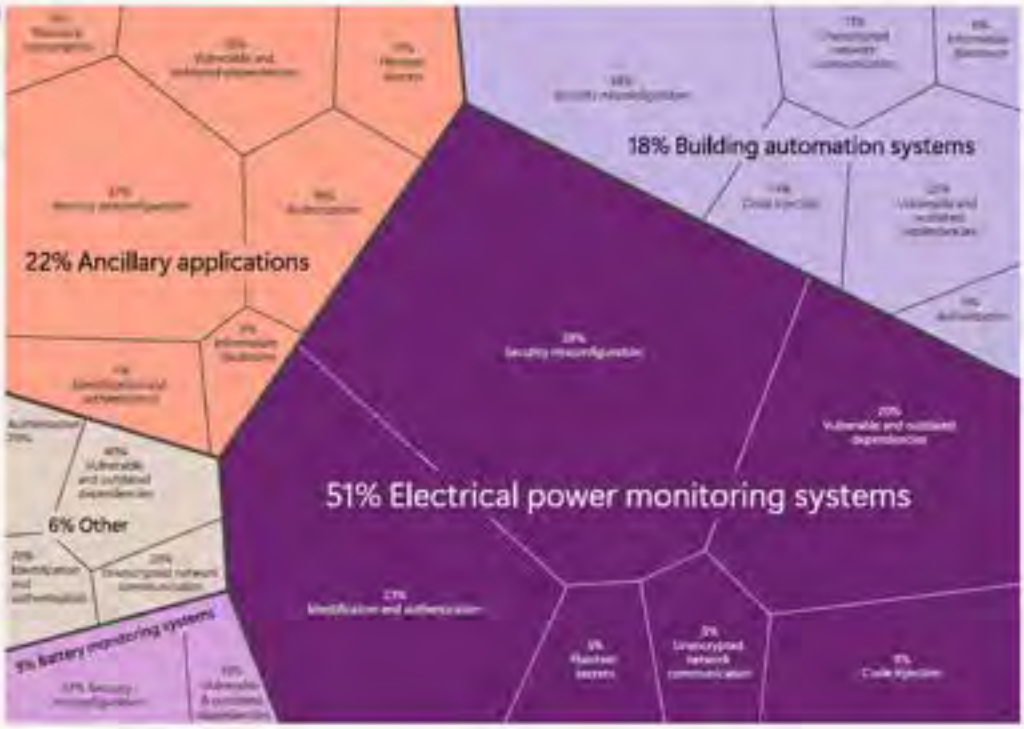
Categorizing the vulnerabilities

As of July 2024, we had identified and disclosed over 300 vulnerabilities to suppliers through our OT application review initiative. This work offers a unique perspective on the types and risks of security weaknesses in OT equipment.

The most common security vulnerabilities we identified, prioritized by risk and impact, are:

- **Outdated authentication:** Vendors should adopt modern authentication methods, such as integrating Windows Server Active Directory for identity management, to enhance security through Kerberos security groups, password complexity, and rotation policies.
- **Insecure configurations:** OT applications should be secure by default and not contain built-in passwords or accounts that could pose a security risk. They should also comply with secure communication protocols like current versions of TLS.
- **Outdated legacy libraries:** Vendors should update legacy software libraries, which are often outdated and contain numerous vulnerabilities. Updating, however, is a significant challenge due to the resources required for updates and the potential impact on application functionality.

Critical and high vulnerabilities discovered in the critical environment disciplines for datacenters



The challenges of securing OT networking protocols

The OT environment is unique place. It has special characteristics and legacies that have resulted in the use of insecure networking protocols. Addressing these challenges is essential for improving the security posture of OT environments and protecting critical infrastructure from emerging threats. For example, the lack of encryption for backend network traffic poses a risk if adversaries gain network access, including operational disruptions and potential sabotage. Other key challenges leading to insufficient security protocols include:

1. Latency and Performance: Implementing security measures like encryption can introduce latency due to the need for additional processing, such as SSL handshakes (Secure Sockets Layer connection to establish an encrypted link between client and server to enable secure data transmission), which can impact real-time operational requirements.
2. Legacy Systems Compatibility: Many OT systems rely on older, inherently insecure protocols that lack modern security features. In some cases, the hardware is not powerful enough to run encrypted protocols. Upgrading these systems to support secure protocols can be challenging and costly.
3. Certificate Management: Secure protocols often require managing digital certificates for authentication and encryption. This can be complex, especially for devices with limited computational resources or in environments with a large number of devices.
4. Operational Priorities: In OT environments, the priority is often on maintaining availability and operational continuity. Security measures that could potentially disrupt operations may therefore be deprioritized.
5. Resource Constraints: Developing and implementing secure protocols requires significant resources, including skilled personnel and financial investment. Organizations may struggle with allocating the necessary resources to enhance security.

Network security of embedded devices

While an application security program effectively secures customer-owned managed devices, datacenters face challenges from unmanaged employee devices, vendor equipment, and IoT/OT devices.

These devices often fall outside of established security policies, presenting risks due to their diverse nature and the organization's limited direct control and necessitating a different strategy to address them.

For example, runtime OT monitoring is an essential solution that helps organizations keep a comprehensive inventory of devices (including all information on operating systems, firmware, vendors, and models), assess the potential risk exposure from these devices, and detect any signs of malicious activity in real-time.

Firmware analysis on embedded devices can also be used to automate the identification of potential security vulnerabilities in these devices and identify and prioritize which devices need to be patched when new vulnerabilities are discovered. The additional layer of network security can be achieved using a non-intrusive tool (passive) and without any impact to the environments using a dedicated network sensor or utilizing agents running on the managed devices that can act as a data source to secure those unmanaged OT devices in the datacenter and without any deployment activity.

Actionable insights

1. Consider adopting a formal application security review program for critical OT assets.
2. Recognize the need to find a balance between timely deployment of security patches and maintaining availability.
3. Leverage solutions that build inventories of OT assets, prioritize risks, and help to identify malicious network activities.

Cyber Point of View: Africa

Increasing the cyber resilience of emerging economies

Emerging economies continue to struggle with the rising tide of cybersecurity threats. As a founding member of the Global Forum on Cyber Expertise (GFCE), Microsoft partnered with like-minded governments to tackle this risk through the first Global Conference on Cyber Capacity Building (GC3B)®, in Ghana.

With over 1,000 delegates including international leaders, decision-makers, and cybersecurity experts, the conference aimed to foster effective, sustainable, and inclusive cooperation for cyber resilience in emerging economies.

We announced our commitment to support GFCE's new Africa Hub, a regional initiative helping to address cybersecurity issues through local and regional means. Additionally, we brought to conclusion our workstream on mainstreaming cybersecurity into international development, issuing sets of recommendations in partnership with the Swedish Ministry of Foreign Affairs, the International Telecommunications Union, and GFCE.

Microsoft was also the first industry supporter of the Accra Call for Cyber Resilient Development, now endorsed by over 50 governments. The Accra Call is a critical commitment, emphasizing the need for global action by focusing on the needs and priorities of developing countries.

Links

Bridging the cybersecurity gap: a collaborative compendium for global development | Mar 2024

Managing software and firmware updates in the critical infrastructure environment

Last year, we highlighted research that used customer telemetry to show that while OT security vendors were patching critical vulnerabilities, there was a significant delay between the patches becoming available and when they were deployed—in some cases, up to 10 years.

In the following section, we explore the OT software update challenge by using our Azure datacenters as a case study, showing that increased software security only works when it's actually deployed.

A fundamental difference between traditional IT and OT is the need to prioritize systems availability. This is because the OT infrastructure is supporting critical services where disruptions and outages could have significant, even life-threatening consequences. At its core therefore, managing the update process is a supply chain integrity issue: managing the equipment, core software, component origin and how they changed between updates. Any deviation from expected operations or incompatibility in an update can cause an outage, and this is difficult for vendors to manage.

For example, we saw variations in firmware versions of a cooling system take out multiple datacenters. Contrast this with our fully cloud-managed system where we control the entire supply chain and regularly update hundreds of thousands of devices in under two days without outages. We highlight this approach as a north star for addressing the OT update challenge, but recognize it is one that will take years to achieve.

Difficulties in updating software in the OT environment

Compared to traditional IT software, there are several key points to consider:

- Availability is paramount: In datacenters, the primary function of OT systems is to ensure the continuous operation of critical infrastructure, such as power management and cooling systems. Any disruption in these systems could lead to significant operational issues, including potential downtime of services provided by the center.
- Security as a component of availability: While availability takes precedence, security is not ignored. It is considered a component of availability, since security breaches can lead to service disruptions. Therefore, security measures are implemented in a way that they do not compromise the availability of the OT systems.
- Updating software in production environments: This is not as simple as just installing an update: it involves extensive testing to ensure updates do not disrupt the operational functionality or introduce new vulnerabilities. Updates must be carefully planned and executed to avoid any disruption in service. This often means that security patches and updates may be delayed or scheduled during maintenance windows to minimize impact on availability. The challenge is to balance the need for security with the imperative of maintaining uninterrupted operations.
- The "infinite loop" challenge: Teams that manage our datacenters face the "infinite loop" problem, where updating to a new version of software to address security vulnerabilities can introduce new vulnerabilities or dependencies, leading to a cycle of continuous updates. This creates a situation where organizations are always carrying some level of security debt, as new updates can potentially introduce new issues.

Managing the OT software supply chain

OT applications and devices can be complex, with hardware components, operating systems, core software, and supporting libraries sourced from a wide variety of suppliers.

For a clean, seamless software update, all these elements need to be controlled so that devices continue to operate as expected—if not, there is a risk of service interruptions and downtime.

This end-to-end control of a device update is very difficult to achieve, and results in the "infinite loop" cycle described above, impacting both vendors and security teams. Vendors often suggest upgrading to the next version as a solution to vulnerabilities found in the current version, but this too can introduce new vulnerabilities. This cycle is challenging for security teams as it becomes very difficult to achieve a state of minimal vulnerabilities.

Actionable insights

Based on our OT experiences with our datacenters, we recommend the following:

1. Vendors should provide clear documentation on the changes each update brings, including any new vulnerabilities introduced. This transparency can help security teams make informed decisions about updates.
2. Encouraging vendors to provide incremental updates that fix current issues without introducing significant new features can minimize the introduction of new vulnerabilities.
3. Security teams should work closely with vendors to understand the impact of updates and prioritize fixing of critical vulnerabilities that do not introduce significant new issues.

Datacenter outages caused by firmware version mismatch

A recent real-life datacenter outage provides a compelling case study on the challenges of software updates in the OT environment.

In our datacenters, OT cooling management systems monitor the temperature, adjusting both fan speed and air flow to keep servers within an acceptable operating range. If they fail the server will overheat, and to prevent physical damage the servers will be shut down.

In this particular case, a firmware update was deployed to these systems, but due to a bug the devices with the new firmware did not communicate properly with the devices running the previous version.

This resulted in a "packet storm" on the network controlling these devices, with high volumes of messages being passed between the devices. As a result, the devices ran out of memory and restarted. As they came back online, the devices operated at a low fan speed. As a result, significant parts of the datacenter computer infrastructure experienced a spike in temperature and had to be shut down.

This outage occurred across multiple centers. While the teams quickly identified and resolved the issue, it highlights how firmware updates implemented without significant testing can potentially cause unexpected outages.

Experiences with fully-managed device updates

Our north star for OT software updates is one where all components in the software and device supply chain can be controlled to minimize unexpected changes in device behavior on update.

A good example is our experience in managing updates for Azure Sphere, which combines hardware, OS, and a fully cloud-managed application and security environment. This allows control and visibility over the end-to-end supply chain of the product. Achieving this requires a comprehensive test and development lifecycle process to ensure updates can be deployed to the entire fleet simultaneously, maintaining high standards of reliability and security.

Approaches like this will not be applicable in every OT environment, and the costs on the vendor side are not insignificant. However, we do see this as a long-term approach to address security vulnerabilities in the OT environment while maintaining the availability promises required to protect critical infrastructure.

We consistently see updates deployed at scale to the entire fleet of devices. Typically, hundreds of thousands of devices are updated within 48 hours of deployment with no production outages or downtime issues reported.

Key steps included:

- **Building verification tests:** These tests, both hardware and emulated software, validate that the update compiles correctly and maintains the expected contract with applications. This ensures that updates do not break existing functionalities.
- **Strict contractual layer:** Azure Sphere maintains a strict contractual layer guaranteeing application compatibility across OS updates. This means applications are assured to run independently of OS updates, allowing for predictable behavior post-update.
- **Customer and Original Equipment Manufacturers (OEM) testing:** Before an update is released to the retail environment it undergoes testing by customers and OEMs in a retail evaluation setting. This step allows for real-world testing and validation, ensuring that updates do not introduce new issues.
- **Scheduled updates:** The Azure Sphere platform allows for updates to be deferred, accommodating operational requirements. This flexibility ensures that updates do not disrupt critical operations.
- **Fleet management at scale:** Azure Sphere supports fleet management capabilities, allowing for updates to be managed and deployed across devices at scale efficiently. This includes the ability to set configurations remotely and manage devices autonomously.

Links

Exposed and vulnerable: Recent attacks highlight critical need to protect internet-exposed OT devices | May 2024

Microsoft to help rural hospitals defend against rising cybersecurity attacks - Stories | Jun 2024

Collective action

The digital transformation of defense and a call for partnership

Hybrid warfare, cyberattacks, and foreign influence operations pose grave risks not just to IT systems, but to the stability, prosperity and national security of society itself.

As cyber risks take on more real-world consequences, digital technologies can be powerful tools to enhance our traditional defense capabilities. However, to make this a reality, we need a deeper partnership between industry and governments to implement the digital transformation of the defense sector.

Technology and initiatives touching on AI and cybersecurity were front and center at the recent 75th summit of the North Atlantic Treaty Organization (NATO). Cloud computing, AI, and quantum computing all have a role in cybersecurity, but their impacts on our collective defense can only be maximized through joint action and collaboration in defense innovation. Initiatives like the NATO's Defense Innovation Accelerator for the North Atlantic (DIANA) and the NATO Innovation Fund (NIF) exemplify the strength of these collaborations.

Beyond NATO, Ukraine's digital transformation of the defense sector has showcased the effectiveness of industry-government collaboration and the use of commercially available technologies in real-world security situations. Microsoft believes that structures like NATO can complement national efforts in advancing the digital transformation of defense, as they help to maintain a broad perspective on standardization and interoperability efforts and ensure scale across an alliance.

RAISE: The Roundtable for AI, Security, and Ethics

The Roundtable for AI, Security, and Ethics (RAISE) exemplifies the power of collective action through strategic partnerships and inclusive dialogue.

Led by the United Nations Institute for Disarmament Research (UNIDIR) and launched in partnership with Microsoft, RAISE is an initiative dedicated to AI for national security, grounded in international legal and normative frameworks.

RAISE launched in March 2024, assembling experts from industry, academia, civil society, and government. Initial participants included representatives from China, Ecuador, India, Israel, Japan, Namibia, Russia, Switzerland, the UK, the US, and others who worked to identify shared interests, enhance cooperation, and generate actionable recommendations. Its goals are to reduce the risks of AI in national security, support multilateral AI governance, and promote AI to enhance security globally.

This is done through six priority themes:

1. Trust-building: Establishing trust in AI development, deployment, and governance is crucial for national security. RAISE's trust-building initiative promotes transparency, accountability, and adherence to international norms, setting the stage for responsible and ethical AI governance.
2. Developing the knowledge base: RAISE aggregates and analyzes authoritative research to inform policy decisions and guide the application of international law and norms in AI for national security, building a comprehensive knowledge base.
3. Integrating the human element: This initiative focuses on the ethical, social, and psychological aspects of human-AI interactions and decision-making, ensuring that AI governance is rooted in principles of human-centered design, inclusivity, and ethical responsibility.
4. Data practices: RAISE's data practices initiative examines how data is sourced, curated, and used in AI systems, addressing issues such as biases, explainability, and auditing to ensure responsible and lawful AI in national security.
5. Lifecycle management: RAISE promotes governance approaches that manage AI technologies across their entire lifecycle, emphasizing ethical and legal considerations "by design" to ensure responsible integration and disposal within existing systems.
6. Destabilization: This initiative explores the security implications of AI as both a force-multiplier and threat-multiplier, aiming to develop governance solutions that mitigate risks of AI-related destabilization and contribute to global stability.

Links

RAISE: The Roundtable for AI, Security, and Ethics | UNIDIR

How Microsoft helps support democratic elections

During this unprecedented period of critical elections taking place around the world, Microsoft has worked to defend democratic institutions by combatting malicious schemes designed to disrupt or influence electoral processes and promoting a healthy information ecosystem. Our initiatives stem from four key principles:

1. Voters have a right to transparent and authoritative information regarding elections.
2. Candidates should have the ability to verify the authenticity of content originating from their campaigns and have access to procedural or legal mechanisms to address instances where their likeness or content is manipulated by AI to mislead the public during elections.
3. Political campaigns should have the resources to safeguard against cyber threats and effectively utilize AI, with access to affordable, easily deployable tools, training, and support.
4. Election authorities should be able to ensure a secure and resilient election process and have access to tools and services that enable this process.

Detection. Microsoft uses advanced tools and capabilities to monitor, analyze, and attribute malicious activities or campaigns that aim to disrupt, influence, or manipulate elections and elections infrastructure. We leverage our global network of partners and sources to gather intelligence on the threat landscape and the actors behind it.

Response. Microsoft responds to and mitigates the threats to elections around the world through several means. The Digital Crimes Unit uses its legal and technical expertise to disrupt the malicious activities and campaigns intended to compromise, sabotage, or interfere with the elections. Microsoft Incident Response and other Microsoft security partners help political and elections customers respond to and recover from active cyber incidents via our Election Security Advisors program.[59] In early 2024, we launched a site where certified candidates in any national or federal election can directly report deceptive AI election content on Microsoft's platforms.

Collaboration. We collaborate with public and private stakeholders globally who share a similar goal of protecting the electoral process. This includes local elections officials and elections commissions, working across the tech sector on initiatives like the Tech Accord to Combat Deceptive Use of AI in 2024 Elections,[60] and with government agencies or law enforcement bodies when appropriate.

Microsoft is also helping protect the online environment surrounding elections by:

Defending the information environment

Identifying disinformation campaigns propagated by nation-state actors and collaborating to mitigate the potential risks of deepfakes.

- Tech Accord: A cross-sector coalition to combat deceptive uses of AI in elections.[41]
- Public election influence operations reports: The Microsoft Threat Analysis Center releases timely public reports about cyber and influence threats.[42]

Protecting data

Protecting elections-focused employees and official systems, including combatting phishing lures using elections-related themes

- Advanced security and productivity tools for political campaigns.[43]
- Advanced support for customers running elections-critical workloads in Azure, like voter registration or results reporting systems.[44]

Identifying and responding to threats

Utilizing our significant threat intelligence capabilities to identify threats and identify mitigations.

- Advanced threat detection and notification against nation-state attacks for high-risk elections customers available in 35 countries.[45]
- Election security advisors providing expert consultation for proactive cybersecurity audits, threat hunting, or remediating cyber incidents.[46]

Boosting public awareness of AI elections risk

These are some ways we contribute to educating the public on the potential misuse of AI in elections and promoting transparency in AI-generated content.

- Societal Resilience grants with OpenAI: $2 million in grants to enhance AI education and literacy among voters and vulnerable communities.[47]
- Content Credentials: Implementation of authenticity markers on AI-generated and authentic images and video to help the public discern if media has been created or edited by AI.[48]
- Security and deepfake trainings for political stakeholders: Ahead of major elections, Microsoft provides cybersecurity hygiene and deepfake response trainings to political organizations.
- Public awareness campaigns: Launch of several public awareness campaigns in the EU, US, and globally, to ensure voters are aware of the risks of deepfakes and to guide users to authoritative election information sources.

Links

Microsoft's efforts to enhance the security of Indian elections | Jun 2024

Microsoft and OpenAI launch Societal Resilience Fund | May 2024

Combatting abusive AI-generated content: a comprehensive approach | Feb 2024

AI Elections accord - To Combat Deceptive Use of AI in 2024 Elections | Feb 2024

Microsoft announces new steps to help protect elections - Microsoft on the Issues | Nov 2023

Combatting the deceptive use of AI in elections – Middle East & Africa News Center (microsoft.com)

Cyber Point of View: UK

A continuously improving security partnership

Microsoft and the UK's National Cyber Security Centre (NCSC) have been in partnership for over 20 years. From securing user devices to covering the UK Government's broader cloud ecosystem, we are building a secure foundation to protect from the most common cyberattacks.

Together, they have developed the "Secure Configuration Blueprint" to help government departments configure Microsoft 365 in a way that helps meet their statutory obligations. The blueprint leverages the service's inbuilt features and capabilities to lower residual risk. It is the leading practice published by the NCSC and Microsoft, drawing on extensive experience across the UK Government and industry.

Additionally, secure configuration practices have been extended beyond the UK Government to include principles for manufacturers of enterprise-connected devices and networking equipment.

Manufacturers can use the principles to determine which security mitigations can be designed and built into their products by default, and, in parallel, organizations can use the same principles as a framework to assist in the procurement and validation of secure, enterprise-connected devices.

Not only does this reduce the complexity of the procurement process, it increases the speed of deployment.

The ultimate goal of this partnership is to enhance the security posture of public and private sector organizations in the UK, ensuring data protection, effective collaboration, and reliable services.

Links

Updated Microsoft 365 security and compliance guidance for the UK public sector - Microsoft Industry Blogs | Feb 2024

National Cyber Security Centre - NCSC.GOV.UK

Device security principles for manufacturers - NSC.GOV.UK | May 2022

Chapter 3

Early insights: AI's impact on cybersecurity

What do we know about new AI challenges and solutions today?

Key developments	83
Introduction	84
Emerging threat landscape	87
AI for defense	94
Advancing global AI security	101

Key developments
Early insights: AI's impact on cybersecurity

AI is reshaping the landscape of cybersecurity, arming defenders with powerful tools to preempt and counteract evolving threats with unprecedented precision. As we explore this transformative era, we are met with both promising advancements and daunting challenges—from sophisticated AI-powered targeting to complex influence operations orchestrated by nation-state threat actors.

As ever, information is power. The more knowledge and understanding an organization has of the emerging threats, the better it can prepare. In this chapter we explore how AI is changing everything from enhancing detection capabilities and operations efficiencies, to customized mitigations. At the same time, governments and industry are collaborating, and using a variety of approaches, to advance global cybersecurity initiatives in the AI era.

AI-enabled human targeting

These threats will be more difficult to detect and defend against—even with AI tools assisting defensive strategies.

Find out more on p89.

Emerging threat actor techniques

AI-enabled spear phishing, résumé swarming, and deepfakes emerge.

Find out more on p90.

Governments and industries working to advance global AI security

While there is a consensus on the importance of security in the development of AI, governments have pursued different approaches in implementing security requirements.

Find out more on p101.

Nation-state threat actors are using AI for influence operations

AI-generated images and audio manipulations are being used to shape audience perception and engagement in conspiratorial narratives.

Find out more on p91.

AI for defense

Defenders are using AI to become more efficient, especially in security operations.

Find out more on p94.

Limiting foreign influence operations in the modern era

Existing limitations of foreign influence operations under international law are no longer sufficient in the modern era.

Find out more on p93.

Staying a step ahead of threat actors in the age of AI

Policy principles can mitigate risks associated with use of AI tools.

Find out more on p106.

Introduction: AI's impact on cybersecurity

We are at the start of what could become one of the most transformative technological eras in modern history. Much has been said and written about how AI can have a significant effect on every industry, but the impact it can have on how businesses secure their most important data and assets in the face of ever-increasing cybersecurity threats will be one of the most critical uses of this technology.

Organizations of all sizes around the world are facing the same challenges: infinite amounts of data to manage, more endpoints to secure, and a shortage of talent to operate security environments that are becoming more complex every day. Cybersecurity is a top priority for businesses of all sizes, but at the same time, cybersecurity is an infinite game that has no winner and no end. Defenders must constantly be vigilant as the landscape becomes more intricate. With threat actor adoption of AI, the economics and sophistication of attacks are changing rapidly, and with that, the sophistication of how we must defend.

Generative AI is ushering in a new era of cybersecurity that can put defenders one step ahead of threat actors. The adoption of large language models (LLMs) tailored for security operation scenarios will see a shift from humans having to write manual automation of repetitive tasks to AI systems capable of detecting and investigating security threats at the skill level of security professionals. AI can help develop a thorough understanding of a security incident and how to respond in a fraction of the time it would take a person to manually process a multitude of alerts, malicious code files, and corresponding impact analysis.

Not only can this significantly reduce the time to identify, investigate, and respond to an incident from days to minutes, but this AI-driven threat analysis provides the opportunity for security teams to learn and train in real-time, helping to reduce the skills gap and freeing up experienced analysts to focus on more important tasks.

Today, the industry has taken the first steps to add assistive agency into products, and more autonomy will be created over time, enabling agents to perform tasks, monitor, and take action proactively and in collaboration with security teams. AI agents will use language models in incredible ways to get much closer to the way security analysts operate in reasoning, decision making, and task completion. Not only will we see security teams supported by these agents, but we will also see agents working together to investigate and resolve incidents. Agents will respond to events when activated or given permission by an analyst, and Microsoft sees a world where soon AI agents will potentially reason, make mistakes, learn from mistakes, and work together like a team of experts.

The deployment and utilization of AI and agents will be vital, especially with threat actors becoming more sophisticated in their tactics every day. But as history has shown, technology can have the ability to elevate our human potential, and through innovation, collaboration and responsible use of generative AI and agents, defenders will be positioned to take on cybersecurity's toughest challenges and work toward making the world safer for all.

Shawn Bice
Corporate Vice President,
Cloud Ecosystem Security

Understanding how generative AI systems work

Generative AI is one of the most impactful technological shifts of the past several decades. Its tremendous range of applications could put it not only into thousands of existing systems and business processes, but into a range of entirely new processes.

However, a rapidly changing world likewise creates opportunities for threat actors, who can often adapt to changes faster than defenders. To protect against this, it's important to understand how generative AI works and how to apply the techniques of safety and security to it.

Predictive vs. generative AI: In traditional predictive AI, people build individualized models using their own data. Since they have control over the process, they try to build controls things like fairness, training data leakage, and data poisoning. Predictive AI is good at analyzing large fields of data, classifying, predicting, and recommending. Generative AI, on the other hand, is best understood as a different technology – one where general-purpose models are shared by millions of users and have no special data access. Generative AI models are good at summarizing or analyzing natural language data and role-playing characters like "customer service rep" or "math teacher."

How they work: First-generation generative AI applications were just bare language models for which users carefully crafted inputs ("prompts") to receive outputs. Because they were creative tools, they were very prone to hallucination and required careful prompting for effective use. Second-generation applications, widespread today, are more complex, with many skills. Some of these skills are ordinary functions written in normal programming languages; these might look up data or call another system to perform some action. Like any function, they require very specific inputs. Such skills are natural points for controls since they are the only way the AI system can access outside data. Other skills call the model—for example, to summarize or analyze data, create content (such as by roleplaying a writer or programmer) or to improve data by roleplaying an editor. Those skills work by fusing user inputs, additional data fetched by other skills, and their own data to create a prompt, automating what people did by hand in the first generation. Skills are then called by a central function that defines the AI system: this might either be an ordinary function, running known steps in turn, or an AI function that roleplays a subject-matter expert and asks it to come up with a plan using those known skills.

How Copilots work

Such systems have natural safety intervention points. Metaprompts are the step that tell the model about the character it's roleplaying, things it should avoid, and so on, and are a key place for defense. "Editor" steps are a kind of metacognition, where a second AI looks at the outputs of the first to see if its statements are grounded in its known list of facts, if they are aligned with its compliance or strategic goals, and so on. Filters pre- or post-process data using ordinary software, predictive AI, or generative AI to catch suspicious situations and handle them differently.

For example, a question-answering system might first use a filter to see if it's being asked something inappropriate or outside its expertise. If not, it role-plays a subject-matter expert to figure out searches it should run; then again as a subject-master expert evaluates the credibility of each page and extracts its key ideas; then role-playing a writer, it combines these to form an answer to the original question. Finally, as an editor, it checks to see if every part of its response is appropriately grounded and if any key points were missed.

Future generative AI systems ("agents") are likely to add capabilities like memory (learning as they work with you), operation over longer timescales than a single conversation, and more autonomy, reacting to events other than user inputs.

Two key insights

As we've developed a large range of generative AI systems, we've found some important insights about the process.

1. Building is easy; testing is hard

Generative AI changes the traditional relationship between development and testing investment. In traditional software, 90% of the work goes into writing software that will function. With generative AI, writing a system is much easier, with significant features being "quick projects" rather than multi-month investments. However, that AI system will work correctly only in the handful of cases that the developers imagined as they worked; the majority of the work will be in testing and tuning as the system is evaluated on uncommon inputs, adversarial inputs, or even just inputs from users who think differently from the developers.

As a system is built, it's important to make a list of the ways in which the system could potentially go wrong and develop a large test suite of example inputs that may trigger those outcomes. Likewise, there should be lists of intended and "uncommon" inputs as well. Team diversity is key at this stage, since without it the team can't adequately imagine how real-world use will look and will miss critical risks. Generative AI can itself amplify a team's ability in this space, turning individual examples into large multi-lingual lists.

Then, using a test framework, these lists can be run against the system in bulk, with generative AI once again helping efficiently evaluate the outputs for correctness. These tests can re-run whenever the system is updated, much like ordinary integration tests.

2. Generative AI security is nondeterministic

Generative AI systems are software, and traditional software security remains important. In addition to that, however, generative AI systems face risks from anomalous natural-language (or media) inputs. These are nondeterministic, especially in that variations in language or phrasing can profoundly change behavior. In fact, most "jailbreak" attacks can be summarized as "social engineering works against generative AI." The resulting vulnerabilities can't be deterministically patched, even in theory.

Fortunately, it is possible to secure systems with nondeterministic components; we call those components "people." Microsoft has found that asking "How would you secure this if it were a person?" scales very effectively to generative AI. Where an organization would vet a person, they should test a system thoroughly and adversarially. Where it would train a person, it should adjust metaprompts and filters so that they behave correctly. Where it would have multiple eyes check and approve sensitive decisions, organizations should do the same—both by having one AI look over the results of another (metacognition) and involving humans in the process.

A surprising fact that makes metacognition more effective is that since generative AI is trained on human language, a brief summary of a character it is meant to roleplay allows it to infer broad aspects of that personality without the user having to specify those qualities manually. That means, for example, telling the system it is an experienced newspaper editor or computer hacker lets it do meaningful edits and safety checks from those perspectives.

Links

- Responsible AI Transparency Report | Microsoft
- AI Content Safety | AI Content Moderation
- PyRIT: Python Risk Identification Tool
- AI Red Team Guidance | Microsoft Learn
- AI jailbreaks: What they are and how they can be mitigated | Jun 2024
- The HAX Toolkit Project - Microsoft Research

These attacks are different

They're nondeterministic:

- Saying the same thing twice won't have the same effect
- Slight changes in phrasing may change the outcome

This means you can't "patch" them the same way you do traditional security vulnerabilities

Map human ideas to generative AI safety

For a person, you might...

- Vet them
- Train them
- Monitor them
- Have multiple eyes check and approve sensitive decisions
- Build trust over time

For a Copilot, you might...

- Test the system thoroughly and adversarially
- Adjust metaprompts so they behave right
- Monitor them
- Have multiple AIs look at a problem (metacognition)
- Have humans in the loop

Integrate the Copilot into your business practices like you would a new person—step-by-step.

Emerging threat landscape

The AI landscape is changing tremendously quickly, and any analysis will therefore inevitably be out of date by the time it is published. While details of any summary will quickly become obsolete, its principles may prove useful for much longer.

The generative AI threat landscape

When discussing AI threats, a first division is between system threats—issues like security vulnerabilities, where securing one system effectively mitigates the risk—and ecosystem threats, where attackers can choose the most vulnerable system with which to achieve their goals.

System threats

The key system threats Microsoft has seen are system compromise, overreliance, and content exposure.

- System compromise: The key threat here is cross-prompt injection attacks (XPIA, also known as indirect prompt injection), where the system is processing data under the control of a third party (for example, email messages or Word documents). Attackers insert malicious payloads to exploit vulnerabilities in the way the system combines inputs to form LLM prompts to do things such as run commands with a victim's credentials, take over systems, and/or exfiltrate data.
- Overreliance: Users tend to overrate the reliability of AI output. The best mitigations for these threats are often in the user experience (UX) or business practice. Overreliance comes in four forms:
  - Naive, where users aren't aware of the limitations of the AI.
  - Rushed, when a lack of time or confirmation blindness means users don't check outputs.
  - Forced, when the user is physically unable to check the output (for example, vision augmentation for the blind, or systems that create apps for non-programmers).
  - Motivated, when a user offers "the AI said so" as an excuse to do what they already wanted.
- Content exposure: A threat exists when operators are exposed to content such as hate speech, violence, radicalization or child sexual abuse material (CSAM) that is directly harmful to them. Fortunately, this threat can be defended against with filters and metaprompts, such as with Azure AI Content Safety.
- Infrastructure compromise: Traditional cybersecurity threats against the underlying storage, network, computing, and supply chain continue to be significant.

Ecosystem threats

Ecosystem threats often require defense outside the AI system.

- **Impersonation:** Use of image, audio, and video deepfakes to impersonate individuals. Specific threats include fraud, blackmail, coercion, defamation, and information warfare. Defense against this threat includes moving communication to authenticated channels.
- **Content production:** Creation of harmful content for dissemination such as CSAM, non-consensual intimate images, disinformation, child grooming scripts, or spam. Threats in this category are diverse and are typically amplifications of existing threats, sometimes on a large scale.
- **Nefarious knowledge acquisition:** Acquisition of information that helps threat actors upgrade their skills, such as how to make drugs or biological weapons. This threat has not emerged at scale and is being actively researched by the security community.
- **Cyber threat amplification:** Automated generation of malware, and, more importantly, attack command and control infrastructure, which could give lower-tier threat actors access to persistent attack capabilities previously limited to advanced actors. There is significant risk that attackers will develop AI techniques faster than defenders adopt AI-powered defense systems, and attackers will take advantage of that gap until the defenders catch up.
- **Direct social attacks:** Malicious activities such as scams, phishing, propaganda, and terrorist recruitment can be automated using generative AI and operated at far larger scales than before. The cyber intelligence community anticipates a large rise in all of these categories, driven by AI enablement. Defenses against them may focus on providing AI support to the recipient and interdicting payloads at the communication system level.
- **Indirect social attacks:** Automated harassment and defamation are very difficult to counter. Because someone can be harassed by targeting their friends, colleagues, and the public, a defense strategy for this threat type is not yet clear.

As defenders, particularly governments, are considering the threats associated with the abuse of AI, it is important to keep in mind that many of the future victims will not have the benefit of automated systems and programs to defend them. Many ecosystem threats will have an immediate impact on the most vulnerable targets—humans. As difficult as it currently is to stop multi-billion-dollar frauds against vulnerable groups like the elderly, AI's impersonation capabilities will make it even harder for victims to identify and resist fraud.

In the coming year, we anticipate the biggest rises in automated fraud and election interference, CSAM and non-consensual intimate image (NCII) production, and the use of XPIA and deepfake impersonation as cyberattack and fraud channels.

Sophisticated AI-enabled human targeting

Behind every bot is a real person. As AI is increasingly used to help people get more efficient, threat actors are learning that they can use the same AI efficiencies as a force multiplier in their targeting efforts.

Targeting high-value individuals

Threat actors target high-value individuals at organizations because they have access to trade secrets, financial systems, key strategies, and other sensitive and proprietary intellectual property. Because AI is very capable of performing most of the time-consuming research needed to identify lucrative targets, it frees the actors up to conduct other activities. This emerging threat landscape of AI-enabled targeting is also aided by the machine learning (ML) aspect of AI. This is because bots can rapidly learn from the sum total of human knowledge documented on the Internet.

We expect two diverging trends pertaining to AI-enabled cyber-threat actors and defenders. Whichever party masters AI faster will have a near-term advantage. However, when it comes to AI-enabled human targeting, threats will be more difficult to detect and defend against—even with AI tools assisting defensive strategies.

The defensive advantage

The defensive advantage against AI-enabled cyber threats comes in the form of defenders' ability to deploy AI into defensive tools and systems. If organizations are early adopters of AI tools, they can use ML to rapidly ingest and infer evolving tactics, techniques, and procedures (TTPs), thus detecting and preventing malware and malicious code. Hesitance to incorporate AI into defensive strategies on the other hand, will open a window of opportunity for threat actors to exploit gaps they identify with AI tools. This means the early AI adopters will enjoy a near-term advantage afforded by the nimbleness of AI.

The offensive advantage

The offensive advantage of AI-enabled human targeting comes from AI's ability to:

1. Rapidly perform functions that previously took humans months or even years.
2. Avoid hallmark mistakes humans make in their targeting operations.

For example, a threat team that previously relied on manual operations to identify targets, research them, develop a social engineering approach, and execute it can assign roughly 90% of this work to AI, freeing up human resources to perform more nuanced tasks AI is not yet effective at performing.

Since AI can perform these labor-intensive tasks far more rapidly than a human, it also reduces the time to target. This operational efficiency is complemented by the fact that AI won't make, for example, spelling and grammar errors that humans make in phishing communications.

The impact of AI in attacks is already being felt in the wider cybersecurity community. Tools in the multi-factor authentication toolbox are becoming vulnerable, and AI has demonstrated it can defeat CAPTCHA,[19] which was specifically designed to stop bots. The use of AI will expand the threat landscape by making bots harder to detect, more pervasive, and more adaptable due to increasingly sophisticated ML capabilities.

With the Internet of Things (IoT) market growing at 42% per year,[20] we also expect pervasive targeting of personal and home-use products. Overall, the democratization of AI will enable unsophisticated threat actors to become more capable and effective without having to become more technically proficient.

Emerging techniques in AI enabled attacks

While some TTPs are in their infancy and little more than proof of concept, others are already being widely used.

This section discusses some of the TTPs threat actors are currently using and evolving for use against their targets in the social engineering phase of attacks. We expect threat actors to rapidly involve and deploy these TTPs in the near term, and the variations will continue to evolve and expand.

AI-enabled spear phishing and whaling

AI is evolving spear phishing and whaling by coupling AI with malware, creating a tool that lies dormant until it identifies its intended target and deploys. Threat actors can focus their attacks on highly specific targets and hone-in on exfiltrating only the most useful information. Without users knowing, the AI uses device cameras, speakers, and GPS for target verification. By the time it is discovered, the malware has already exfiltrated the target information.

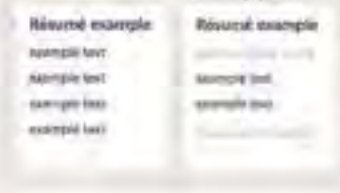
Links

Digital Safety | Report a concern

"Résumé swarming" and steganography

Threat actors can use AI to scrape keywords and qualifications from job postings and develop "perfect" candidates in the virtual world. AI can then generate hundreds or thousands of variations of highly qualified—but imaginary—candidates' résumés to apply for open positions at unsuspecting companies.

These résumés can even use steganography techniques to embed invisible information to increase their chances of passing automated screening tools, getting the applicant selected for interviews and ultimately hired. Threat actors can use this technique in their attempts to emplace insiders within an organization to steal trade secrets, intelligence, or other sensitive information. In another variation of this technique, threat actors may create a limited number of ideal candidates alongside a swarm of AI-generated unqualified résumés to break screening processes.



Deepfakes and other variations on social engineering

Using AI's capability to rapidly conduct expansive research, threat actors can discover massive amounts of information about targeted individuals and programs.

This means they can then develop highly tailored social media profiles with which to contact thought leaders, subject-matter experts, and other high value targets for social engineering. Further enhancing this false persona technique, AI-enabled deepfake tools can also be used to create fake social media profiles impersonating people known to the target.

Threat actors can establish the impersonating persona's bona fides by using video teleconferencing or phone calls to deploy real-time deepfake contact with voice and video synthesis. Or, using AI bots, threat actors can automate a substantial portion of communication before actual human interaction is required. All these AI-assisted approaches act as a force multiplier that can help threat actors simultaneously approach a virtually unlimited number of potential targets to identify the most viable targets for further development.

With the increasing sophistication and quality of deepfakes, we anticipate that it is highly likely that criminals will also use this TTP for fraud, identity theft, blackmail, and extortion. Nearly flawless deepfake video with audio can generate extremely convincing (false) evidence to compel and coerce victims to comply with criminals' demands.

Even knowing the artifacts are fake, many victims may choose to comply simply to avoid embarrassment or potential negative perceptions.

A strong mitigation strategy will seek to reduce the threat landscape through predictive and preventative activities. Incorporating AI into risk mitigation activities means defenders can evolve at the same or a greater rate as threat actors. As discussed in the data security section of this report, discovering and prioritizing data assets is foundational. Threat actors rely on disorganization, poor communication, lack of consensus, and unwillingness to invest in non-revenue generating activities within organizations. We therefore recommend mapping identified gaps to key stakeholders responsible for managing the associated mitigation strategy. Lastly, one of the best mitigation strategies is robust training and awareness campaigns.

Actionable insights

1. Report criminal and suspicious activity to the appropriate law enforcement organization in your region.
2. Reporting suspicious activity, whether or not you fall victim to it, enables defenders to better understand the threat, identify what's being targeted, take action to protect those targets, and educate the population about protecting against those threats.

Nation-state threat actors using AI for influence operations

Nation-state threat actor groups, such as those backed by Russia, Iran, and China, are increasingly incorporating AI-generated or enhanced content into their influence operations in search of greater productivity, efficiency, and audience engagement.

We assess this content has had a limited effect on the impact of nation-state influence operations thus far, but if integrated into otherwise creative and multifaceted influence operations, AI may prove to offer a significant capability in reaching and engaging audiences in the future.

Adversarial use of AI in influence operations

Capability	China	Russia	Iran & proxies
Text	MEDIUM / LOW	MEDIUM / LOW	LOW
Image	HIGH	HIGH	MEDIUM / LOW
Audio/video	HIGH	HIGH	LOW
Example	May 2024: Bespoke Taizi Flood AI-generated cartoon	June 2024: AI-generated audio of Elon Musk narrating fabricated documentary	April 2024: Likely AI-generated video leading up to Iranian military operation

China-affiliated influence actors favor AI-generated imagery

China-affiliated threat actors' increasing use of AI to enhance influence campaigns, especially those targeting elections around the world, distinguishes them from other nation-states using AI.

In the past year, Microsoft observed China-linked threat actors utilizing various generative AI technologies to create sleek, compelling visual narratives. Microsoft uncovered a series of AI-generated memes aimed at the United States that emphasized domestic discord and criticized the Biden administration.

Taizi Flood is the most prolific threat actor in this arena, using third-party AI technology, including technology that generates virtual news anchors, for its online campaigns. With influence operations spanning over 175 websites and 58 languages, Taizi Flood has continuously mounted reactive messaging campaigns around high-profile geopolitical events, with a focus on portraying the United States in an unfavorable light and furthering Beijing's interests in the Asia-Pacific region. During the Maui, Hawaii wildfires in August 2023, the actor used AI-generated images of burning coastal roads and residences to augment the conspiratorial narratives about US Government complicity it spread across social media platforms.

More recently, the actor attempted to fan the flames of discord around the Israel-Hamas war by circulating photorealistic AI-generated images of purported protests, as Israel-Palestine related university campus protests surged across the United States in late April to May 2024.

Taizi Flood's "photorealistic" AI-generated images intended to portray protests at a named US university.

Russia-affiliated influence actors using audio-focused AI across mediums

Russia-affiliated threat actors often adopt a more nuanced strategy in their AI tactics, though the effectiveness of their campaigns has had mixed results.

For example, they create fully synthetic deepfake videos of prominent political figures but the videos struggle to gain significant online engagement because they are quickly exposed as fake.

Audio manipulations have proven more influential in shaping audience perception. Two days before Slovakia's 2023 election—a tight race between pro-Western and pro-Kremlin parties—AI-generated audio of the pro-Western party leader discussing how to rig the election appeared online.[7] The incident represented a test case of how vulnerable elections around the world could be to the malicious use of AI by nation-state threat actors.

In February 2024, pro-Russian social media accounts circulated a fabricated video, falsely claiming that Ukrainian authorities planned to assassinate French President Emmanuel Macron. While the visual component of the video appeared to be from an authentic France24 broadcast, the audio component was AI-generated.[8] The video gained traction online and former Russian President Dmitry Medvedev later repeated the false narrative in a post to X, without explicitly referencing the video itself.[9] Although attribution is unclear in both the Slovak and French examples, the targets, narratives, themes and tactics are consistent with pro-Russia influence activities.

A still image from the fabricated video: The image features a well-known French news anchor with likely AI-generated audio of his voice. The overlaid text graphic appears slightly [illegible]. The original, observed version of the video included Russian subtitles, as reported [illegible].

Russia also used a malicious application of AI in influence operations surrounding the 2024 Paris Summer Olympics. In mid-2023, Microsoft identified a fake documentary titled "Olympics Has Fallen" disseminated by Russian-affiliated influence actor Storm-1679 on Telegram.

The video featured AI-generated audio that mimicked the voice of American actor Tom Cruise criticizing the International Olympic Committee and its leadership. This was Storm-1679's first use of AI-enhanced content for influence efforts. In June 2024, the actor launched a sequel, "Olympics Has Fallen II," this time featuring AI-generated audio impersonating businessman Elon Musk. For both videos, Storm-1679 appears to have allocated significant time and resources. This ongoing initiative reflects a persistent effort to target Western audience information spaces where this actor has traditionally struggled to effectively amplify its content.

Iran-affiliated influence actors are in the early stages of AI integration

In contrast to actors supporting Russia and China, pro-Iran groups have so far employed AI more sparingly. Nevertheless, they are gradually increasing use of AI-generated or enhanced images and videos as key components of their messaging campaigns, particularly against Israel.

We observed Cotton Sandstorm disrupting streaming television services in the UAE and elsewhere in December 2023 under the guise of a persona called "For Humanity." For Humanity published videos on Telegram showing the group hacking into three online streaming services and replacing several news channels with a fake transmission featuring a likely AI-generated anchor that claimed to show images of Palestinians injured and killed by Israeli military operations.

News outlets and viewers in the UAE, Canada, and the UK reported disruptions in streaming television programming, including BBC, that matched For Humanity's claims.[10] In April 2024, amid Iran's airstrikes on Israel, a new Iranian cyber persona, "Mostaghemoun," posted threatening messages in Hebrew, English, and Farsi that included videos and images that were likely created with AI.[11]

Limiting foreign influence operations in the modern era

Influence operations have been used throughout history by both state and non-state actors to shape public opinion and achieve strategic goals. Because they are recognized tools of soft power, there are established boundaries for such activity under international law. The principle of nonintervention, for example, safeguards national autonomy and, in certain cases, prohibits direct interference in the external and internal affairs of sovereign states. Activity which covertly manipulates the economic or political systems of another country, for example, could cross that line.

However, existing limitations of foreign influence operations under international law are no longer sufficient in the modern era. The emergence of social media and the advancements in generative AI have significantly changed the landscape. Therefore, it is imperative to reassess the impact and boundaries of these activities. Similar to the norms established by the United Nations to restrict state-sponsored cyberattacks, there should be comparable norms to regulate foreign influence operations in the online space.

Montaghemoun* (meaning Avengers in Arabic) posted threatening messages in Hebrew, English, and Farsi across 10 social media accounts in the days leading up to the Iranian attack against Israel, including posting multiple threatening videos and images. Microsoft assesses were created with AI.

Microsoft recommends that states embrace the following limitations on foreign influence operations:

Limits on targets

- **Crisis/emergency scenarios:** In an emergency or crisis – including wildfires, floods, extreme weather events, and chemical/radiation spills – foreign influence operations should not seek to manipulate civilians with respect to the crisis. When lives are at stake, reliable information is critical for safety.
- **Emergency/humanitarian response organizations:** Undermining public trust in organizations involved in humanitarian or emergency response missions is unacceptable. Governments deliberately spreading or promoting misleading information about medical First Responders or humanitarian assistance efforts abroad should equally be prohibited.
- **Elections:** Covert interference in elections via foreign influence operations online must be prohibited. Such a commitment was already included in the 2018 Paris Call for Trust and Security in Cyberspace, which has the support of 80 national governments from around the world.
- **Vulnerable/marginalized communities:** States should refrain from foreign influence campaigns that advocate national, racial or religious hatred or which incite violence against protected groups, including racial and ethnic minorities and LGBTQ+ populations.

Limits on tools and techniques

- **Covert use of AI:** States should not secretly create or knowingly use synthetic audio, images or video content generated by AI, to covertly mislead or coerce citizens of other countries.
- **Theft/abuse of social media data:** States should refrain from stealing or misusing data on foreign citizens held by private companies for the purpose of developing covert influence operations targeting a foreign populace.

Links

Protecting the public from abusive AI-generated content | Jul 2024

AI jailbreaks: What they are and how they can be mitigated | Jun 2024

How Russia is trying to disrupt the 2024 Paris Olympic Games | Jun 2024

Russian US election interference targets support for Ukraine | Apr 2024

China tests US voter fault lines and ramps AI content to boost its interests | Apr 2024

Staying ahead of threat actors in the age of AI | Microsoft Security Blog | Feb 2024

AI for defense

Microsoft's significant investment in AI innovation is aimed at providing cybersecurity defenders with an asymmetric advantage over attackers in the realm of defense.

In our efforts, we prioritize cutting-edge research and the development of groundbreaking solutions like Copilot for Security. These solutions amplify defenders' efforts by optimizing resources and scaling cybersecurity endeavors. This is particularly crucial considering the significant shortage of skilled cybersecurity workers, which poses one of the biggest challenges in the field of cybersecurity.

Currently, cybersecurity teams operate at their limits, facing staffing constraints, escalating regulatory compliance demands, and an ever-growing number of increasingly sophisticated adversaries. However, the introduction of AI will change this workload, offering various benefits to both attackers and defenders.

For defenders, the "automated ingenuity" of generative AI can now be applied across the entire defense chain, from initial detection of anomalies to prompt triage and response. Beyond merely enhancing existing security operations centers (SOC), AI holds the potential to introduce entirely new methods of defense. For instance, it enables persistent systems that constantly monitor for vulnerabilities and promptly address any breaches. Additionally, AI streamlines the sharing of information among defenders, transforming it from a labor-intensive manual process into a continuous, automated one.

"AI holds the potential to be as much of a transformative technological revolution for human beings as things like electricity or modern computing, if not possibly more so; a tool that opens up benefits across the board, transforming zero-sum problems into non-zero-sum opportunities and creating massive net long-term gains for humanity.

But, as we've seen repeatedly throughout the course of history, when in the wrong hands, any sufficiently new and powerful tool that people are given can be used by those people to cause harm. The good news is that these same AI tools, when paired with creativity, innovation and diligence, can put those of us on the side of defense and security ahead of disruptive threat actors, and allow everyone a chance to fully realize the tremendous benefits that AI can bring."

Kevin Scott, Chief Technology Officer

Harnessing AI to detect cyberattacks

Our researchers are developing a novel AI approach to detect and disrupt cyberattacks and "endpoint stories."

Endpoint stories are narratives of endpoint activities generated from data collected from physical devices that connect to a network system. These include mobile devices, desktop computers, virtual machines, embedded devices and servers. The data source for these stories is Microsoft Defender for Endpoint (MDE).

Detecting hidden attacks with AI

Hands-on-keyboard (HOK) attacks, where cybercriminals directly interact with compromised systems, are a major concern for enterprises. These attacks are hard to detect because attackers often use common administrative tools and techniques to blend in with legitimate activities, and attackers are able to move through networks in real-time and respond to what they find in the environment. To detect these attacks, we use LLMs that are fine-tuned to analyze endpoint story narratives and identify anomalous or suspicious activities. These models can learn from the context and semantics of the stories and flag potential threats that might otherwise go unnoticed.

Disrupting attacks by combining endpoint detection and response with AI

Our AI models are integrated with MDE, a cloud-based security solution that provides comprehensive protection for endpoints. MDE collects and processes data from millions of devices and uses it to generate endpoint stories. AI models are then automatically invoked, and when a model detects a HOK attack, an alert is created in the MDE portal. Based on the AI decision, MDE can automatically[16] isolate an affected device, temporarily disable compromised user accounts, and take additional actions to disrupt the attack. This way, MDE can thwart the attack before it causes more harm.

Extending AI across cybersecurity

Our approach is not only effective for detecting HOK attacks, but also has wider implications for other areas of cybersecurity. Leveraging the understanding capabilities of LLMs, AI models such as ours can be used to analyze and find malicious activities using large and complex data sources such as network logs, email communications, web traffic, and social media. This can help us uncover hidden patterns, trends, and insights that can inform our security strategies and policies. We are also exploring the latest methods, such as leveraging the Phi family of models,[17] to improve our AI models for detection of attacks and suspicious activities.

AI's early impact on the security operations center (SOC)

Scale, efficiency, and speed are key components affecting defenders' ability to detect and respond to incidents. On average, it takes 277 days to identify and contain a breach, with 207 days for identification and 70 days for containment.[14] By leveraging AI, defenders can significantly reduce this lag.

Microsoft has invested heavily in AI to help SOCs upskill and operate at speeds beyond human capability to tackle threat actors. In a 2023 study we found that novice users were able to perform 26% faster and were 44% more accurate across all tasks when using Copilot for Security.

Examples

During advanced human-operated ransomware attacks, we have seen the time from initial pre-HOK (hands on keyboard) alert to the encryption event averaging a mere 16 hours, underscoring the importance of operating fast to remediate the actor from the network. As mentioned, prioritizing incidents is a significant challenge that impacts time to resolve/mitigate. AI security solutions provide more than just a graphical representation of events; they generate a comprehensive incident summary that allows SOC analysts to quickly understand the situation and identify human-operated ransomware targeting mission-critical devices and users, enabling swift and decisive action.

To address the incident, the analyst must dive into indicators of compromise. Using AI, the analyst can instead assess an encoded command line run on a suspicious device from the incident. What would have taken a junior analyst dozens of minutes and several tools can now be achieved at machine-speed.

Seven areas of efficiencies in Microsoft security operations

AI has demonstrated significant benefits to cybersecurity by enhancing threat detection, response, analysis, and prediction. AI can also be used for various other tasks within a security organization, which often involves processing large volumes of unstructured data to gain insights, answer questions, and make informed decisions. Microsoft is leveraging AI in seven key areas of security operations.

1. Triaging requests and tickets. Teams in a security organization receive large volumes of requests and tickets. Depending on the complexity of the logic that determines how these items are dispositioned, large language models (LLMs) can speed up the triage process and increase the efficiency and effectiveness of responding teams. LLMs can use the specifics of a new request and, comparing them to how similar requests were dispositioned in the past, decide what to do. LLMs can additionally use relevant policies, controls, and other material to inform these decisions. At Microsoft, one of our internal response teams receives on average 25 requests each week. This volume is expected to double over the next six months. Without LLMs, initial triage of a request takes approximately three hours. The team developed an LLM solution, which takes seconds to recommend response actions based on information provided in the requests and guidelines on when each action is appropriate. The LLMs can also generate follow-up questions if the information in the request is insufficient to recommend an action. The use of LLMs in this scenario is estimated to save at least 20 hours per person, per week.

2. Prioritizing work items. Keeping an organization secure and compliant involves a constant stream of work items of varying importance and time-criticality. AI can assess the priority of a given item based on how similar items were prioritized in the past. As with the previous use case, LLMs can use relevant policies, procedures, and other material to determine these priorities. Additionally, AI can ensure that the prioritization criteria are up to date with the ever-evolving compliance requirements where hundreds of regulatory changes happen on a daily basis.

3. Knowledge gathering from diverse external sources. Augmenting proprietary in-house datasets with online content (such as threat intelligence and information on recent vulnerabilities) enables an organization to make better decisions. AI can scrape online content and extract security-related information at scale. At Microsoft, one of our internal teams identifies and processes 50 articles per week. While this used to take two hours per article on average, using AI, the team is now able to generate concise reports from these articles in minutes.

AI is not only useful for first line of defense operations, but its capability to transform behind-the-scenes daily processes is also significant and promising. Modernizing these processes is essential for scaling up security operations and making the best use of human expertise.

One notable example is the use of AI for triaging requests, which is saving at least 20 hours per week per person on one of our internal response teams.

4 **Knowledge retrieval.** A large part of keeping an organization secure depends on how well-informed its employees are on security policies, best practices, and the remediation actions necessary for compliance. However, this information is usually fragmented across multiple locations, forcing an employee to search for and extrapolate it. LLMs can greatly improve this experience and generate complete and accurate answers, even allowing the user to ask follow-up questions. If integrated with an organization's data on devices and services, the answers can be tailored to a specific situation.

5 **Risk assessment.** AI can assimilate information from diverse sources, whether proprietary or publicly available, to bear on the risk of a given entity, service, account, etc. AI can leverage unstructured organizational knowledge and historical precedents to enrich the set of factors determining risk.

6 **Learning from the past.** Security operations constantly generate large volumes of diverse artifacts (tickets, reports, playbooks). Looking at the evolution of this data over time can provide valuable insights into themes, anomalies and recurring issues. Much of this historical content is unstructured and impractical to manually sift through. LLMs can ingest data pertaining to previous incidents, violations, remediations and other events to uncover valuable learnings that help the organization get a comprehensive view of past events. For example, analyzing historical data from post-incident reviews can answer questions like: 1) What were the main themes in past incidents? 2) For a given theme, did the associated incidents happen over a large span of time (indicating an unaddressed root cause) or did they happen and then stop (indicating successful remediation)? 3) Have we historically seen anything similar to a new incident?

7 **Reporting.** As a security organization's size grows, so do reporting needs. AI can help combine, consolidate, and distill artifacts such as documents and slides into reports whose content, level of detail, tone, and length can be adjusted depending on the audience and the report goal.

Using generative AI to understand cyberattacks and create tailored mitigations

As discussed throughout this report, the frequency and severity of cyberattacks have increased significantly in recent years. Addressing large volumes of attacks requires automation engines beyond the current rules-based approach. But volume isn't the only thing changing.

There is also a huge growth in the types and complexity of attacks. Microsoft Defender for Endpoint has seen a significant increase in the number of indicators of attack (IOAs) from January 2020 to today, there has been a 79% growth in IOAs.

79%

growth in number of indicators of attack since 2020

Growth in complexity of MITRE ATT&CK tactics and techniques

Today's challenges require a way to process alerts, precisely yet practically without resorting to defined and nuanced differentiated treatment to numerous of types of attacks.

Source: MITRE

The growth and increasing complexity of attacks is also evident in the evolution of the MITRE ATT&CK framework. The changing nature of attacker TTPs compounds the difficulty defenders face in confronting and remediating attacks. In 2015, it was possible to bucket attacks into nine tactics and 96 techniques and differentiate their treatment with rules. Today, the diversity of TTPs requires hundreds of differentiated rules and nuanced treatment, making it harder than ever to alleviate the volume of incidents by automation alone.

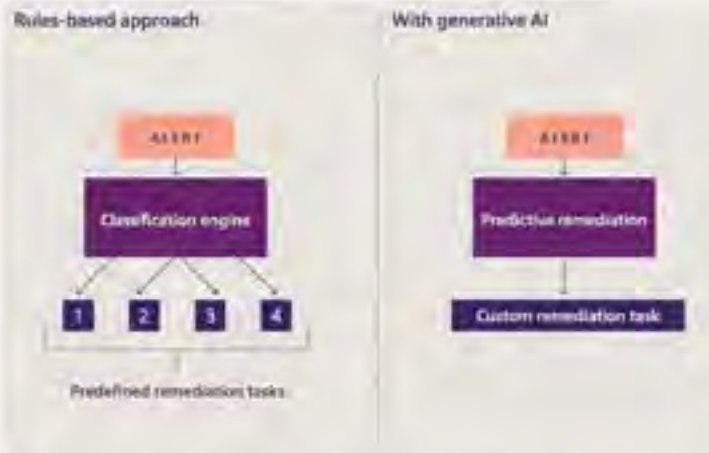
From categories to context

Generative AI allows defenders to use the narrative context of the threat as a qualifier to defensive actions and remediation. Instead of classifying an alert into a known set of categories, the differentiation is now built from all surrounding contextual information, with remediation dependent on the factual findings and not by abstraction into a bucket (categorization).

The technical difference is in moving from classification, which is a methodology that abstracts similar attacks, to a high-dimensional proximity engine, where the remediation is the statistically best next step. The outcome of this method is very specific to the collection of all entities and facts of the event. This means all the nuances of an event are handled and considered without loss of resolution, which happens when an event is classified into a bucket.

A good example of the challenges organizations are facing and how generative AI could help them is in "User Submitted Phish." Security operations centers struggle with a high volume of "user-submitted phish," alerts that are based on users reporting emails they suspect of being phishing attempts. An analyst's attention is needed to determine if the email is indeed malicious but due to the volume of alerts—which can total hundreds or even thousands of emails in a month—some companies use a service to prequalify the user-submitted phish report before an analyst investigation.

The decision whether an email is malicious is not easily discerned since many factors come into play with variability in what makes it a phishing attempt. Since it is not easy to categorize an email as a phishing attempt, a method that looks at the specifics of the email without a predetermined rule is more advantageous in determining a verdict.



A rules-based approach limits your remediation options to predefined tasks. With generative AI, alert treatment is generalized to produce a unique remediation that is the predictive next step of the specific facts of the alert. Each remediation will be unique to the facts of the alert, and it does not rely on a predetermined classification of remediations.

Additional automated enrichments can then be provided including:

- Finding a domain's reputation.
- Associating domain, sender, and return path to threat articles.
- Checking the same email subject line sent to other users across the organization.

With all added enrichments, we can then leverage generative AI to help with a summary, a verdict recommendation, and a containment plan.

Typical prompting examples include:

- "Triage the following email and point out what you find suspicious? Investigate the Message-ID for any inconsistencies or signs of spoofing. I'm specifically interested in a sense of urgency, generic greetings, spelling or grammar mistakes, requests for personal information..."
- "Based on the above email investigation, summarize the investigation steps that were taken and provide supporting evidence on the percentage of certainty that this is a true positive phishing incident."
- "Based on your investigation, create a containment plan."

The culmination of these elements, without a pre-determined rule, is the foundation for a factual verdict and the certainty around it. This is where generative AI comes into play, as the True Positive/False Positive decision is reflective of the specific set of contexts and findings for the email. One may argue that task-based AI such as Bayesian tools could achieve the same result. However, based on current research our hypothesis is that generative AI will offer more flexibility to manage the ever-growing variety of events and cases, while Bayesian tools have a narrower scope and diminished flexibility.

A high volume of alerts, which also contain false positives, forces SOC analysts to focus on reactive tasks and takes focus away from proactive efforts to improve security posture, which would result in fewer alerts. Previously, this loop was hard to break. However, the advent of a new AI-based methodology to apply to incoming volume and pre-qualify which alerts may need an investigation and which are not likely to require one is a positive step forward that will allow SOC teams to allocate more time to proactive tasks.

How governments and industries are advancing global AI security

AI is not new in the cybersecurity field. For many years it has been used to detect malware by using ML, but recent breakthrough advancements are now changing the depth and breadth of its impact.

We are now facing key questions such as: how to harness the power of AI to turbocharge our cybersecurity defenses while deterring adversaries from exploiting it for malicious cyber activity? Or how do we protect AI models against cyber threat actors?

Governments worldwide have recognized that AI offers both benefits and risks for society. As they pursue AI regulatory approaches that seek to balance those benefits and risks, their efforts vary in scope and scale. These differences among governments' policy initiatives are not surprising; they reflect the core values of the governments' leadership, the countries' legal and constitutional frameworks, and the state of the technology industry and its potential for future growth. Despite these differences, safety and security are emerging as core principles pursued by the majority of governments as they encourage the safe and responsible development, deployment, and use of AI.

Government approaches to AI security

While there is a consensus on the importance of security in the development of AI, governments have pursued different approaches in implementing security requirements.

These approaches mainly focus on the secure design, development, deployment, and operation of AI. Examples of security measures that target the secure design and deployment of AI systems include preliminary risk assessments to identify potential vulnerabilities and to design mitigation measures, adversarial testing such as red teaming to address unidentified vulnerabilities, and data management systems to guarantee quality and trusted data.

Security measures that target the secure deployment and operation of AI systems include mechanisms to protect them against misuse by users or third-party attackers, such ongoing auditing and monitoring mechanisms, incident reporting, and automatic record-keeping systems.

The United States

The 2023 Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence[60] directs US federal agencies to implement the policies set forth in the E.O, including taking a series of actions focused on safety and security of AI technology. The US approach is notable in two ways: first, it imposed mandatory cybersecurity measures on federal agency use of AI without extending them to the private sector. Second, it leverages government action to enhance AI capabilities for cyber defense. For example, EO 14110 directs the Department of Defense (DOD) and Department of Homeland Security (DHS) to plan and conduct pilot projects for how AI capabilities can aid in the discovery and remediation of vulnerabilities in critical US Government software, systems, and networks. The DOD has been tapped to spearhead actions for national security systems, while the DHS will spearhead actions for US Government civilian systems.

Furthermore, the FY24 National Defense Authorization Act (NDAA)[61] includes several provisions designed to strengthen the DOD's use of AI in its defense operations. Under the NDAA, the DOD must: develop a bug bounty program for foundation AI models being integrated into the "missions and operations" of the Department to strengthen cyber defense resiliency; establish a prize competition designed to evaluate technology for generative AI detection and watermarking to support the DOD's warfighting requirements; establish and review guidance around the Department's near-term and long-strategies for the adoption and use of AI; and assess the potential vulnerabilities of AI-enabled military applications, including assessments of research and development efforts needed to advance AI-enabled military applications. The US Government administration has also announced it will release a National Security Memorandum (NSM)[62] that addresses the regulation of AI systems for national security, military, and intelligence purposes.

The European Union

The EU's Artificial Intelligence Act (AI Act), the first ever horizontal legal framework on AI, requires providers of high-risk AI systems and general-purpose AI (GPAI) models with systemic risk, to implement security measures. The AI Act requires providers of high-risk AI systems to ensure that such systems achieve an appropriate level of accuracy, robustness, and cybersecurity, and perform consistently in those respects throughout their lifecycle. Providers of GPAI models with systemic risk are required to ensure an adequate level of cybersecurity protection of the model, as well as its physical infrastructure. The AI Act also requires providers (and deployers in some cases) of high-risk AI systems, and providers of GPAI models with systemic risk, to report serious incidents to relevant governmental authorities as well as relevant actors in the AI value chain.

Other legislative initiatives

Brazil and Costa Rica have proposed legislation that would impose on all AI systems certain security requirements (for example, parameters for separating and organizing training data; information security measures; human rights impact assessments), with additional requirements for high-risk systems. Meanwhile, China has adopted the most stringent approach imposing security requirements on all covered AI systems. These requirements include technology ethics reviews; user registration and verification; measures to counter telecommunication network fraud; and the use of accurate and lawful training data.

Finally, other countries have published voluntary guidelines and codes of conduct that suggest security measures for private sector entities. For example, under the UK National Cyber Security Centre (NCSC)'s guidelines, companies should consider complying with measures such as identification of threats and risks; acquisition of well-secured and well-documented hardware and software; and documentation of models and datasets. Canada, Japan, and Singapore have published similar codes of conduct.

Cyber Point of View: Albania

Transparency, advanced technologies, and generative AI to combat malicious state-sponsored cyberattacks

In July 2022, Iran launched a devastating cyberattack designed to cripple Albania's digital infrastructure. The National Agency for Information Society (AKSHI), responsible for managing approximately 95% of the government's digital services, was the biggest target.

In response, AKSHI acted decisively, stopping the attackers from causing additional damage and embarking on a journey to invest in enhancing its cybersecurity maturity. AKSHI partnered with diplomatic partners and industry leaders to gather intelligence and implement cutting-edge technologies and innovative strategies to protect its digital assets from ongoing and continuous attacks.

One of the key components in AKSHI's success was transparency throughout the process and the swift adoption of automation, advanced technologies, and generative AI. This enabled AKSHI to fortify its defenses, boost cybersecurity resilience, and detect and respond to cyber threats more effectively. By providing real-time insights and predictive analytics, AKSHI was able to stay ahead of the attackers.

This success story is a testament to the transparency, resilience, and determination of AKSHI's leaders and cybersecurity professionals. By turning a crisis into an opportunity for growth and innovation, AKSHI has set a new standard for cybersecurity excellence and is now taking full advantage of generative AI capabilities to enhance its cybersecurity infrastructure and improve services for Albania's citizens.

Collaborative policy initiatives for AI security

Organizations around the world are collaborating to advance government policy initiatives on enhanced AI security.

July 2023

- Microsoft, Anthropic, Google and OpenAI launched Frontier Model Forum, an industry body focused on ensuring safe and responsible development of frontier AI models.[63]

August 2023

- The White House announces the AI Cyber Challenge, for cybersecurity researchers to spur the use of AI to identify and fix software vulnerabilities.[64] Microsoft committed to host competition on Microsoft Azure.

November 2023

- The UK launched the world's first safety institute to spur collaboration on AI's safety with leading AI companies and nations.[65]
- The US Department of Commerce, through National Institute of Standards and Technology (NIST) announced the US Artificial Intelligence Safety Institute (USAISI) to lead the US Government's efforts on AI safety and trust, including working with partners in academia, industry, government, and civil society to advance AI safety.[66]
- The Bletchley Agreement for collaboration resulted from an AI Safety Summit convened by the UK and including the US, EU, and China, likeminded AI companies, and 28 country delegations.[67]
- Microsoft contributed to the development of secure AI system guidelines alongside the UK National Cyber Security Centre (NCSC), and the US Cybersecurity and Infrastructure Security Agency (CISA),[68] among others. It was co-sealed by 23 domestic and international cybersecurity organizations. This publication marked a significant step in addressing the intersection of AI, cybersecurity, and critical infrastructure.

January 2024

- CISA's cross-sector analysis of sector-specific AI risk assessments completed by sector risk management agencies. Microsoft provided recommendations through the IT Sector Coordinating Council - a public private partnership for collaboration between IT sector and the Department of Homeland Security (DHS).

February 2024

- The Japanese government launched a new AI Safety Institute within the Information-technology Promotion Agency (IPA) in collaboration with relevant ministries and agencies.[69] The Institute aims to examine evaluation methods and standards related to AI. Japan plans to collaborate with the UK and the US.

March 2024

- The US Department of Treasury released a report on the current state of AI-related cybersecurity and fraud risks in financial services, including an overview of current AI use cases, trends of threats and risks, best-practice recommendations, and challenges and opportunities.[70]

April 2024

- In April 2024, building on the NCSC secure AI development guidelines release in 2023, the US National Security Agency's Artificial Intelligence Security Center published the joint Cybersecurity Information Sheet Deploying AI Systems Securely[71] in collaboration with CISA, the US Federal Bureau of Investigation, the Australian Signals Directorate's Australian Cyber Security Centre, the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre, and the United Kingdom's National Cyber Security Centre.
- The US Department of Homeland Security (DHS) released Safety and Security Guidelines for Critical Infrastructure Owners and Operators.[72] Microsoft contributed to the cross-sector risk assessments that informed the DHS guidance.
- Microsoft joined the DHS AI Safety and Security Board (AISSB).[73] The AISSB advises the DHS Secretary, the critical infrastructure community, other private sector stakeholders, and the broader public on the safe, secure, and responsible development and deployment of AI technology in our nation's critical infrastructure.

May 2024

- The second global AI summit secured safety commitments from companies. It is a new agreement[74] between 10 countries and the EU to establish an international network similar to the UK's AI Safety Institute,[75] the world's first publicly backed organization to accelerate the advancement of AI safety science. The network will promote a common understanding of AI safety and align its work with research, standards, and testing. Australia, Canada, the EU, France, Germany, Italy, Japan, Singapore, South Korea, the UK, and the US have signed the agreement.[76]
- Microsoft released a blueprint for mutual prosperity through AI governance in Korea.[77]

June 2024

- Microsoft funded the Securing Critical Infrastructure in the Age of AI workshop led by Georgetown University's Center for Security and Emerging Tech (CSET). CSET will publish a report based on findings from the workshop offering policy recommendations for AI security in critical infrastructure. Expected publication date: September 2024.
- Microsoft hosted and participated in the first federal AI security tabletop exercise led by CISA JCDC.AI,[78] convening more than 50 AI experts from US and international agencies and industry partners focused on effective and coordinated responses to AI security incidents.

International standards for AI security

Security vulnerabilities and risks arising from adversarial manipulation of AI systems can be exploited and impact everything from confidentiality to human safety. Therefore, standards are becoming essential to improve awareness and understanding of AI, address regulatory concerns and requirements, and extend good practice and consistency across the industry. Standards can also help build trust and confidence in AI systems among stakeholders such as users, customers, regulators, and society at large.

The benefits of international standards

The AI regulatory landscape is evolving almost as fast as AI itself. Just as there is a demand for regional and national standards, there are also many benefits to international standards.

International standards can mitigate fragmentation and ensure more consistency, good practice, controls, and even conformity assessment, especially where supply chains, threat actors, and applications are of a global nature. International standards can also help to facilitate cooperation, innovation, and competition.

ISO/IEC 42001

Under ISO/IEC 42001, organizations are guided in establishing continually improving risk-based processes to support responsible use of AI throughout the AI system lifecycle.

There are also crosswalks[79] available to map the NIST AI Risk Management Framework. Many responsible AI practices were born out of information security practices. Responsible AI red teaming is one such practice, where real-world adversarial behaviors are emulated in an attempt to expose AI system vulnerabilities which can lead to harmful outputs, especially through prompt injection attacks.

The requirements in ISO/IEC 42001 are intended to be auditable to achieve certification including helping to manage responsible AI across supply chains as well as provide a foundation that can help with regulatory compliance.

ISO/IEC 27090

ISO/IEC 27090 is being developed to provide guidance for addressing security threats to AI systems. The standard aims to help organizations better understand the consequences of security threats specific to AI systems throughout their lifecycle, such as evasion attacks, data poisoning, model stealing, and membership inference attacks. The document also describes how to detect and mitigate such threats. ISO/IEC 27090 starts with the premise that AI systems are information systems. Therefore, conventional cybersecurity measures—including those in international standards such as ISO/IEC 27002 information security controls, and zero trust principles—are the foundation to mitigating security risks to AI systems and for securing the datasets associated with AI systems.

Actionable Insights

1. Existing cybersecurity standards provide good practice to secure all types of information systems, including AI systems, throughout their lifecycle. As controls to address risks specific to AI systems mature, new standards will be developed. A multi-stakeholder approach is essential for the development of pragmatic and useful standards to help all types of organizations to manage security.
2. Security underpins a responsible AI approach; international standards can be used to demonstrate an overall responsible AI approach, accountability, and effective mitigation against harm and safety risks.
3. International standards can help mitigate fragmentation, ensure consistent practices globally, and facilitate trust and cooperation. International standards continue to uphold the accountability of trust even while regional standards are in demand to support regulatory frameworks.

Staying a step ahead of threat actors in the age of AI

Our experts and automated systems analyze and correlate across the thousands of threat actors we track, uncovering efforts to evade detection or expand their capabilities by leveraging new technologies like AI.

In February, Microsoft and OpenAI released publications[89] discussing the emergence of nation-state threat actors utilising AI for malicious purposes. Microsoft also released a set of policy principles to mitigate the risks associated with the use of AI tools and application programming interfaces (API) by nation-state advanced persistent threats (APT), advanced persistent manipulators (APM), and cybercriminal syndicates.

To stay ahead of threat actors in the age of AI, Microsoft's policy follows the principles below:

- **Identification and action against malicious use of Microsoft AI:** Upon detection of the use of any Microsoft AI APIs, services, or systems by an identified malicious actor, Microsoft will take appropriate action to disrupt their activities, for example by disabling the accounts used, terminating services, or limiting access to resources.
- **Notification to other AI service providers:** When we detect a threat actor's use of another service provider's AI, AI APIs, services, and/or systems, Microsoft will notify the service provider and share relevant data. This enables the provider to independently verify our findings and take action in accordance with their own policies.
- **Collaboration with other stakeholders:** Microsoft will collaborate with other industry and civil society stakeholders to regularly exchange information about threat actors' use of AI. This collaboration aims to promote collective, consistent, and effective responses to ecosystem-wide risks.
- **Transparency:** Microsoft will inform the public and stakeholders about threat activity, including the nature and extent of threat actors' use of AI detected by our systems and the measures taken against them, as appropriate.

These principles reflect Microsoft's commitment to prioritizing security and responsible AI innovation, which includes the safety and integrity of our technologies with respect for human rights and ethical standards.

These principles build on our Responsible AI practices, our commitments to advance responsible AI innovation, and the Azure OpenAI Code of Conduct. We also follow these principles as part of our broader commitments to strengthening international law and norms and to advance the goals of the Bletchley Declaration.

Links

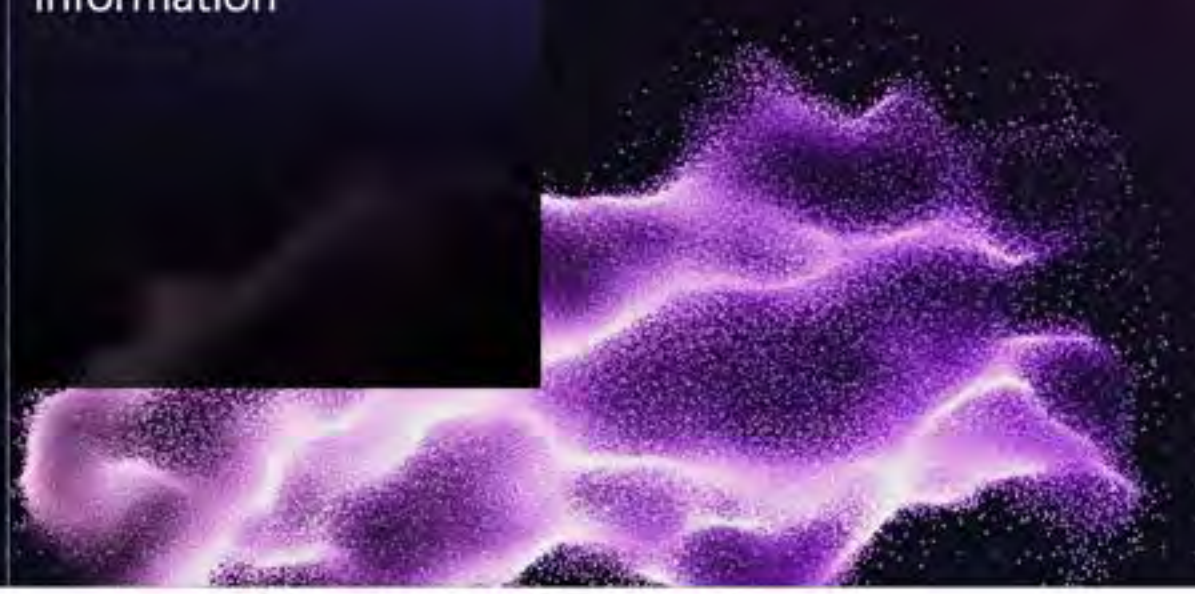
Staying ahead of threat actors in the age of AI | Microsoft Security Blog | Feb 2024

Global Governance: Goals and Lessons for AI | Sep 2024

Appendix

Additional information

References 108
Contributing teams 110



References

Overview

1. Expanding Microsoft's Secure Future Initiative (SFI) | Microsoft Security Blog | May 2024

Chapter 1. The evolving cyber threat landscape

2. National Security Strategy of Japan | Dec 2022
3. Publications | Japan Ministry of Defense
4. ISMAP Overview
5. Exclusive: UN experts investigate 58 cyberattacks worth $3 bln by North Korea | Reuters
6. Half of North Korean missile program funded by cyberattacks and crypto theft, White House says | CNN Politics
7. Microsoft Digital Defense Report 2022
8. IRGC-Affiliated Cyber Actors Exploit PLCs in Multiple Sectors, Including U.S. Water and Wastewater Systems Facilities | CISA
9. Telegram channel "KARMA" — @karmabelow80 statistics — TGStat
10. https://darktrace.com/blog/amadey-info-stealer-exploiting-n-day-vulnerabilities
11. A report on NOBELIUM's unprecedented nation-state attack | Microsoft Security Blog
12. Exposed and vulnerable: Recent attacks highlight critical need to protect internet-exposed OT devices | Microsoft Security Blog
13. Microsoft Content Integrity tools are available in private preview for political campaigns and newsrooms to provide more transparency into who created the image, whether it's AI generated, the publisher, when and where the image was created, and whether the image has been edited.
14. Breeker US V10 16x9 VO2 QR 2 (youtube.com)
15. Source: The original YouTube account has since been removed: youtube.com/@truetjl
16. https://archive.is/H1HgA (Taiwan)
17. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/11/MTAC-Report-2024-Election-Threat-Assessment-11082023-2-1.pdf
18. https://www.youtube.com/watch?v=kbLBJb3UpYQ;

    https://web.archive.org/web/20240423173006/

    https://sanfranchron.com/2024/04/21/17/web.archive.org/web/20240423034950/

    https://bostontimes.org/2024/04/21/do-everything-to-prevent-donald-trump-from-winning-the-elections-leak-from-ukrainian-troll-factory/
19. https://t.me/DonaldJTrump29/1612
20. China tests US voter fault lines and ramps AI content to boost its geopolitical interests - Microsoft On the Issues)
21. Source: globaltimes.cn/page/202407/1315977.shtml
22. Technology boosting global financial crime, INTERPOL warns | World Economic Forum (weforum.org)
23. KPMG 2022 Fraud Outlook Survey - KPMG Global
24. https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf
25. Microsoft, I4C block 1,000 Skype accounts tied to cyber criminals (msn.com)
26. Indian call centre fraudster is jailed for 28 months for Microsoft scam | Daily Mail Online
27. Amazon, Microsoft, and India crack down on tech support scams – The Verge
28. Losses from Online Payment Fraud to Exceed $362 Billion Globally Over Next 5 Years | Press (juniperresearch.com)
29. https://usa.visa.com/partner-with-us/payment-technology/visa-tokenization.html
30. Top 15 Phishing Stats to Know in 2024 | Trend Micro News
31. The Coalition for Content Provenance and Authenticity (C2PA), the global standards body responsible for Content Credentials, continues to gain momentum with over 150 members, adding Google, OpenAI, and many others in 2024.
32. Account Takeover Incidents are Rising: How to Protect Yourself in 2024 | Security.org
33. How effective is multifactor authentication at deterring cyberattacks? - Microsoft Research
34. Identity Reveal: The Threat Actor Behind ONNX Store and Caffeine Phishing Kit | Blog | Dark Atlas | Dark Web Monitoring Platform | Compromised Credentials Monitoring | Account Takeover Prevention Platform | Threat Intelligence | Buguard
35. Cybersecurity Threats in Online Gaming: Learnings for India (orfonline.org)

References continued

Chapter 2. Centering our organizations on security

36. Passkeys (Passkey Authentication) (fidoalliance.org)
37. FIDO Alliance – Open Authentication Standards More Secure than Passwords
38. The Accra Call for Cyber Resilient Development - GC3B – Global Conference On Cyber Capacity BuildingGC3B – Global Conference On Cyber Capacity Building
39. Election Security Advisors One-Pager (microsoft.com)
40. AI Elections accord - A Tech accord to Combat Deceptive Use of AI in 2024 Elections
41. Meeting the moment: combating AI deepfakes in elections through today's new tech accord - Microsoft On the Issues
42. Microsoft announces new steps to help protect elections - Microsoft On the Issues
43. Microsoft 365 for Campaigns
44. Keeping your vote safe and secure: A story from inside the 2020 election – On the Issues (microsoft.com)
45. Microsoft AccountGuard
46. Election Security Advisors One-Pager (microsoft.com)
47. Microsoft and OpenAI launch Societal Resilience Fund - Microsoft On the Issues
48. Expanding our Content Integrity tools to support global elections - Microsoft On the Issues

Chapter 3. Early insights: AI's impact on cybersecurity

49. Completely Automated Public Turing test to tell Computers and Humans Apart
50. Internet of Things Market Overview 2024-2028, Internet Of Things Industry 2024 (reportlinker.com)
51. Slovakia's Election Deepfakes Show AI Is a Danger to Democracy | WIRED
52. FRANCE 24 journalist impersonated in new deepfake video - Truth or Fake
53. Dmitry Medvedev on X: "Macron seems to have been so scared of a real, or presumed assassination in nazi Kiev that not only has he cancelled his trip there, but also decided to share the nuclear capacity with other Europeans. Sure, such trifles as the Nuclear Non-Proliferation Treaty are of no concern" / X
54. UAE: Cyberattack disrupts TV services, rattles some residents with graphic content from Gaza - News | Khaleej Times

    UAE: A Cyberattack Imitates TV Services And Unnerves Some Locals With Explicit Material From Gaza - The Emirates Times

    Jadoo tv hacked by "For humanity 2023" (youtube.com)
55. twitter.com/montaghemoun/status/1778585175552561344
56. Automatic attack disruption in Microsoft Defender XDR - Microsoft Defender XDR | Microsoft Learn
57. Introducing Phi-3: Redefining what's possible with SLMs | Microsoft Azure Blog
58. Randomized Controlled Trials for Microsoft Copilot for Security by Benjamin G. Edelman, James Bono, Sida Peng, Roberto Rodriguez, Sandra Ho :: SSRN
59. Updates | MITRE ATT&CK®
60. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence | The White House
61. H.R.2670 - National Defense Authorization Act for Fiscal Year 2024
62. USGA announces a national security memorandum
63. Microsoft, Anthropic, Google, and OpenAI launch
64. White House launches AI cyber challenge to identify and fix open-source software vulnerabilities | FedScoop
65. Prime Minister launches new AI Safety Institute - GOV.UK (www.gov.uk)
66. At the Direction of President Biden, Department of Commerce to Establish U.S. Artificial Intelligence Safety Institute to Lead Efforts on AI Safety | U.S. Department of Commerce
67. Countries agree to safe and responsible development of frontier AI in landmark Bletchley Declaration - GOV.UK (www.gov.uk)
68. CISA and UK NCSC Unveil Joint Guidelines for Secure AI System Development | CISA
69. Launch of AI Safety Institute (meti.go.jp)
70. U.S. Department of the Treasury Releases Report on Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Sector | U.S. Department of the Treasury
71. CSI-DEPLOYING-AI-SYSTEMS-SECURELY.PDF (defense.gov)
72. Safety and Security Guidelines for Critical Infrastructure Owners and Operators | Homeland Security (dhs.gov)
73. Artificial Intelligence Safety and Security Board | Homeland Security (dhs.gov)
74. Seoul Declaration for safe, innovative and inclusive AI: AI Seoul Summit 2024 - GOV.UK (www.gov.uk)
75. The AI Safety Institute (AISI)
76. In Seoul summit, heads of states and companies commit to AI safety | TechCrunch
77. New Digital Order: A blueprint for mutual prosperity through AI governance in Korea - Microsoft Stories Asia
78. CISA, JCDC, Government and Industry Partners Conduct AI Tabletop Exercise | CISA
79. NIST AIRC - Crosswalk Documents
80. Staying ahead of threat actors in the age of AI | Microsoft Security Blog ; Disrupting malicious uses of AI by state-affiliated threat actors | OpenAI

Contributing teams

The Microsoft Digital Defense Report (MDDR) has been a collaborative effort. The data and insights it pulls together have been compiled by a diverse group of security-focused professionals across various Microsoft teams. Their common goal is to protect Microsoft, its customers, and the world from the threat of cyberattacks, and we are proud to share what we found as we work towards building a safer environment for everyone.

AI for Good Research Lab is a philanthropic, applied research and data visualization lab that is committed to leveraging the transformative power of AI to address some of the world's most pressing challenges. In collaboration with subject matter experts in academia, NGOs, and all levels of government, the Lab leverages Microsoft's cloud technology and data science talent to create solutions across many disciplines and around the world.

AI Safety and Security is responsible for all aspects of AI safety, including pre-launch evaluation, incident response, building safety infrastructure, training, research, and policy.

Azure DDoS Protection is responsible for safeguarding Microsoft's cloud infrastructure from distributed denial of service (DDoS) attacks. The team develops and maintains advanced network security solutions to detect, mitigate, and prevent DDoS threats, ensuring high availability and reliability for Azure services and customers' applications by minimizing the impact of malicious traffic.

Azure Edge + Platform is responsible for Microsoft's operating systems, IoT and edge products, engineering systems, and health platforms from the chip level to the cloud. E+P is the platform team for the company and the foundation upon which virtually every Microsoft product and service is built.

C+E Governance leads and manages compliance and regulatory programs and initiatives for the C+E organization, including payments compliance. The Commerce Risk Engineering Team harnesses cutting-edge AI, strategic risk containment solutions and engineering excellence to safeguard transactions across all of Microsoft and Xbox.

Central Fraud and Abuse Risk detects and responds to Nation-state actors, criminal syndicates, and common hackers who wish to cause financial and reputational harm to Microsoft, its customers, and partners. To make the world safer for all, the team also partners with law enforcement, industry affiliates, and customers to share fraud insights.

Cloud Ecosystem Security is responsible for the core cloud security platform, data security, compliance, governance and privacy. The team also leads AI-powered threat and data intelligence, as well as AI security research and development.

Core Datacenter Services is responsible for global availability by implementing global standard processes and delivering programs that maximize efficiency while optimizing safety, security, availability across our global datacenter portfolio.

Corporate Standards Group represents Microsoft in multistakeholder organizations that are establishing standards on issues such as cybersecurity, artificial intelligence, and data. The team works with governments, civil society, academia, and industry to create coherent international practices that can be used to develop, evaluate, and manage trustworthy technology.

Critical Infrastructure Networking & Cyber Defense is a global organization that provides safe, reliable connectivity and protection for operational technology assets required for Microsoft data center operations.

Customer Experience Engineering (CxE) drives better security outcomes by engaging directly with customers throughout the product development process. By incorporating real-world feedback, CxE ensures that Microsoft Security products are tailored to meet customer needs and deliver enhanced satisfaction.

Contributing teams continued

Customer Security and Trust drives continuous improvement of customer security in Microsoft products and online services. Working with engineering and security teams across the company, the team ensures compliance, enhances security, and drives transparency to protect customers and the global ecosystem.

Customer Success security teams collaborate with customers to accelerate their security transformation and modernization by sharing best practices, lessons learned, and expert guidance.

Data Intelligence collaborates with partners in the security organization to enhance the efficiency and effectiveness of processes related to risk and resilience, findings analysis, standards and compliance and device security, among others. The team uses machine learning and Generative AI to learn from structured and unstructured data.

Data Security & Privacy provides comprehensive solutions that empower customers to protect, govern, understand, and manage their enterprise data across the Microsoft cloud – and beyond.

Democracy Forward works to preserve, protect, and advance the fundamentals of democracy by safeguarding open and secure democratic processes, promoting a healthy information ecosystem, and advocating for corporate civic responsibility.

Digital Crimes Unit is an international team of technical, legal, and business experts that has been fighting cybercrime, protecting individuals and organizations, and safeguarding the integrity of Microsoft services since 2008, through strategic partnerships and engagements, the seizure of criminal infrastructure, and the disruption of global cyber threats and criminal networks.

Digital Diplomacy is an international team of former diplomats, policy makers, and legal experts working to advance a peaceful, stable, and secure cyberspace in the face of rising nation-state conflict.

Digital Security & Resilience is the organization led by our Microsoft CISO, and is dedicated to enabling Microsoft to build the most trusted devices and services, while keeping our company and customers protected.

Enterprise & Security provides platform technologies and solutions to manage and harden platforms against attacks. The team also empowers company-wide security initiatives in Zero Trust, secure identity, secure devices, secure supply chain, and scale management from cloud.

European Government Affairs represents Microsoft's positions towards European political institutions, governments and other political actors. The team oversees a large variety of digital policies across Europe, including AI, Cloud, Sustainability and Cybersecurity policy.

Extended Security Posture Management builds cross-domain pre-breach security solutions for attack surface management and threat exposure reduction. The team brings together posture management capabilities for devices, identities, cloud, and applications into a set of consolidated products serving security leaders and their teams.

Global Cybersecurity Policy team focuses on developing and advancing public policy that strengthens customer and ecosystem-wide cybersecurity and resiliency at the intersection of geopolitics and emerging technologies.

Global Hunting Oversight and Strategic Triage identifies threat actor victims across the Microsoft Ecosystem, orchestrates rapid, effective, and iterative improvements to reduce attack surface, and develops automated, repeatable solutions to security and analysis problems.

Identity & Network Access teams innovate and build solutions that manage and govern identities and access, including the consumer sign-in experience.

Insights, Data Engineering, and Analytics Momentum and Storytelling curates metrics used in non-financial public disclosures; helps craft the messages around those metrics, and ensures that the messages align with Microsoft's perspectives.

Contributing teams (continued)

Microsoft Counterintelligence Program is a team that assesses threat and vulnerability information to inform leadership and formulate mitigation strategies to predict, deter, and investigate threat activity directed against Microsoft. The team also advises on how to improve related security and business practices to minimize or prevent exploitable vulnerabilities.

Microsoft Defender Experts is a managed Threat Hunting and Extended Detection and Response service that proactively looks for threats 24/7/365 using Microsoft Defender data.

Microsoft Incident Response (Detection and Response Team) is an organization of security experts with deep technical and industry skills who provide incident hunting, cyber resilience and threat intelligence services to customers. Microsoft Incident Response maintains strategic partnerships with security organizations, governments, and many internal Microsoft groups.

Microsoft Threat Analysis Center is a team of experts who analyze nation-state threats, including cyberattacks and influence operations, by combining cyber threat intelligence with geopolitical analysis, and provide insights to customers and Microsoft for effective response and protection.

Microsoft Threat Intelligence Center (MSTIC) identifies, tracks, and disrupts the most sophisticated nation-state and financially motivated threat actors impacting Microsoft and its customers. To deliver on this mission, MSTIC collects and analyzes threat information to produce actor-centric cyber threat intelligence and delivers high quality finished intelligence analysis, detections, and insights across Microsoft's security solutions.

Microsoft Threat Protection Research is a team that combines the trillions of signals we see daily with world class security research into highly sophisticated and emerging threats to deliver prevention, detection, response and automated disruption capabilities to more than 1 billion devices across all domains (Endpoint, Identity, Office, Cloud, IoT/OT.)

National Security Officers A team of globally based senior cybersecurity experts working with government stakeholders, ranging from advising on best practice cyber guidelines, support with driving compliancy and, certification of Microsoft's services and products in countries with particular national requirements.

Office of Responsible AI (ORA) collaborates with stakeholders across Microsoft to develop policies, practices, and governance systems to uphold our AI principles. ORA also helps to shape the new laws needed to ensure that the promise of AI technology is realized for the benefit of society at large.

Office of the Chief Scientific Officer leads strategic initiatives at the confluence of the sciences, technology, and society, including frontier efforts in AI.

Operational Threat Intelligence Center (OpTIC) is responsible for managing and disseminating cyber threat intelligence that supports the investigation and mitigation of threats impacting Microsoft. OpTIC delivers actionable intelligence to security teams, leadership, and engineering groups including proactive and reactive technical analysis of adversary behaviors, and strategic reporting.

The **US Government Affairs** team advances collaborative discussions with US federal and state government representatives, policymakers, and third-party groups, as well as the UN and other international organizations. The team oversees a large variety of policy priorities including AI, Cybersecurity, Cloud, Sustainability and Competition.

Worldwide Public Sector empowers people, societies, and public sector organizations around the world with cutting-edge technology and services for effective digital transformation.

Microsoft

Microsoft Digital Defense Report

The foundations and new frontiers of cybersecurity

Learn more: https://microsoft.com/mddr

Dive deeper: https://blogs.microsoft.com/on-the-issues/

Follow us for MDDR insights and more:
https://www.linkedin.com/showcase/microsoft-security/

For more news on cybersecurity policy follow us on:
https://www.linkedin.com/showcase/microsoft-on-the-issues/

A Microsoft Threat Intelligence report
October 2024