

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent of: Roy Shkedi Attorney Docket No.: 58717-0007IP1
U.S. Patent No.: 10,984,445
Issue Date: Apr. 20, 2021
Appl. Serial No.: 15/893,354
Filing Date: Feb. 9, 2018
Title: PROVIDING COLLECTED PROFILES TO MEDIA
PROPERTIES HAVING SPECIFIED INTERESTS

DECLARATION OF DR. MICHAEL D. MITZENMACHER

I declare that all statements made herein on my own knowledge are true and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable under Section 1001 of Title 18 of the United States Code.

Date: 7/17/25

By: Michael Mitzenmacher
Dr. Michael Mitzenmacher

TABLE OF CONTENTS

I.	Introduction.....	3
II.	Qualifications and Background Information	3
III.	Overview of the '445 Patent.....	6
IV.	Teachings of Beyda	7
V.	Teachings of Herz.....	21
VI.	Combination of Beyda and Herz	25
	A. Explanation of At Least One Way a POSITA Would Have Known to Combine Beyda and Herz	26
	B. A POSITA's Motivation to Combine Beyda and Herz	44
VII.	Manner in Which The Prior Art References Render the '445 Patent Claims Obvious.....	47
	A. [GROUND 1] Beyda-Herz Combination Renders Obvious Claims 1-14..	47
VIII.	Level of Ordinary Skill in the Art	57
IX.	Legal Principles	59
	A. Terminology.....	59
	B. Legal Standards.....	60
	C. Anticipation.....	60
	D. Obviousness	61
X.	Conclusion	65

I. INTRODUCTION

1. I have been retained on behalf of LiveIntent, Inc (“Petitioner”) to offer technical opinions related to 10,984,445 (“the ’445 Patent”) (EX1001). I understand that Petitioners are requesting that the Patent Trial and Appeal Board (“PTAB” or “Board”) to institute an *inter partes* review (“IPR”) proceeding of the ’445 Patent. I have been asked to provide my independent analysis of the ’445 Patent based on the prior art publications cited in this declaration, and my own knowledge, education, experience, and expertise in the relevant fields.

II. QUALIFICATIONS AND BACKGROUND INFORMATION

2. My qualifications are summarized here and explained in more detail in my curriculum vitae, which is attached as Appendix A to this report.

3. I am currently employed as a Professor of Computer Science at Harvard University. Specifically, I am the Thomas J. Watson, Sr. Professor of Computer Science in the School of Engineering and Applied Sciences. I joined the faculty of Harvard as an Assistant Professor in January 1999. I was promoted to Associate Professor in 2002 and to Professor in 2005. In 2010, I began a three-year term as Area Dean, which is essentially equivalent to what other schools call Department Chair of Computer Science, and held that position through June 2013. I also served as the Area Co-Chair for Computer Science from July 1, 2018, to June 30, 2019. My work address is 150 Western Ave, Allston, MA 02134.

4. I received my undergraduate degree in Mathematics and Computer Science from Harvard College in 1991. I received a Certificate of Advanced Study in Mathematics from Cambridge University in 1992. I received a Ph.D. in Computer Science from the University of California at Berkeley in 1996. From August 1996 to January 1999, I was employed as a Research Scientist at Digital Systems Research Center.

5. I have published over 250 research papers in computer science conferences and journals. My general area of research includes algorithms and data structures, as well as the analysis of probabilistic processes. I am listed as an inventor or co-inventor on 19 issued patents and am the co-author of a textbook, entitled *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, published by Cambridge University Press. I am a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and a Fellow of the Association for Computing Machinery (ACM). In addition, I have consulted for several companies while I have been a professor, such as Microsoft, Mitsubishi, Akamai, ITA software, Fluent Mobile, Adverplex, and e-Harmony.

6. I teach courses on various computer science topics to undergraduate and graduate students at Harvard. Among other topics, I have regularly taught a class on algorithms and data structures for information, networking, and communication to Harvard undergraduate and graduate students.

7. I have both done research in the area of electronic commerce, as well as worked for companies related to electronic commerce. For example, I have written papers regarding how the use of Groupon can affect ratings (such as Yelp ratings) for businesses, on the issue of information asymmetry in pay-per-bid auction sites, and on algorithms for finding statistically significant frequent item sets. Also, as examples, I have served on the scientific advisory board for e-harmony (a dating service with millions of users), and worked at companies such as Google, ITA software (which was bought by Google and serves as the basis for Google flights) and Adverplex/Cogolabs (which works in the space of web and Internet advertising).

8. Much of my work involves issues relating to networking and communications. I regularly serve on program committees for conferences in networking, algorithms, and communication. For example, I have served on the program committee multiple times for the SIGCOMM conference, which is described on the conference homepage as follows: “SIGCOMM is the flagship annual conference of the ACM Special Interest Group on Data Communication (SIGCOMM) on the applications, technologies, architectures, and protocols for computer communication.” <https://sigcomm.org/events/sigcomm-conference>. Similarly, I have served several times on the Program Committee for NSDI, the USENIX Symposium on Networked Systems Design and Implementation. I have

written papers on networking that have been published in the IEEE/ACM Transactions on Computer Networking, the SIGCOMM conference, the INFOCOM conference, and other major venues for networking research. My graduate course entitled “Algorithms at the end of the wire” covers many subjects at the intersection of networking and algorithms.

III. OVERVIEW OF THE '445 PATENT

9. The '445 Patent “generally relates to profile-based behavioral targeting advertisement placement services.” EX1001, 1:19-20. The '445 Patent “more specifically relates to an automatic system operated by a profile owner company (‘PO company’) for identifying media properties interested in having a collected profile made available to them.” EX1001, 7:19-22. The profile is collected from a “profile supplier,” who “transfers profile information, such as behavioral information, demographic information, etc.” EX1001, 2:24-27. “The automatic system makes the identification by recognizing a collected profile as one that a media property has previously requested or one that is within a category that a media property has previously requested.” EX1001, 7:22-26.

10. “Thereafter, the PO company arranges for the visitor to be tagged with a tag readable by the media property that requested such a profile.” EX1001, 7:26-28. “When an identified media property (*e.g.*, MP2 in FIG. 2) recognizes (at time *t*₂) a tagged visitor (tagged previously, at time *t*₁) within its ad space, the media

property will serve an ad 30 to the visitor (*e.g.*, visitor 20) based on the profile provided by the PO company 10.” EX1001, 10:24-31.

For example, following a redirect (not shown) from a server 10 of a PO company #1, weather.com might place a weather.com cookie on a visitor’s computer 20 marking the visitor as interested in “travel” (profile category “A”) and indicating that the profile originated with PO company #1. Later, if and when the weather.com ad server (MP2) recognizes that the same visitor 20 had happened to visit weather.com’s site (weather.com can recognize the visitor by reading its own weather.com cookie), the weather.com ad server (MP2) can choose to deliver a travel-related ad 30 to the visitor 20 within the weather.com ad space.

EX1001, 11:1-11.

11. I have not analyzed any of the priority claims made by the ’445 Patent. However, I have applied the earliest possible priority date of the ’445 Patent—June 19, 2006—as the invention date (Critical Date) in the following analysis.

IV. TEACHINGS OF BEYDA

12. Like the ’445 Patent, Beyda “relates generally to audience targeting and more particularly to matching an audience with deliverable content such as advertising.” EX1004, [0003]. Beyda recognized that “an effective marketing system may appreciate the characteristics and preferences of a specific user

regardless of the number or type of channels through which contact with the user is made.” EX1004, [0006]. Therefore, “it may be desirable to have a system and method for delivering content that integrates with and aggregates data from various sources, including the underlying systems that deliver content to the consumer.” EX1004, [0007]. “These sources may include databases and servers, as well as multiple web properties within a network of affiliated websites.” EX1004, [0006].

13. Data collection and profile generation are the first step in Beyda’s process for targeting content. EX1004, [0043]. “Initial profile information about individual audience members may be collected from available databases, such as a registration database [] and a subscription database” EX1004, [0044]. Profile data may also be collected “directly from the audience member computer [] and/or [a] targeting engine [] that indicates the website pages visited, the web searches conducted, and the emails received by the audience member.” EX1004, [0046]. “By combining the profile data obtained from the offline databases with the profile data updates that occur as a result of the audience member visiting website pages, a complete set of profile data may be collected for an audience member, reflecting both offline and online behavior and characteristics for the audience member.” EX1004, [0049].

14. Beyda’s system tracks audience members’ online behavior by utilizing unique identifiers for each audience member. EX1004, [0050], [0074].

“This tracking may be accomplished by combining a unique identifier for each audience member with website pages in the network that the audience member has visited.” EX1004, [0050]. “The unique identifier may be so used by including it in a domain cookie associated with each website page visited by the audience member.” EX1004, [0049]. Thus, as an audience member interacts with a webpage or a network of related webpages, the cookies stored in the audience member’s browser can be used to associate the audience member with a profile, and to update that profile as the audience member interacts with the webpages. *See* EX1004, [0051]-[0054], [0074].

15. In Beyda’s system, each network of related webpages (*i.e.*, a domain) collects profile information with respect to audience members and develops its own profiles for its audience members. *See* EX1004, [0133]-[0135]; *see also id.*, [0047]-[0054].

The profile data for audience members may be collected across numerous sites having the audience targeting functionality (e.g., ATS 1440-1444). Each of these sites offers potential segments and population of such segments with members in their respective domains. These domains may also be referred to as “local” domains. These domains may, for example, each correspond to a separate publisher.

EX1004, [0135]. In Beyda's system, each publisher's domain or set of domains includes its own targeting system, called an Audience Targeting System (ATS), that collects and stores profile information for audience members who visit the publisher's websites, segments members of the audience into groups based on information in their profiles, and targets content and ads based on the segmentation. *See* EX1004, [0067]-[0074], [0133]-[0136]. "For example, ATS 1440 may uniquely profile a.com audience member '123', under a unique identifier denoted as PRID₁₂₃." EX1004, [0137].

16. These domain-level targeting systems each include a data warehouse "for storing audience member profile data." EX1004, [0042]-[0048], [0055], [0127]-[0128]. "Initial profile information about individual audience members may be collected from available databases, such as a registration database [] and a subscription database" EX1004, [0044]. "These registration and subscription databases may include a wide variety of profile information such as name, gender, age, birth date, home and work addresses, telephone numbers, credit and charge card information, marital status, income level, number and age of children, employment history, hobbies, recent purchases, educational status, interests, preferences, and the like, for example." EX1004, [0044]. Profile data may also be collected "directly from the audience member computer [] and/or [a] targeting

engine [] that indicates the website pages visited, the web searches conducted, and the emails received by the audience member.” EX1004, [0046].

17. However, the ATS for any given publisher’s domain is only able to collect information about audience members when they visit the publisher’s websites. *See* EX1004, [0136]-[0141]. To broaden the ability of its system to collect the most complete information about audience members and better target content to them, Beyda further describes an Audience Matching Network (AMN) for collecting and managing profile data across different publishers’ targeting systems. *See* EX1004, [0133]-[0135]. In effect, the AMN is a central targeting system that gathers audience profile information from all participating domains and creates a network-level profile for audience members. *See* EX1004, [0133]-[0135]. “As such, the AMN 1420 is able to organize and manage segments based upon attributes that collectively traverse the network, and that otherwise would not have been identifiable by individual sites.” EX1004, [0136]. Like the domain-level targeting system, the network-level targeting system also includes a data warehouse that “is populated with and stores the various profile data as previously described, but does so at the network level.” EX1004, [0163].

18. In order to populate the domain-level and network-level profiles, each of the domains place code in their respective websites, called a “data agent,” that facilitates collection of profile data from audience members who visit the sites.

See EX1004, [0077], [0145], [0159], [0165]-[0167], [0171]-[0172]. This data agent code directs a visiting audience member's browser to both the domain's and network's targeting systems, passing relevant data to the targeting systems about the audience member's online activity (*e.g.*, details about the page the audience member is viewing) for inclusion in their profile. *See id.* In this regard, both the targeting system for the domain from which a visiting audience member requests a website and the network targeting system both receive and store profile information associated with the audience member's visit. *See id.* By collecting information in this fashion from visitors to all of the participating domains, the network-level targeting system "is able to recognize that an audience member that is (separately) uniquely profiled in different domains is in fact the same audience member, and to assemble segments that collectively traverse the numerous domains in the network, so as to determine that such an audience member is a member of a complex segment based upon information collected in the local domains." EX1004, [0141].

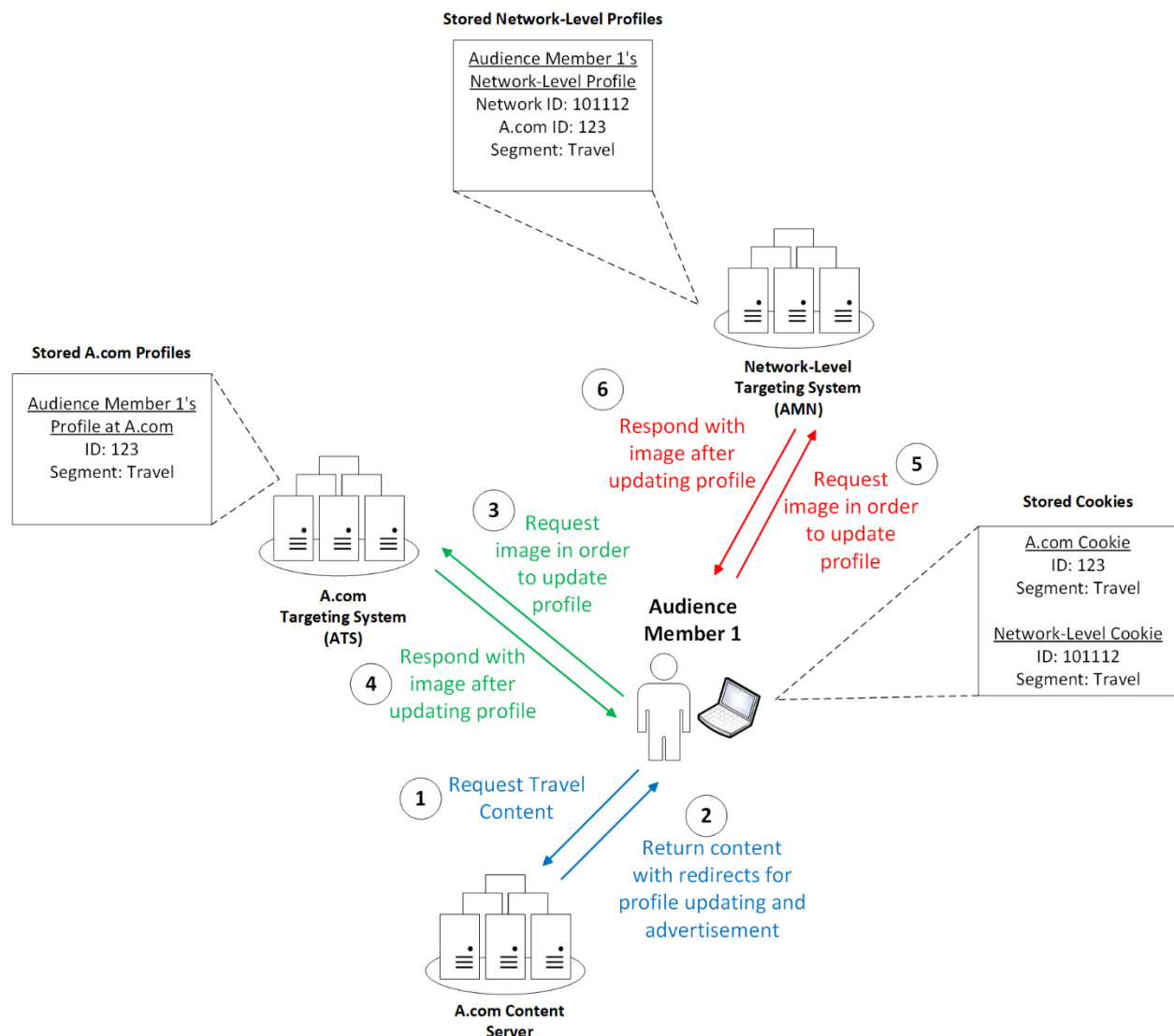
19. Each of the domains are incentivized to participate in the network for at least two reasons. First, each publisher domain can rely upon the network-level targeting system to utilize its greater level of profile data and more "complex" segmentation therefrom to place higher value advertisements and increase the publisher's revenue. *See* EX1004, [0141]-[0142] (describing the ability of the

network-level targeting system to determine membership in “complex” segments as a result of the “combination of segments defined in separate local domains in the network”); [0184] (describing that the network-level targeting system makes “publisher (web site) revenue paramount in [advertisement] placement determinations”); *see also id.*, [0005]-[0011] (describing that increased targeting information leads to delivery of more relevant content to visitors and maximization of revenue for web properties). Second, each publisher domain that contributes profile data to the network-level targeting system that is used to place advertisements is considered a “data provider” and the network-level targeting system includes a revenue sharing module that compensates data providers for contributing profile data for the visitors to their domain. *See* EX1004, [0197]-[0199].

20. In FIGS. 14A, 14B, 16A, and 16B, Beyda provides examples of how the network-level targeting system and domain-level targeting systems work together to target advertisements to audience members. *See generally* EX1004, [0133]-[0149], [0164]-[0176]. In effect, every time an audience member visits a site of one of the domains participating in the network, the “data agent” causes the audience member’s browser to provide the domain-level targeting system and network-level targeting system with information about the visit so that the user’s

domain-level profile and network-level profile may be updated with relevant information about the visit. *See* EX1004, [0077], [0159], [0165]-[0167].

21. To help illustrate this process, I provide the following diagram based on FIGS. 16A-1 to 16A-3 that shows an abstracted view of the message passing that happens in order to update an audience member's profiles when they visit a webpage.



22. As shown, the process starts when an audience member visits a webpage at one of the domains (*e.g.*, A.com) within the network. EX1004, [0165]. In this example, at (1), the audience member requests a page related to travel. EX1004, [0165]. At (2), the page is returned including data agent code that assists “in the gathering of information from website visitors” by redirecting the visitor’s browser to the domain’s targeting system (*i.e.*, ATS) and the network’s targeting system (*i.e.*, AMN). *See* EX1004, [0077], [0145], [0159], [0165]-[0167].

23. At (3), the audience member’s browser is redirected to the domain’s targeting system in order to update that member’s profile with information extracted from the visited webpage and the member’s interaction with it, usually by requesting a small transparent image. EX1004, [0165]-[0166]; *see also id.*, [0052]-[0053] (explaining the use of an image request for “blank.gif” in order to update a visitor’s profile). This update process may include several redirects to different constituent parts of the domain’s targeting system (*see* EX1004, [25], [53], [149], [0165]-[0166], [0169], [0176]), but Beyda clarifies that the targeting system may be made of any number of components executed on any number of computers (*see* EX1004, [0067]-[0069]). Whether through multiple redirects or not, the domain’s targeting system attempts to identify the audience member by determining whether the audience member has previously visited the domain by

checking for the domain's cookie in the user's browser. EX1004, [0166]; *see also id.*, [0049] (describing the setting of cookies).

24. If the user's browser is storing a domain cookie, the domain's targeting system determines the audience member's unique identifier from the cookie and retrieves the profile associated with the unique identifier. EX1004, [0165]-[0166], *see also id.* [0049]-[0054] (describing that a "unique identifier may be so used by including it in a domain cookie associated with each website page visited by the audience member," and "[b]ecause these domain cookies include the unique identifier that identifies the particular audience member, the extractor may use these cookies to modify the profile data for a particular audience member to reflect that the audience member visited the website pages associated with the cookies"), [0058] (describing that "the Targeting Engine determines whether or not a domain cookie, associated with the requested website page, includes a unique identifier for the audience member"). If the user's browser is not storing a domain cookie, the domain's targeting system creates a new profile with a new unique identifier and stores a new domain cookie on the user's browser. *See id.* In either case, the user's profile and cookies are updated with targeting information (*e.g.*, audience segments) used to target content to the audience member. EX1004, [0165]-[0166], *see also id.* [0055]-[0061] (describing a "method of associating an audience member with a segment" and that a "wide variety of content may be

provided to the audience member as a result of the segment-targeting cookie being stored on the audience member computer”).

25. At (4), the domain’s targeting system returns the requested transparent/blank image, and at (5) and (6), the data agent causes the user’s browser to initiate a similar profile updating process with the network’s targeting system. *See* EX1004, [0165]-[0168]; *see also id.*, [0052]-[0053]. As shown in FIG. 15, the network targeting system (*i.e.*, “AMN”) is similar in structure and operation to the individual domain targeting systems but is able to gather profile information whenever the audience member visits any of the domains within the overall network. *See* EX1004, [0159]-[0164] (“when an audience member logs onto a page for the first time in the domain, the [AMN] places a cookie on their browser, which contains a unique identifier”).

26. As illustrated in FIGS. 16 and 20, each audience member that visits the AMN is assigned a network-level unique identifier (*i.e.*, “NPRID”), which is set in a network cookie and allows the network targeting system to track audience members across all of the domains participating in the network. *See* EX1004, [0167]-[0168]; *see also id.*, [0142] (describing the NPRID), [0145] (“One example uses a local cookie corresponding to the local domain and a network cookie corresponding to the network domain” and the AMN may “retain a listing of PRIDs corresponding to an NPRID”). Because the data agent code for each

participating domain directs the audience member's browser to visit the AMN, the AMN "is able to recognize that an audience member that is (separately) uniquely profiled in different domains is in fact the same audience member, and to assemble segments that collectively traverse the numerous domains in the network, so as to determine that such an audience member is a member of a complex segment based upon information collected in the local domains." EX1004, [0141]. Thus, for example, if an audience member visits a travel related page in the a.com domain and a business page in the b.com domain, the network-level targeting system is able to identify the audience member as part of a "business traveler" segment, while each domain is only able to identify the audience member in the less specific segments of "travel" and "business," respectively. EX1004, [0146], [0174].

27. In addition to accumulating robust network-level profiles, the network-level targeting system also sets "an AMN cookie (AMNC)," which is an "indication whether an audience member currently has any values in the network segment cookie that may be targeted against." EX1004, [0148], [0157]. The AMN cookie effectively "indicate[s] that an AMN network ad should be served." EX1004, [0176]. When it is set, the data agent code causes the audience member's browser to request advertisements from the network-level targeting system. EX1004, [0148]-[0149], [0168]-[0169], [0175]-[0176]. Otherwise, the domain level targeting system serves its own advertisement. EX1004, [0149], [0169].

28. Beyda leaves open to the designer those instances where the AMN cookie is set to ‘N,’ indicating that an AMN network ad should not be served. See EX1004, [0148]-[0149], [0157], [0168]-[0169], [0175]-[1076]; *see also id.*, [0144] (teaching that in another context “[t]he network designer is free to establish the system as desired,” which a POSITA would have understood to apply more generally). However, a POSITA would have understood or at least found obvious that, because the purpose of the AMN is “to organize and manage segments based upon attributes that collectively traverse the network, and that otherwise would not have been identifiable by individual sites” (*see* EX1004, [0136], [0141]) and the AMN cookie indicates “whether an audience member currently has any values in the network segment cookie that may be targeted against” (*see* EX1004, [0148], [0157]), the AMN cookie would typically be set to Y where the network-level profile has more information from other domains than is available in the domain-level profile. “As is well known, publishers may display online advertisements on pages, and advertisers typically pay publishers to place these advertisements.” EX1004, [0135]. Thus, a POSITA would have understood or at least found obvious that a domain would only want to share revenue with the network-level targeting system to serve an advertisement where the network-level targeting system has more specific/valuable targeting information than the domain already has available. *See* EX1004, [0196]-[0197] (describing the “standard practices in

the industry where by two or more partners agree to split the revenue generated by an ad placement”).

29. In the above-described examples provided by Beyda, the network-level targeting system is described as receiving information from participating domain-level targeting systems in order to build a cross-domain profile. However, Beyda also describes that the network-level targeting system may act as “a clearing house for global, non-personally identifiable consumer data that may be shared across some or all publishers within the network.” EX1004, [0204]. “The assets of each publisher are [] protected by not sharing this data with other publishers, unless specific agreements exist to share data within the network.” EX1004, [0208]. In the presence of such agreements, the network-level targeting system may “share limited anonymous profile data across the network, based upon the publisher’s willingness to opt-in certain information and also in compliance with the publisher and network privacy policies.” EX1004, [0212]. In other words, the network-level targeting system may share profile information received from a domain with other domains within the network, under appropriate privacy constraints. *See* EX1004, [0204], [0208], [0212]. As a result, a domain-level targeting system not only receives profile information directly from visitors to its domain but may also receive profile data from other domains via the network-level targeting system. *See id.*

30. Beyda does not limit the techniques by which profile information contributed by one domain is shared by the network-level targeting system with other domains in the network. *See* EX1004, [0201]-[0213]. And Beyda explains in the context of defining segments that “[t]he network designer is free to establish the system as desired,” which a POSITA would have understood reflects the overall flexibility of Beyda’s disclosure. *See* EX1004, [0144]; *see also id.*, [0090] (“The designer is free to structure the reporting options as desired”), [0216] (“Although the present invention has been described in considerable detail with reference to certain embodiments thereof, the invention may be variously embodied without departing from the spirit or scope of the invention”). Accordingly, a POSITA would have understood or at least found obvious that, in designing Beyda’s system to share profile information amongst domains, he/she would have relied upon his/her own knowledge as well as techniques used in similar systems to provide this functionality. One similar system that provides example mechanisms by which a central profile collecting entity may share profile information amongst different participating publishers was described by Herz.

V. TEACHINGS OF HERZ

31. Herz describes “a system to allow a privacy-protected market for data exchange between multiple self-interested parties.” EX1005, [0003]. Generally, Herz provides “a central data warehouse that maintains data submitted by different

users, and executes queries and programs on the data.” EX1005, [0003]. “In application, the invention enables new systems for the delivery of personalized information, profiling and automated matchmaking and information exchange, all within a framework that protects privacy and maintains data security.” EX1005, [0043].

32. Every actor within Herz’s system is referred to as an “agent.” *See* EX1005, [0044]. “Agents may represent any party with individual goals, autonomy of control, and a persistent identity.” EX1005, [0044]. One agent may be “an individual browsing the Internet” who interacts with other agents “representing vendors and content providers.” EX1005, [0032]. “Examples in business-to-consumer e-commerce include business such as newspapers, book stores and travel companies, that wish to receive data about the profiles and buying habits of users so that they can personalize the information, products and services that are sold to users.” EX1005, [0044].

33. “The central SDI [Secure Data Interchange] data warehouse is the core system in SDI, managing data records on behalf of agents.” EX1005, [0078]. The central SDI data warehouse contains data submitted by agent computers, including, for example, “data that represents user profiles (i.e. what types of books they like to read, what there [sic] political interests are, etc.), or data that represents business rules (i.e. what type of customer a business is seeking, what types of

products it offers, what type of strategic agreements it is looking to make with other businesses, etc.).” EX1005, [0083]. Herz teaches that “providers of information provide the central SDI warehouse with a certain amount of autonomy to control how the information is used: the types of uses that are permitted, the type that are not prohibited, and what types of rewards (financial or otherwise) are required for access to the information.” EX1005, [0084].

34. “Agents can submit queries to the SDI data warehouse, [which] are placed in the pending queries queue if they are one-time and to be executed as quickly as possible, or in the persistent queries queue if they are to be executed periodically, as the data in the repository changes.” EX1005, [0086]. Herz describes a query execution module that is part of the central data warehouse and receives the queries from the agents. EX1005, [0557], [0808], [0810]. “Persistent queries check for certain information to reside in the data warehouse, and notify a requesting agent whenever these conditions exist, perhaps automatically making an action.” EX1005, [0093]. “One typical use of such a query is to say ‘introduce me to other users with property P’, such that the system introduces user A_1 with user A_2 if A_1 has the property required by A_2 and also A_2 has the property required by A_1.” EX1005, [0556].

35. In one use case for queries to the SDI data warehouse, Herz describes that “[v]endors can use information in the central SDI data warehouse for users’

cross vendor and within vendor browsing and purchasing habits, and also with respect to profiling information about a user.” EX1005, [0796]. “One key application is first-time personalization, so that information and products can be targeted to a user when he/she first hits a web page, based on profile information that the user is willing to release.” EX1005, [0797]. To facilitate this type of “first-time personalization,” Herz teaches that the SDI data warehouse allows “[v]endors and users [to] sell this information to other vendors.” EX1005, [0797]. In this way, vendors can “perform collaborative-filtering style data-mining across different data sets from multiple vendors, but only with the explicit authorization of owners of that information, as presented by price-rules associated with that data in the main data warehouse.” EX1005, [0422]. And user’s (*i.e.*, website visitors) can indicate whether they are “willing to release profile information to enable useful advertising as specified in their price rules for regular data.” EX1005, [0808].

36. Therefore, consistent with Beyda’s description of sharing profile information between publishers’ domain-level targeting systems, a POSITA would have understood or at least found obvious that Herz’s system was configured to allow vendors to exchange profile information with one another consistent with their respective privacy policies and user agreements. *See* EX1005, [0422], [0796]-[0797], [0808]. And, as described above, one mechanism Herz provides for

the exchange of information between agents, such as two vendors, was the use of “a persistent query [that] allows an agent to maintain a permanent ‘searching’ presence in the central database, always on the lookout for useful information.” EX1005, [0037]; *see also id.*, [0030] (“We also propose to allow ‘persistent queries’, which reside on the server for a fixed period of time and return a response to the querying agent whenever conditions are met in the data.”), [0086], [0093], [0147], [0271], [0557], [1166]. In other words, a persistent query causes a response to occur whenever profile data is updated and the update leads to the criteria of the query being matched. *See id.* For example, where a vendor is interested in tailoring content to first-time website visitors with specific interests (*see, e.g.*, EX1005, [0012], [0667], [0719], [0738]), Herz’s system was configured to permit the vendor to submit a persistent query for users with profiles that indicate the users have those interests.

VI. COMBINATION OF BEYDA AND HERZ

37. As I discuss above, Beyda’s network-level targeting system may share profile information received from a domain with other domains within the network, under appropriate privacy constraints. *See* EX1004, [0204], [0208], [0212]. However, Beyda does not require or otherwise limit a methodology for sharing by the network-level targeting system profile information contributed by one domain with other domains in the network. *See* EX1004, [0201]-[0213]. As

of the Critical Date, a POSITA would have found obvious that at least one way of facilitating the sharing of profile information between publisher domains would have been the use of the queries—and particularly the persistent queries—described by Herz.

A. Explanation of At Least One Way a POSITA Would Have Known to Combine Beyda and Herz

38. In the proposed combination, a POSITA would have found it obvious that Beyda’s network-level targeting system would have used persistent queries as at least one of the mechanisms to allow domains to request profile information for website visitors who meet certain criteria (*e.g.*, visitors with certain interests), consistent with the teachings of Herz. EX1004, [0212] (teaching that the network-level targeting system may “share limited anonymous profile data across the network, based upon the publisher’s willingness to opt-in certain information and also in compliance with the publisher and network privacy policies.”). Persistent queries in the combination would have passed updated profile information (gathered from one domain) from the network-level targeting system to other interested domain targeting systems whenever a profile is updated. *See* EX1005, [0034], [0093], [0147], [0557]. Implemented in the context of Beyda’s system, Herz’s persistent queries would be submitted by publisher domains to Beyda’s network-level targeting system, which plays a similar role to Herz’s SDI as a

central data warehouse for network-wide consumer/visitor profile data. *See, e.g.*, EX1004, [0136] (describing the capability of the AMN to organize and manage profile data “based upon attributes that collectively traverse the network, and that otherwise would not have been identifiable by individual sites”), [0163] (describing the AMN as including a “data warehouse” that “is populated with and stores the various profile data . . . at the network level”); EX1005, [0371] (“An important application of SDI is to a system for business-to-consumer (B2C) e-commerce, where SDI allows individuals to provide vendors with access to profile information that is collected client-side about an individual as he/she interacts with vendors . . .”), [0574] (“SDI is intended to function as the intermediary between a vast web of vendors . . .”), [0719] (teaching that the SDI storing user profile data is “particularly important for first-time visitors to the vendors site and particularly within vendor domains in which content-based merchandise . . . are sold”), [0728] (describing that the SDI “can track users across multiple domains, to allow fine level control over ads that the user sees”).

39. A POSITA would have found it obvious that, in the context of Beyda’s system of redirects and cookies, at least one way to implement persistent queries would have been for the network-level targeting system to cause the audience member’s browser to be directed to a querying domain’s targeting system when that audience member’s profile data satisfies a query submitted by the

querying domain. *See* EX1004, [0135]-[0145] (describing the identification of new audience members and collection of their profile data through the use of cookies).

40. More specifically, applied in the context of Beyda, where the central data warehouse is part of the network-level targeting system (*see, e.g.*, EX1004, [0163]), a POSITA would have at least found it obvious that, in order to facilitate the described sharing of profile data between publishers (*see, e.g.*, EX1004, [0204], [0208], [0212]), a query execution module would receive and periodically evaluate persistent queries submitted by domains seeking “to maintain a permanent ‘searching’ presence in the central database.” *See* EX1005, [0034], [0271]. As data agent code on publisher’s websites directed an audience member’s browser to provide updated profile information to the network-level targeting system (*i.e.*, when the condition exists that certain information would come to reside in the data warehouse), a POSITA would have found it obvious for a query execution module associated with the network-level targeting system to evaluate any pending persistent queries that would be satisfied by the newly added profile information. *See* EX1004, [0163]-[0168] (describing data agent code directing a user’s browser to the network-level targeting system to provide profile information); EX1005, [0034], [0093], [0271].

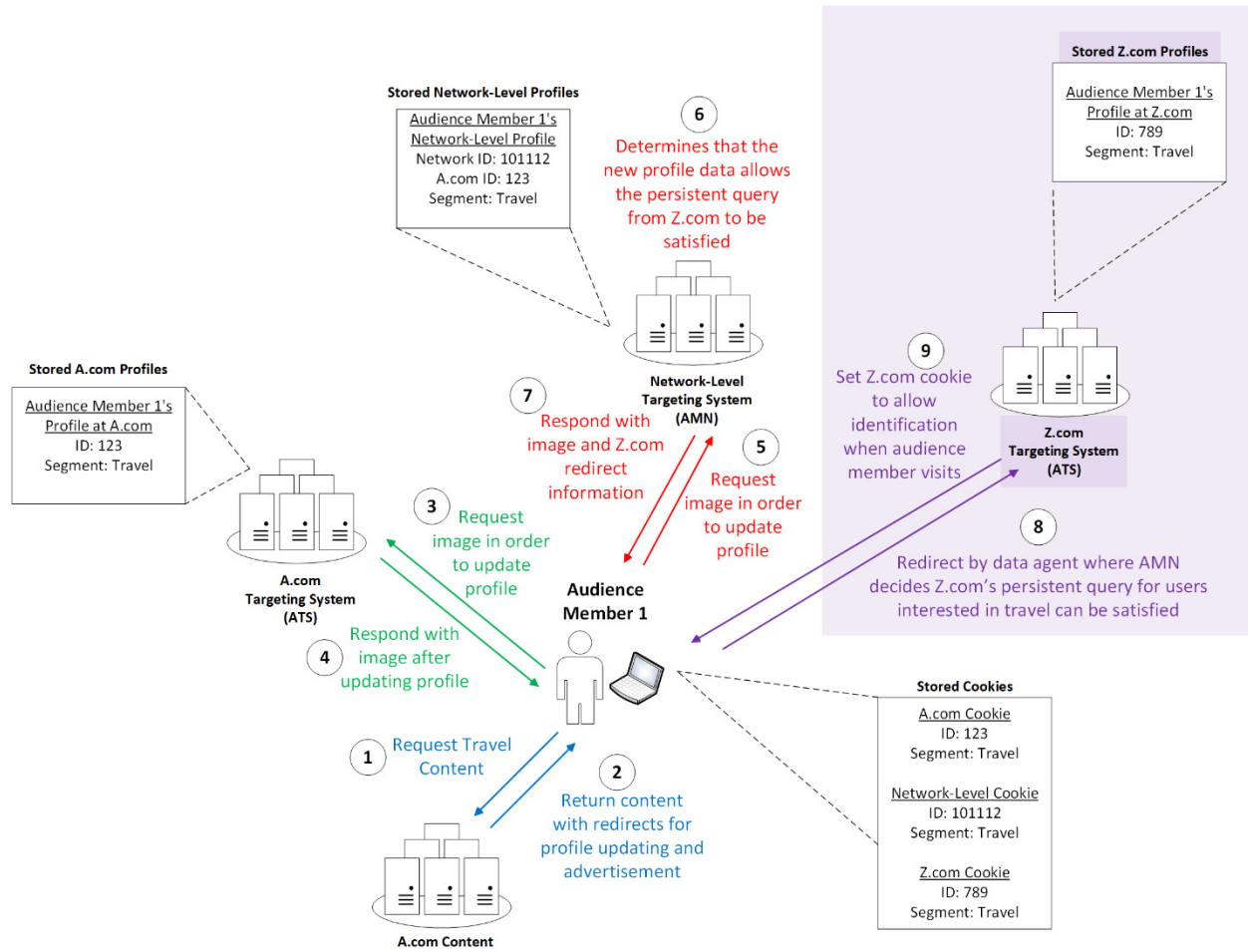
41. In order to evaluate persistent queries, the query execution module periodically checks the queries and check for certain information to reside in the data warehouse consistent with the query. *See* EX1005, [0093], [0271]. A POSITA would have understood or at least found obvious that such evaluation would have involved comparing the profile information newly added to the data warehouse of the network-level targeting system with the information identified in the persistent requests as being of interest to a requesting domain. *See id.* For example, if an audience member's visit to A.com results in provision of profile information indicating the audience member is interested in travel, the query execution module would compare this profile information against the pending persistent queries to determine whether any requesting publisher domains were seeking a "match" with audience members interested in travel. *See* EX1004, [0165]-[0168] (describing a use case in which an audience member visiting A.com is determined to belong to the "travel" segment and this information is added to the audience member's domain-level and network-level profiles); EX1005, [0057] (describing that persistent queries may provide a "matchmaking" function by which an agent's query says "introduce me to other users with property P"), [0093], [0271].

42. As I noted above, Beyda describes that the redirects that provide profile information to the various targeting systems are facilitated by the data agent

code included in each participating domain's websites. *See* EX1004, [0145] (describing that the "the data agent associated with ATS 1440[] directs the particular audience members browser to the AMN 1420"), [0159] (describing components of the data agent "sending data" within various domains), [0165] (a "data agent for the domain causes a communication to be sent"), [0173] (describing the data agent "prompt[ing] the visitor's browser to provide 1656 information to the AMN"). By redirecting the audience member's browser to the querying domain's targeting system, the audience member's browser would be able to provide the matching profile information to the querying domain's targeting system and would allow the querying domain's targeting system to set a cookie to identify the user during future visits to that domain—exactly as is done for the visited domain's targeting system and the network-level targeting system. *See* EX1004, [0164]-[0175].

43. In other words, in order to satisfy persistent queries in the context of Beyda's system, a POSITA would have found it obvious that, when the network-level targeting system received updated profile information as a result of a redirect by a data agent, the network-level targeting system (using, for example, a query execution module like the one described by Herz) would have determined whether the new profile information satisfied any pending persistent queries. *See* EX1004, [0145] (describing how a data agent directs an audience member's browser to

targeting systems), [0165]-[0166] (describing same); EX1005, [0147]-[0156] (describing how the central SDI's query execution module "[m]onitor[s] conditions for the persistent queries, and provide a response to a query if conditions are satisfied"). If the new profile information satisfied a pending persistent query, the network-level targeting system would have returned an indication to the data agent to direct the audience member's browser to the domain that submitted the persistent query, which would allow the querying domain to create a profile for the audience member with the profile information and set a cookie on the audience member's computer so that the audience member can be identified during future visits. *See id.* This additional redirect is illustrated in the following diagram that updates the diagram based on Beyda's FIGS. 16A-1 to 16A-3 that I provided above.



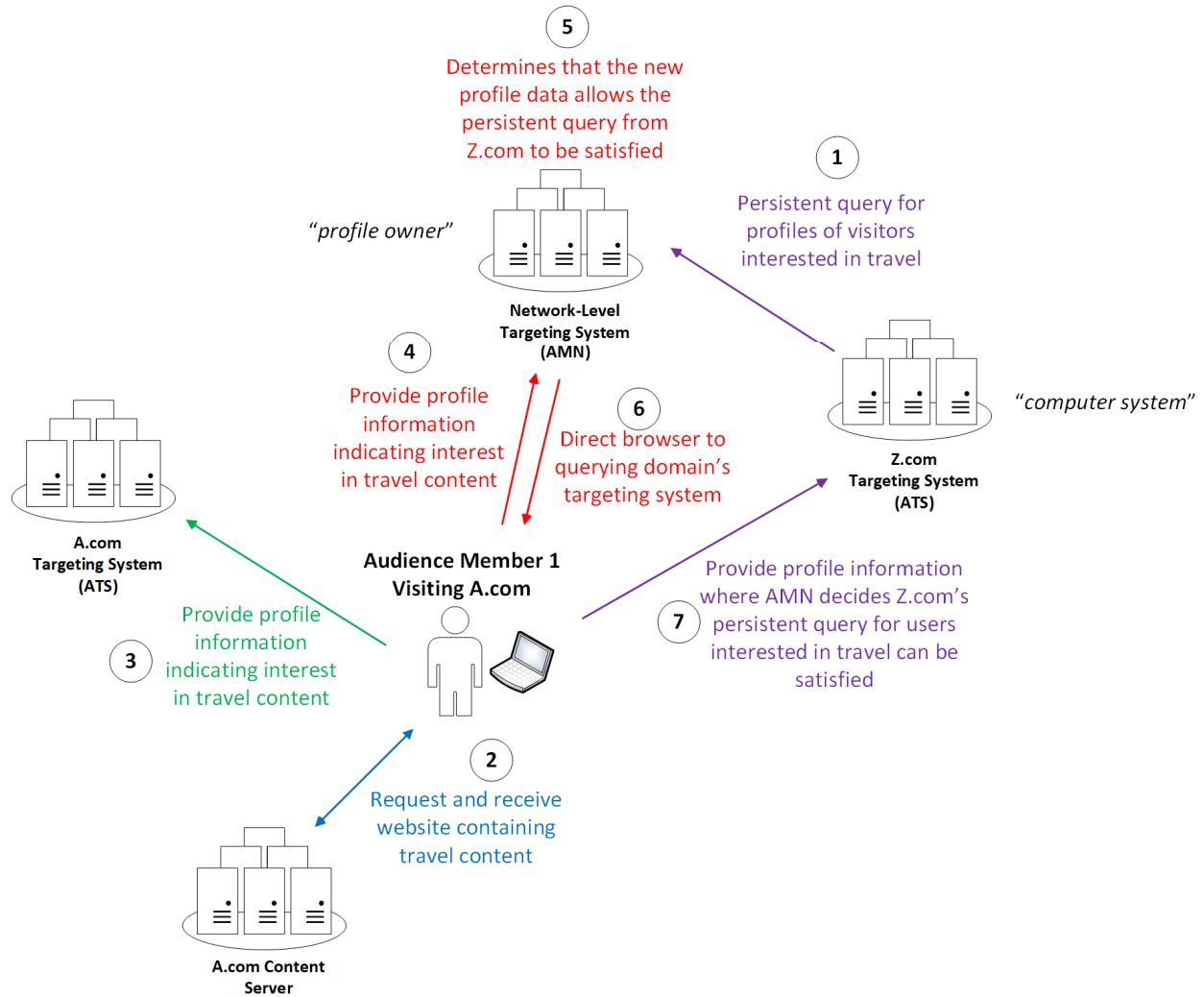
44. As shown in the updated diagram, when the data agent code directs audience member's browser to the network-level targeting system to update the audience member's network-level profile (5), the network-level targeting system determines whether the new profile data (*e.g.*, that the audience member is interested in travel) satisfies any persistent queries submitted by other domains, and, in this example, determines that the new profile data allows the persistent query from Z.com to be satisfied (6). *See* EX1005, [0147]-[0156]. Where the new profile data satisfies a persistent query, the network-level targeting system returns

information that causes the data agent code to direct the user's browser to the domain-level targeting system associated with the satisfied persistent query (7). *See* EX1004, [0159], [0165]-[0167] (describing data agent code directing an audience member's browser to multiple targeting systems). Similar to the redirects to the A.com targeting system and network-level targeting system, the data agent code causes the audience member's browser to pass the profile data to the Z.com targeting system (8), which creates a profile for the user with a new ID for the audience member (*e.g.*, "789") and the profile information, and sets a cookie on the audience member's browser (9) to allow the Z.com targeting system to identify the audience member in the future. *See* EX1004, [0135]-[0145] (describing the identification of new audience members and collection of their profile data through the use of cookies).

45. The '445 Patent describes a "profile owner" as an entity "who either own[s] the profiles or ha[s] the right to resell the profiles on behalf of another profile owner." EX1001, 5:50-62; *see also id.*, 7:45-49 (describing that the profile owner company either "owns profiles directly or acts as a databank representing other profile owners"). In the context of the Beyda-Herz combination, the network-level targeting system is a "profile owner," as that term is used in the '445 Patent, at least because the network-level targeting system (*i.e.*, AMN) acts "as a clearing house for global, non-personally identifiable consumer data that may be

shared across some or all publishers within the network.” EX1004, [0204]. Beyda describes that the data warehouse of the network-level targeting system “is populated with and stores the various profile data . . . at the network level,” resulting in the network-level targeting system being “able to recognize that an audience member that is (separately) uniquely profiled in different domains is in fact the same audience member, and to assemble segments that collectively traverse the numerous domains in the network, so as to determine that such an audience member is a member of a complex segment based upon information collected in the local domains.” EX1004, [0141], [0163]. And, in combination with Herz, Beyda’s network-level targeting system is further able to resell the profiles on behalf of another profile owner. *See* EX1005, [0797] (describing that Herz’s central SDI data warehouse allows “[v]endors and users [to] sell this [profile] information to other vendors”).

46. Below is a further focused diagram that illustrates the submission and fulfillment of a persistent query based on the combined teachings of Beyda and Herz.



47. Implemented in the context of Beyda’s system, Herz’s persistent queries would be submitted (1) by publisher domains to Beyda’s network-level targeting system in order to receive audience member profile information that meets certain criteria. *See, e.g.*, EX1004, [0136], [0163]; EX1005, [0371], [0574], [0719], [0728]. In order to satisfy persistent queries in the context of Beyda’s system, a POSITA would have found it obvious that, when the network-level targeting system received updated profile information as a result of a redirect by a

data agent (4), the network-level targeting system (using, for example, a query execution module like the one described by Herz) would have determined whether the new profile information satisfied any pending persistent queries (5). *See* EX1004, [0145], [0165]-[0166]; EX1005, [0147]-[0156]. If the new profile information satisfies a pending persistent query, the network-level targeting system would have returned an indication to the data agent to direct the audience member's browser to the domain that submitted the persistent query (6), which would allow the querying domain to create a profile for the audience member with the profile information and set a cookie on the audience member's computer so that the audience member can be identified during future visits (7). *See id.* Beyda describes that a targeting engine may direct an audience member's computer through the use of browser redirects to URLs associated with other targeting systems, and that the URLs sent to an audience members browser to facilitate the redirect may themselves include URLs to other locations (e.g., "http://te.primarydomain.com/blank.gif?te.domain1.com"). *See* EX1004, [0052]-[0053].

48. After having set the cookie in the user's browser, the targeting systems of the Beyda-Herz combination would have used the cookies to serve targeted ads. Specifically, "[w]henever that audience member returns to the site, the unique identifier is sent back to the" targeting system, and "can be used for the

delivery of targeted content such as ads.” EX1004, [0074]. Specifically, in the context of the “method of delivering targeted content to an audience member based on the segment affinity data” illustrated in FIG. 6, “[o]nce a website domain cookie is provided with a unique identifier, . . . the Targeting Engine may query the data warehouse for any segment affinity data associated with the audience member.” EX1004, [0058]-[0059]. For example, “[i]n connection with visiting site A, the audience member browser is also directed 1628 to the a.com domain Ad Server A to for advertisements (these may be those populating the currently visited page).” EX1004, [0169]. If, for example, the AMN cookie is set to ‘N’, the “Ad Server A would simply serve a local advertisement.” EX1004, [0169].

49. In the Beyda-Herz combination, profile building and advertisement targeting would have happened across all of a user’s devices. For example, Beyda describes that its system “is adapted to segment and target audience members for delivering content to an audience member across a plurality of digital mediums.” EX1004, [0062]. “The digital mediums may be heterogeneous, and may include, but are not limited to, a website network, a cable system, a non-web based internet network, a wireless communications system, such as a cellular phone or RF network, and/or any digital medium in which the means for interfacing the audience member with the digital content is uniquely addressable.” EX1004, [0062]. As part of tracking audience member activity across multiple mediums,

Beyda describes that audience members may be assigned unique profile identifiers in each medium and associate the audience members' unique profile identifiers across the various mediums. *See* EX1004, [0065]-[0066]. Further, “[a]nother aspect of [Beyda’s] invention provides profile synchronization” allowing the association of multiple profiles where an audience member “access[es] various computers throughout the day and week, such as a home computer, office computer, mall kiosk, or the like.” EX1004, [0114], [0118].

50. In one exemplary use case, an audience member may use his/her computer to visit a website provided by a web server, where “a unique audience member identifier related to the website network is assigned to the audience member, and profile data is collected and stored.” EX1004, [0065]. When the user visits another website associated with a cable network, the cable network “may read the audience member's web network identifier, and associate the audience member's cable network identifier with this identifier.” EX1004, [0066]. Using segment information associated with the audience member’s web network identifier, an “appropriate digital signal is then sent to the cable head-end 450, and the head-end 450 delivers the audience member targeted content via the set-top box 480 and the audience member's home television.” EX1004, [0066]. “In this manner, the audience member’s preference and behavioral data is synchronized across a plurality of mediums into a common profile, and the content delivered to

the audience member via those mediums may be customized based upon the characteristics of the profile.” EX1004, [0066].

51. Beyda suggests using the initially collected registration information to assist in profile synchronization. *See generally* EX1004, [0115]-[0132]. “Profile synchronization entails a recognition that audience members, and the potential multiple identifiers that they may become associated with, may be associated with an authoritative identifier (ID).” EX1004, [0123]. “In one embodiment, the Authoritative ID is associated to registration (e.g., login credentials, REGID) for the user web site.” EX1004, [0123]. While Beyda describes that, “[p]referably, the demographic data will be non-personal data,” (*see* EX1004, [0134]) Beyda’s system nonetheless provides for the collection and use of personally identifiable information (*e.g.*, registration data) to form an Authoritative ID as part of profile synchronization. EX1004, [0123]. For example, as part of the registration data, Beyda describes that the targeting system may collect “telephone numbers.” *See* EX1004, [0044]. A POSITA would have at least found it obvious that one Authoritative ID would include the audience member’s telephone number.

52. Beyda describes that one approach to network-level synchronization involves the audience member’s client side browser storing the Authoritative ID using plug-in technology. *See* EX1004, [0209]. Such implementations of “authoritative identifiers along with the collective information of the network []

provide[s] a robust network profile that can help overcome the problems associated with cookie blocking technologies.” EX1004, [0203]. Based on these teachings, a POSITA would have at least found it obvious to supplement or replace cookie-based identification with storage of Authoritative IDs using plug-in technology such that an audience member can be definitively identified as part of data gathering. *See* EX1004, [0206]-[0210]. Additionally or alternatively, a POSITA would have found it obvious that the Authoritative ID could be stored in the cookie of the domain that collected the registration information used to create the Authoritative ID, as a cookie set by a domain is only accessibly by that domain and therefore would not be viewed by other domains.

53. The ’445 Patent describes that “[a] tag generally is a unique identifier used to mark a person electronically visiting a media property, such as a web site, TV channel, radio show, or the like, using a computer, a mobile device, a TV set, a TV set top box, or any other device.” EX1001, 3:25-29. Thus, consistent with the ’445 Patent, the Authoritative ID—which may include personally identifiable information—stored on the audience member’s browser may be a tag, or at least a portion of a tag, used to associate an audience member with their profile data when visiting publisher’s sites. EX1004, [0206]-[0210].

54. With respect to FIGS 18 and 19, Beyda describes the network-level targeting system collecting and analyzing information about the delivery of

advertisements by the network-level targeting system itself, and does not focus on information collected when a domain-level targeting system “simply serve[s] a local advertisement.” *See generally* EX1004, [0177]-[0200]; *see also id.*, [0169]. However, as part of the standard audience monitoring functionality, any time that an audience member requested and interacted with a website of one of the participating domains, Beyda describes collecting audience member “behavioral data” for inclusion in the audience member’s profiles. *See* EX1004, [0017], [0049], [0056], [0066], [0072], [0093]-[0094], [0134].

55. Beyda describes that the collected behavioral data “may include attributes such as which sections were viewed on a site (e.g., sports, entertainment, health), which advertisements were seen (e.g., mortgage rates, allergy medication), referrers (e.g., AOL, Yahoo), the time of visiting the site (point in time, or range), and the frequency of visits to the site.” EX1004, [0072]; *see also id.*, [0093] (“Behaviors (e.g., Page Views, Ad Clicks)”). “Audience segments may be defined based upon” this collected behavioral data, and the segments “form the basis for the information that is extracted for analysis, reporting and targeting audience members in relevant segments.” EX1004, [0072]. Beyda describes that the mechanism for gathering and reporting this behavioral data to a targeting system is through data agent code included on the domain’s website with respect to which the behavior relates. *See* EX1004, [0046] (describing that the data warehouse of a

targeting system “may be used to store the profile and segment affinity data relating to audience members”), [0049] (describing that the extractor of the targeting system “may use [] cookies to modify the profile data for a particular audience member to reflect that the audience member visited the website pages associated with the cookies” such that “a complete set of profile data may be collected for an audience member, reflecting both offline and online behavior and characteristics for the audience member”), [0099] (describing that the data warehouse of a targeting system is populated “with profile data that identifies various attributes “), [0144] (describing that the “the local domains may collect behavioral and characteristic information”), [0159] (“a data agent [] is employed to assist in the gathering of information from website visitors, again provided in the form of code that is added to those pages in connection with which data collection is sought”).

56. Because Beyda describes that the data agent code included in each participating website is “employed to assist in the gathering of information from website visitors” both at the domain-level and the network-level, a POSITA would have understood or at least found obvious that any profile information collected by the domain-level data agent would also have been collected by the network-level data agent—including the advertising exposure and interaction data described above. *See* EX1004, [0077], [0159]; *see also id.*, [0206] (describing example

network-level data agent functionality that allows the data agent “to know what data points each publisher has authorized for capture and where to find that data” and for this data to be passed “via HTML tags”). Thus, when the domain-level and network-level data agents of the Beyda-Herz combination collected and directed reporting of behavior data to their respective targeting systems for inclusion in audience member profiles, a POSITA would have understood or at least found obvious that the behavior data would have reflected advertising exposure and interaction data. *See* EX1004, [0072], [0077], [0093], [0159], [0206]. And when a data agent directs reporting of profile data, Beyda describes that it does so through redirects of an audience members browser to the targeting system to which profile data is being reported. *See* EX1004, [0077], [0145], [0159], [0165]-[0167], [0171]-[0172]; *see also id.*, [0052]-[0053] (describing the use of redirects to targeting systems for assignment of unique identifiers and placement of cookies).

57. Examples of parameters that the data agent of the Beyda-Herz combination would have collected and reported included “the version of the data agent, the page referrer, the page URL, time information, and the [network profile identifier] NPRID.” EX1004, [0077]. In other words, when the data agent reported audience member data to the network-level targeting system, the data agent would have identified the profile with which the audience member data is associated via the network profile identifier. *See* EX1004, [0077]; *see also id.*

[0142] (“The AMN 1420 uses a unique identifier referred to as a network PRID (NPRID) to manage the unique identification of audience members at the network level”).

B. A POSITA’s Motivation to Combine Beyda and Herz

58. There are several reasons that a POSITA would have been motivated to make the above-described Beyda-Herz combination.

59. First, Herz provides one example of implementation details that Beyda does not otherwise limit to any specific approach. As discussed above, Beyda teaches that the network-level targeting system may act as “a clearing house for global, non-personally identifiable consumer data that may be shared across some or all publishers within the network.” EX1004, [0204]. However, Beyda does not further limit how this sharing occurs and leaves the implementation details to POSITAs. *See* EX1004, [0144]. Herz provides one example of implementation details a POSITA would have found obvious to utilize when implementing the described sharing of profile data. *See* EX1004, [0204], [0208], [0212], [0216]. A POSITA would have understood or at least found obvious from Herz’s description that one benefit of persistent queries was that they facilitated the sharing of data based on the needs/desires of specific domains, and permitted fast updating across domains based on those specific data interests. *See* EX1005, [0034], [0093], [0147], [0557].

60. Second, consistent with the teachings of Herz, a POSITA would have recognized that utilizing persistent queries would have permitted a domain to provide “first-time personalization, so that information and products can be targeted to a user when he/she first hits a web page, based on profile information that the user is willing to release.” EX1005, [0797]. By allowing the sharing of this information, audience members receive “personalized content that reflects his/her interests based on the types of information that other agents with similar profiles like to receive.” *See* EX1005, [0365]; *see also id.*, [0371].

61. Third, consistent with the teachings of Herz, a POSITA would have recognized that allowing the sharing of profile information between domains using, for example, persistent queries would have opened a new source of revenue for vendors. *See* EX1005, [0796]-[0797]. Beyda describes a revenue-sharing model in which domains that provide profile data to the network-level targeting system that is used to place ads are allotted a share of the ad placement revenue. *See* EX1004, [0183] (describing that “revenue may also be allocated 1808 corresponding to the delivery of the advertisement based upon participation as a data provider in the collection of information”), [0196]-[0199] (describing how a “data provider is compensated based upon the fraction of the data used in the targeted delivery provided by the data provider”). However, the revenue from each of these ad placements was known to be comparatively small, and Beyda

further describes that the value of contributed profile data decreases with time. *See* EX1004, [0193] (describing “data half-life analysis” and how it was “used to influence the value of a particular data value when evaluating a segment that contains that data value”); EX1006, 10:59-11:13 (describing a contemporaneous example in which an advertiser determines that an advertisement impression is worth between ten and fifty cents); EX1007, [0094] (describing a contemporaneous example in which an advertiser set a price of twenty cents “per some level of performance (e.g., click through, conversion, etc.)”). Allowing a domain to sell its profile data to other domains would have allowed the domain to set a higher price for the data and sell it before its half-life reduces its value. *See* EX1005, [0530]-[0532] (describing the financial motivation of sharing data across the “main SDI service”), [0598] (describing a vendor pricing its data for sale), [0762]-[0768] (describing an example where a vendor is “willing to sell its entire database [of profile data] for a flat fee”), [0797] (describing vendors selling profile information to other vendors and an example in which “information that a user likes a particular type of music is very valuable for vendors that sell content-based products, for example books, and CDs”). And a POSITA would have understood and found obvious that persistent queries provide a fast and easy mechanism for the exchange of data between domains and the resulting production of revenue for

the profile supplier, because persistent queries are fulfilled automatically by the query engine. *See* EX1005, [0034], [0093], [0147].

VII. MANNER IN WHICH THE PRIOR ART REFERENCES RENDER THE '445 PATENT CLAIMS OBVIOUS

A. [GROUND 1] Beyda-Herz Combination Renders Obvious Claims

1-14

62. A POSITA would have found claims 1-14 obvious in view of the Beyda-Herz combination that I describe above, as indicated in the table below. Beyda and Herz were analogous to and in the same field as the '445 Patent, as all relate to profiling internet users and targeting advertisements. *See, e.g.*, EX1001, abstract; EX1004, abstract; EX1005, [0016], [0019]-[0020].

Element	Mapping of Relevant Portion of Beyda-Herz Combination to Claim Element
[1.pre] A method of causing delivery of electronic advertisements based on provided profiles comprising:	Beyda's "invention relates generally to audience targeting and more particularly to matching an audience with deliverable content such as advertising." EX1004, [0003]. Accordingly, a POSITA would have understood and found obvious that the Beyda-Herz combination provides <i>a method of causing delivery of electronic advertisements based on provided profiles.</i>
[1.a.i] with a computer system automatically storing, in a central database associated with the computer system, profile information associated with a visitor, as a result of electronic receipt from a profile owner computer of indicia of the profile	A POSITA would have at least found obvious that the Beyda-Herz combination would <i>with a computer system</i> [e.g., the domain-level targeting system of a first publisher domain] <i>automatically stor[e], in a central database associated with the computer system</i> [e.g., in the data warehouse of the first publisher domain's

information, which profile information associated with the visitor matches a profile or kind of profiles requested by an entity controlling the computer system,	targeting system], profile information associated with a visitor [e.g., audience member profile data received from the network-level targeting system in response to a persistent query by the first publisher domain], as a result of electronic receipt from a profile owner computer of indicia of the profile information [e.g., as a result of receiving new profile data from a second, data-providing domain's targeting system, the network-level targeting system evaluates persistent queries and directs sale/provision of responsive profile information by the data agent code], which profile information associated with the visitor matches a profile or kind of profiles requested by an entity controlling the computer system [e.g., the data is provided to the first publisher's targeting system as a result of a persistent query, which "check[s] for certain information to reside in the data warehouse" and is satisfied "whenever these conditions exist"]].
[1.a.ii] wherein a tag associated with a visitor device serves as a link to the profile information associated with the visitor;	A POSITA would have at least found obvious that, in the Beyda-Herz combination, a tag associated with a visitor device serves as a link to the profile information associated with the visitor [e.g., a cookie set by the querying domain's targeting system uniquely identifies the audience member associated with the profile information that satisfies the persistent query].
[1.b.i] wherein the profile owner computer is programmed: (i) to automatically select a media property entity of a plurality of media property entities based on a comparison of (A) profile	A POSITA would have at least found obvious that, in the Beyda-Herz combination, the profile owner computer is programmed [e.g., the network-level targeting system, which "is preferably provided by software that may be executed

information about a visitor who electronically accessed electronic apparatus of a profile supplier with (B) a plurality of stored requests from the plurality of media property entities, one of which media property entities is the entity controlling the computer system, each stored request being for requested profiles or kinds of profiles, which plurality of stored requests includes the profile or kind of profiles requested by the entity controlling the computer system,	on any conventional processing system”]: <i>(i) to automatically select a media property entity of a plurality of media property entities</i> [e.g., selecting the first publisher domain as a targeting system to which an audience member’s browser should be redirected as a result of determining that a persistent query submitted by the first domain is satisfied by profile information provided by the audience member, where the first domain’s persistent query is located in a “queue” that includes a “set of persistent queries”] <i>based on a comparison of (A) profile information about a visitor who electronically accessed electronic apparatus of a profile supplier with</i> [e.g., the query execution module associated with the network-level targeting system evaluates newly received profile information collected when an audience member visits a website associated with the second, data-providing domain] <i>(B) a plurality of stored requests from the plurality of media property entities, one of which media property entities is the entity controlling the computer system</i> [e.g., the query execution module associated with the network-level targeting system evaluates the newly received profile information with respect to the “queue” with a “set of persistent queries” that includes the first domain’s persistent query], <i>each stored request being for requested profiles or kinds of profiles</i> [e.g., the persistent queries “check for certain information to reside in the data warehouse” and allow “an agent to maintain a permanent ‘searching’ presence in the central database,” where the data warehouse of the network-level targeting system “is
--	--

	populated with and stores the various profile data”], <i>which plurality of stored requests includes the profile or kind of profiles requested by the entity controlling the computer system</i> [e.g., like all other queries, the first domain’s persistent query would “check for certain information to reside in the data warehouse” and allow “an agent to maintain a permanent ‘searching’ presence in the central database”].
[1.b.ii] wherein the profile owner computer is programmed: . . . (ii) when the media property entity is the entity controlling the computer system, to automatically arrange for electronic storage of a requested profile linked to the tag that is associated with the visitor device, the tag being readable by equipment that is part of the computer system of part (a);	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the profile owner computer is programmed, when the media property entity is the entity controlling the computer system, to automatically arrange for electronic storage of a requested profile linked to the tag that is associated with the visitor device</i> [e.g., when new profile information satisfies a pending persistent query, the network-level targeting system would have returned an indication to the data agent code included in the website of the second, data-providing domain to direct the audience member’s browser to the domain that submitted the persistent query, which would allow the querying domain to create a profile for the audience member with the profile information and set a cookie on the audience member’s computer so that the audience member can be identified during future visits], <i>the tag being readable by equipment that is part of the computer system of part (a)</i> [e.g., because the cookie is set by the querying domain, the querying domain is able to read the cookie on the user’s browser].

[1.c] later, when the visitor device is available to receive an advertisement, with the computer system, (i) using the tag of part (a) that is associated with the visitor device to access the profile information stored in the central database, and (ii) using the profile information linked to the tag of part (a) to automatically cause delivery of an electronic advertisement to the visitor device, wherein the electronic advertisement is dependent on the profile information associated with the visitor.	A POSITA would have at least found obvious that the Beyda-Herz combination, <i>later, when the visitor device is available to receive an advertisement</i> [e.g., “[w]henver that audience member returns to the site” of the first, querying domain], <i>with the computer system, (i) us[es] the tag of part (a) that is associated with the visitor device to access the profile information stored in the central database</i> [e.g., the querying domain’s targeting system checks the audience member’s browser for the cookie stored as a result of the persistent query, and uses the unique profile identifier contained in the cookie to access the audience member’s profile], <i>and (ii) us[es] the profile information linked to the tag of part (a) to automatically cause delivery of an electronic advertisement to the visitor device</i> [e.g., the querying domain’s targeting system can use segment data associated with the audience member’s profile to deliver targeted advertising], <i>wherein the electronic advertisement is dependent on the profile information associated with the visitor</i> [e.g., “the profile data includes attributes that are correlated to audience members, and is the basis of the audience segment definitions that are used to target audience members with advertisements and/or other content,” as described by Beyda].
[2] The method of claim 1 wherein the profile information associated with the visitor is derived from the visitor’s activity using a first device, and wherein the visitor device that is available to receive an	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the profile information associated with the visitor is derived from the visitor’s activity using a first device</i> [e.g., Beyda’s system collects profile information from an audience member when

advertisement is a second device of the visitor.	the audience member uses a “computer 180” to visit a website provided by a web server], and wherein the visitor device that is available to receive an advertisement is a second device of the visitor [e.g., Beyda’s system “target[s] audience members for delivering content . . . across a plurality of digital mediums” including, for example, “a wireless communications system, such as a cellular phone” or a set-top box in a cable system].
[3] The method of claim 2 wherein the first device is a computer of the visitor and the second device is a mobile phone of the visitor.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, the first device is a computer of the visitor [e.g., Beyda’s system collects profile information from an audience member when the audience member uses a “computer 180” to visit a website provided by a web server] and the second device is a mobile phone of the visitor [e.g., Beyda’s system “target[s] audience members for delivering content . . . across a plurality of digital mediums” including, for example, “a wireless communications system, such as a cellular phone”].
[4] The method of claim 3 wherein the tag identifies the mobile phone.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, the tag identifies the mobile phone [e.g., the cookie stored on the audience member’s computer by the querying domain includes an Authoirtative ID that is based on registration information, which Beyda describes may include a telephone number].
[5] The method of claim 3 wherein the profile owner computer is programmed to automatically select a media property entity using profile	A POSITA would have at least found obvious that, in the Beyda-Herz combination, the profile owner computer [e.g., the network-level targeting system] is programmed to automatically select a

information about the visitor that was derived from the visitor's activity using the computer of the visitor, which profile information was received by the profile-owner computer as a result of URL redirection from a website that is served by the electronic apparatus of the profile supplier, which website is accessed by the computer of the visitor.	<i>media property entity using profile information about the visitor that was derived from the visitor's activity using the computer of the visitor</i> [e.g., selecting the first publisher domain as a targeting system to which an audience member's computer browser should be redirected as a result of determining that a persistent query submitted by the first domain is satisfied by profile information provided by the audience member as a result of accessing and interacting with the website of a second, data provider domain], <i>which profile information was received by the profile-owner computer as a result of URL redirection from a website that is served by the electronic apparatus of the profile supplier, which website is accessed by the computer of the visitor</i> [e.g., when the audience member accesses and interacts with the website of a second, data provider domain, the data agent code on the website redirects the audience member's browser to the URL of the network-level domain's targeting system, and to the URL of the first querying domain's targeting system when the audience member's profile information satisfies the first querying domain's persistent query].
[6] The method of claim 2 wherein the second device is a computer and further comprising, after parts (a) and (c), with the computer system electronically reporting information about delivery of the advertisement to the second device, which report is provided	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the second device is a computer</i> [e.g., at least one of the devices utilized by an audience member to access the first domain after the persistent query has been fulfilled is a "computer"] <i>and further comprising, after parts (a) and (c), with the computer system electronically reporting information about delivery of the</i>

to a profile owner controlling the profile owner computer.	<i>advertisement to the second device</i> [e.g., after the first domain's targeting system serves a local advertisement to the audience member based on the profile information returned in response to the persistent query, data agent code associated the first domain's targeting system included in the site viewed by the audience member collects behavior data including information about delivery of the advertisement, such as whether the advertisement was viewed or clicked by the audience member], <i>which report is provided to a profile owner controlling the profile owner computer</i> [e.g., because the first domain is a participant in the network-level targeting system, the data agent code included in the first domain's website causes profile data to be passed to both the first domain's targeting system and the network-level targeting system].
[7] The method of claim 6 wherein electronically reporting comprises automatically placing the second device in communication with computer equipment controlled by the profile owner, via URL redirection from the computer system, when the computer system utilizes the tag to cause delivery of the advertisement.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>electronically reporting comprises automatically placing the second device in communication with computer equipment controlled by the profile owner</i> [e.g., the data agent code included in the first domain's website collects behavior data of the audience member for reporting to the network-level targeting system in order to update the audience member's network-level profile], <i>via URL redirection from the computer system, when the computer system utilizes the tag to cause delivery of the advertisement</i> [e.g., the data agent code causes reporting of profile information by redirecting the user's browser to the

	domain-level and network-level targeting systems in conjunction with displaying the first domain's website, including any advertisements delivered therewith].
[8] The method of claim 1 further comprising, after parts (a) and (c), with the computer system electronically reporting information about delivery of the advertisement to the visitor device that is available to receive an advertisement, which report is provided to a profile owner company controlling the profile owner computer.	Claim [8] is substantially the same as at least portions of Claim [6].
[9] The method of claim 8 wherein reporting further comprises automatically electronically identifying to the profile owner company one of a plurality of profiles or kind of profiles used in association with delivery of the advertisement.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>reporting further comprises automatically electronically identifying to the profile owner company one of a plurality of profiles or kind of profiles used in association with delivery of the advertisement</i> [e.g., when the data agent reports the behavior data, it includes a profile identifier that identifies the profile used to deliver the targeted advertisement].
[10] The method of claim 1 wherein the profile owner computer is programmed to automatically select a media property entity using profile information about the visitor that was derived from the visitor's activity using the visitor device.	Claim [10] is substantially the same as at least a first portion of Claim [5].
[11] The method of claim 10 wherein the profile information was received by the profile-	Claim [11] is substantially the same as at least a second portion of Claim [5].

owner computer as a result of URL redirection from a website that is served by the electronic apparatus of the profile supplier, which website is accessed by the visitor device.	
[12] The method of claim 1 wherein the tag comprises personally identifiable information identifying the visitor.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the tag</i> [e.g., the Authoritative ID stored through the use of plug-in technology] <i>comprises personally identifiable information identifying the visitor</i> [e.g., the Authoritative ID is based on registration information that Beyda describes as including personally identifiable information, like name and address].
[13] The method of claim 1 wherein the profile owner computer being programmed to automatically arrange for electronic storage of the profile information linked to the tag associated with a visitor device comprises the profile owner computer being programmed to perform automatic redirection to a URL provided by the selected entity to place the visitor device in communication with computer equipment controlled by the selected entity to enable storage, in a central database associated with the computer equipment of the selected entity, of the profile information linked to the tag that is associated with the visitor device.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the profile owner computer being programmed to automatically arrange for electronic storage of the profile information linked to the tag associated with a visitor device</i> <i>comprises the profile owner computer</i> [e.g., the network-level targeting system] <i>being programmed to perform automatic redirection to a URL provided by the selected entity</i> [e.g., when the network-level targeting system determines that newly received profile information satisfies a persistent query, the network-level targeting system redirects the audience member's browser to a URL associated with the querying domain's targeting system] <i>to place the visitor device in communication with computer equipment controlled by the selected entity</i> [e.g., as part of the redirect to

	the querying domain's targeting system, the audience member's browser and computer communicate with the querying domain's targeting system by providing the profile data to the querying domain] <i>to enable storage, in a central database associated with the computer equipment of the selected entity, of the profile information linked to the tag that is associated with the visitor device</i> [e.g., in response to receiving profile information from the audience member's computer, the querying domain's targeting system creates a profile for the audience that includes the received profile information and returns a cookie to be stored on the audience member's computer that includes a profile identifier associated with the new profile].
[14] The method of claim 1 wherein the apparatus of the profile supplier operates in the Internet and the equipment of the selected entity utilizes the profile data to cause delivery of the electronic advertisements on a television of the visitor.	A POSITA would have at least found obvious that, in the Beyda-Herz combination, <i>the apparatus of the profile supplier operates in the Internet and the equipment of the selected entity utilizes the profile data to cause delivery of the electronic advertisements on a television of the visitor</i> [e.g., Beyda describes collecting profile information and serving digital advertisements "across a plurality of digital mediums," one of which may include a TV connected to a set-top box].

VIII. LEVEL OF ORDINARY SKILL IN THE ART

63. My analysis and conclusions set forth in this Declaration are based on the perspective of a person of ordinary skill in the art having the level of knowledge and skill as of the Critical Date of the '445 Patent (POSITA).

64. I was asked to consider the patent claims and the prior art through the eyes of a POSITA at the Critical Date. I understand that the factors considered in determining the ordinary level of skill in a field of art include the level of education and experience of persons working in the field; the types of problems encountered in the field; the teachings of the prior art, and the sophistication of the technology at the Critical Date. I understand that a POSITA is not a specific real individual, but rather is a hypothetical individual having the qualities reflected by the factors above. I understand that a POSITA would also have knowledge from the teachings of the prior art, including the art cited in this Declaration.

65. Taking these factors into consideration, based on my knowledge and experience in the field and my review of the '445 Patent and file history, I believe that a POSITA in this matter would have had at least a combination of (1) at least a bachelor's degree in computer science, computer engineering, electrical engineering, or a related field, and (2) at least two years of industry experience with e-commerce computing systems, including systems that support targeted advertising. More education could substitute for work experience, and *vice versa*.

66. Based on my experiences, I have a good understanding of the capabilities of a POSITA. Indeed, I have taught, participated in organizations with, and worked closely with many such persons over the course of my career. Based on my knowledge, skill, and experience, I have an understanding of the capabilities

of a POSITA. For example, from my industry experience, I am familiar with what a POSITA would have known and found predictable in the art. From teaching and supervising my undergraduate, graduate, and post-graduate students, I also have an understanding of the knowledge that a person with this academic experience possesses, and, more generally, people with various levels of academic experience possess. Furthermore, I possess those capabilities myself.

IX. LEGAL PRINCIPLES

A. Terminology

67. I understand that claim terms are generally given their plain and ordinary meaning in light of the patent's specification and file history as understood by a POSITA as of the invention date—here, the Critical Date. In that regard, I understand that the best indicator of claim meaning is its usage in the context of the patent specification as understood by a POSITA. I further understand that the words of the claims should be given their plain meaning unless that meaning is inconsistent with the patent specification or the patent's history of examination before the U.S. Patent and Trademark Office (Office). I also understand that the words of the claims should be interpreted as they would have been interpreted by a POSITA at the Critical Date.

B. Legal Standards

68. I have been informed by Counsel and understand that documents and materials that qualify as prior art can render a patent claim unpatentable as being anticipated or obvious.

69. I have been informed by Counsel and understand that all prior art references are to be looked at from the viewpoint of a person of ordinary skill in the art at the time of the invention, and that this viewpoint prevents one from using his or her own insight or hindsight in deciding whether a claim is anticipated or rendered obvious.

C. Anticipation

70. I understand that patents or printed publications that qualify as prior art can be used to invalidate a patent claim as anticipated or as obvious.

71. I understand that, once the claims of a patent have been properly construed, the second step in determining anticipation of a patent claim requires a comparison of the properly construed claim language to the prior art on a limitation-by-limitation basis.

72. I understand that a prior art reference “anticipates” an asserted claim, and thus renders the claim invalid, if all limitations of the claim are disclosed in that prior art reference, either explicitly or inherently (i.e., necessarily present).

D. Obviousness

73. I understand that even if a patent is not anticipated, it is still invalid if the differences between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a POSITA.

74. I have been informed by Counsel and understand that a claim is unpatentable for obviousness and that obviousness may be based upon a combination of references. I have been informed by Counsel and understand that the combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results. However, I have been informed by Counsel and understand that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art.

75. I have been informed by Counsel and understand that when a patented invention is a combination of known elements, a court determines whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue by considering the teachings of prior art references, the effects of demands known to people working in the field or present in the marketplace, and the background knowledge possessed by a person having ordinary skill in the art.

76. I have been informed by Counsel and understand that a patent claim composed of several limitations is not proved obvious merely by demonstrating that each of its limitations was independently known in the prior art. I am informed by Counsel and understand that identifying a reason those elements would be combined can be important because inventions in many instances rely upon building blocks long since uncovered, and claimed discoveries almost of necessity will be combinations of what, in some sense, is already known. I am informed by Counsel and understand that it is improper to use hindsight in an obviousness analysis, and that a patent's claims should not be used as a "roadmap."

77. I have been informed by Counsel and understand that an obviousness inquiry requires consideration of the following factors: (1) the scope and content of the prior art, (2) the differences between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) any so called "secondary considerations" of non-obviousness, which include: (i) "long felt need" for the claimed invention, (ii) commercial success attributable to the claimed invention, (iii) unexpected results of the claimed invention, and (iv) "copying" of the claimed invention by others.

78. I have been informed by Counsel and understand that an obviousness evaluation can be based on a single reference or a combination of multiple prior art references. I understand that the prior art references themselves may provide a suggestion, motivation, or reason to combine, but that the nexus linking two or

more prior art references is sometimes simple common sense. I have been informed by Counsel and understand that obviousness analysis recognizes that market demand, rather than scientific literature, often drives innovation, and that a motivation to combine references may be supplied by the direction of the marketplace.

79. I have been informed by Counsel and understand that if a technique has been used to improve one device, and a person of ordinary skill at the time of invention would have recognized that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill.

80. I have been informed by Counsel and understand that practical and common sense considerations should guide a proper obviousness analysis, because familiar items may have obvious uses beyond their primary purposes. I have been informed by Counsel and understand that a person of ordinary skill looking to overcome a problem will often be able to fit together the teachings of multiple prior art references. I have been informed by Counsel and understand that obviousness analysis therefore takes into account the inferences and creative steps that a person of ordinary skill would have employed at the time of invention.

81. I have been informed by Counsel and understand that a proper obviousness analysis focuses on what was known or obvious to a person of

ordinary skill at the time of invention, not just the patentee. Accordingly, I understand that any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed.

82. I have been informed by Counsel and understand that a claim can be obvious in light of a single reference, without the need to combine references, if the elements of the claim that are not found explicitly or inherently in the reference can be supplied by the common sense of one of skill in the art.

83. I have been informed by Counsel and understand that there must be a relationship between any such secondary considerations and the invention, and that contemporaneous and independent invention by others is a secondary consideration supporting an obviousness determination.

84. In sum, my understanding is that prior art teachings are properly combined where one of ordinary skill having the understanding and knowledge reflected in the prior art and motivated by the general problem facing the inventor, would have been led to make the combination of elements recited in the claims. Under this analysis, the prior art references themselves, or any need or problem known in the field of endeavor at the time of the invention, can provide a reason for combining the elements of multiple prior art references in the claimed manner.

85. I have been informed by Counsel and understand that in an IPR, “the petitioner shall have the burden of proving a proposition of unpatentability,” including a proposition of obviousness, “by a preponderance of the evidence.”

X. CONCLUSION

86. I am over the age of 18 and am competent to write this declaration. I have personal knowledge, or have developed knowledge of these technologies based upon education, training, or experience, of the matters set forth herein.

87. I am not and never have been, an employee of LiveIntent. I received no compensation for this declaration beyond my normal hourly compensation based on my time actually spent analyzing the '445 Patent, the prior art publications cited below, and issues related thereto, and I will not receive any added compensation based on the outcome of any IPR or other proceeding involving the '445 Patent.

88. In writing this Declaration, I have considered the following: my own knowledge and experience, including my work experience in the fields of networking, algorithms and data structures, electronic commerce, artificial intelligence, computer systems, and databases; my experience in teaching those subjects; and my experience in working with others involved in those fields.

89. My opinions, as explained above, are based on my education, experience, and expertise in the fields relating to the '445 Patent. Any figures that

appear within this document have been prepared with the assistance of Counsel and reflect my understanding of the '445 Patent and the prior art discussed above. In addition, I have analyzed the following publications and materials, in addition to other materials I refer to or cite in my declaration:

Exhibit Number	Citation
EX1001	U.S. Patent No. 10,984,445 to Shkedi (“the ’445 Patent”)
EX1002	Excerpts from the Prosecution History of the ’445 Patent (“the Prosecution History”)
EX1004	U.S. Pub. No. 2005/0165644 to Gil Beyda <i>et al.</i> (“Beyda”)
EX1005	U.S. Pub. No. 2009/0254971 to Frederick S. M. Herz <i>et al.</i> (“Herz”)
EX1006	U.S. Pat. No. 7,949,562 to Robert J. Collins
EX1007	U.S. Pub. No. 2005/0096979 to Ross Koningstein

90. I currently hold the opinions set expressed in this declaration. But my analysis may continue, and I may acquire additional information and/or attain supplemental insights that may result in added observations.

APPENDIX A

Michael Mitzenmacher
michaelm@eecs.harvard.edu
33 Cary Avenue, Lexington MA 02421
617-496-7172

Research Interests	Design and Analysis of Algorithms; Machine Learning Systems; Networks and Data Transmission; Databases.
Education	UNIVERSITY OF CALIFORNIA AT BERKELEY, Berkeley, CA Ph.D. in Computer Science awarded December, 1996. Dissertation: The Power of Two Choices in Randomized Load Balancing. Advisor: Alistair Sinclair. GPA: 4.0/4.0 CAMBRIDGE UNIVERSITY, Cambridge, England Attended as one of ten recipients of the Churchill Fellowship. Cambridge C.A.S. in Mathematics with highest distinction awarded June 1992. HARVARD COLLEGE, Cambridge, MA B.A. in Mathematics with Computer Science, summa cum laude, awarded June 1991.
Employment	HARVARD UNIVERSITY, Cambridge, MA <i>Spring 1999-present</i> Assistant professor (from Jan. 1999 -July 2002), Associate professor (from July 2002-January 2005), Professor (from Jan. 2005-present), Area Dean for Computer Science (from July 2010-June 2013), Area Co-Chair for Computer Science (from July 2018-June 2019). Teach undergraduate courses including “Introduction to Algorithms and Data Structures” and “Discrete Mathematics for Computer Science”, and graduate courses covering topics in randomized algorithms, algorithms for networks, compression, coding, cryptography, and information retrieval. DIGITAL SYSTEMS RESEARCH CENTER, Palo Alto, CA <i>Fall 1996-Winter 1998</i> Research scientist. Projects included work on information retrieval on the Web, erasure codes, error-correcting codes, on-line algorithms, and load balancing. Co-inventor for twelve submitted patents. SANTA CLARA UNIVERSITY, Santa Clara, CA <i>Spring 1997</i> Guest professor for the undergraduate class “Introduction to Algorithms.” Consultant: I consult on intellectual property issues as an expert witness and in other capacities. As an expert witness, I have testified in multiple trials. I have also consulted for several technology companies and research laboratories, including Adverplex (Cogolabs), Akamai, AT&T, Digital Fountain, eHarmony, Fluent Mobile (Fiksu), Google, Huawei, ITA Software, JobSync, Microsoft, Mitsubishi Research Laboratories, Weta Digital, and Yahoo.
Funding	NSF CCF-2101140: AF: Small: Algorithms and Data Structures with Predictions. PI: Michael Mitzenmacher. Total grant: \$400,000. 7/21-6/24. NSF CNS-2107078: CNS Core: Medium: Approximation and Randomization in the Programmable Data Plane. co-PIs: Minlan Yu. Total grant: \$1,200,000. 7/21-6/25.

NSF DMS-2023528: Foundations of Data Science Institute (Harvard University). 9/20-8/25.

NSF CCF-1563710: CIF: NeTS: Medium: Collaborative Research: Unifying Data Synchronization. co-PIs: David Starobinski, Ari Trachtenberg. Total grant: \$400,000. 7/16-6/20.

NSF CCF-1535705: AitF: FULL: Collaborative Research: Better Hashing for Applications: From Nuts & Bolts to Asymptotics. co-PIs: David Andersen. Total grant: \$250,000. 9/15-8/19.

NSF CCF-1320231: AF: Small: Data Synchronization : Theory, Algorithms, and Practice PI: Michael Mitzenmacher. Total grant: \$399,370. 9/13-8/16.

NSF CNS-1228598: TWC: Medium: Collaborative: Privacy-Preserving Distributed Storage and Computation. co-PIs: Michael Goodrich, Roberto Tamassia. Total grant: \$400,000. 8/12-8/16.

NSF IIS-0964473: HCC: Medium: Collaborative Research: Data-Parallel Hash Tables: Theory, Practice and Applications. co-PIs: Nina Amenta. Total grant: \$171,095. 8/10-7/13.

NSF CCF-0915922: AF : Small : The Theory and Practice of Hash-Based Algorithms and Data Structures. PI: Michael Mitzenmacher. Total grant: \$441,956. 8/09-7/12.

NSF CNS-0721491: NeTS FIND: A Network-Wide Hashing Infrastructure for Monitoring and Measurement. PI: Michael Mitzenmacher. Total grant: \$330,000. 9/07-8/10.

NSF CCF-0634923: Towards a Basic Understanding of Channels with Synchronization Errors. PI: Michael Mitzenmacher. Total grant: \$200,000. 9/06-8/09.

NSF CCR-0121154: ITR/SY Algorithmic Issues in Large Scale Dynamic Networks. co-PIs: Eli Upfal, Brown. Subcontract to Harvard. Total grant: \$502,000. 9/01-8/06.

NSF CCR-0118701: Low Density Parity Check Codes for Channels with Memory. co-PIs: Alek Kavcic. Total grant: \$510,000. 9/01-8/04.

NSF CCR-9983832: Dynamic Processes and Network Algorithms (CAREER). \$200,000. 7/00-6/04.

Dean's Competitive Fund for Promising Scholarship. 2018. \$49,333.

Google University Research Program. 8/13-8/14. PIs: Eddie Kohler and Michael Mitzenmacher \$56,000.

Google University Research Program. 12/11-12/12. \$25,000.

Yahoo! University Research Program. 7/11-6/12. \$10,000.

Google University Research Program. 12/09-12/10. \$60,000.

Yahoo! University Research Program. 9/09-8/10. \$10,000.

Google University Research Program. 8/08-7/09. \$75,000.

Yahoo! University Research Program. 9/07-8/08. \$25,000.

Yahoo! University Research Program. 9/06-8/07. \$50,000.

Cisco University Research Program. 8/08-7/09. \$80,000.

Cisco University Research Program. 8/07-7/08. \$83,000.

Cisco University Research Program. 8/06-7/07. \$80,000.

Cisco University Research Program. 8/05-7/06. \$72,000.

Alfred P. Sloan Research Fellowship. \$40,000. Awarded in 2000.

IBM Faculty Research Grant. \$10,000 Awarded in 2005.

IBM Faculty Research Grant. \$10,000 Awarded in 2003.

Mitsubishi Electronic Research Laboratory. \$10,000 for undergraduate research projects. Awarded in 2002.

Honors

Applied Cryptography and Network Security Conference Test of Time Paper Award (2025)

ACM Paris Kanellakis Theory and Practice Award (2020)

IEEE Fellow (2020)

IEEE Senior Member (2018)

ACM Fellow (2014)

ACM Symposium on Parallelism in Algorithms and Architectures Best Paper Award (2014)

World Wide Web Conference Best Paper Award (2014)

Royal Academy of Engineering Distinguished Visiting Fellowship (2010)

ACM SIGCOMM Test of Time Paper Award (2009)

IEEE Information Theory Society Best Paper Award (2002)

Alfred P. Sloan Research Fellowship (2000)

Sakrisson Award (for Ph.D. thesis at Berkeley) (1997)

NDSEG Graduate Fellowship (1992-95)

NSF Graduate Fellowship Winner (1991)

Churchill Fellowship (1991-92)

Hoopers Prize (for senior thesis at Harvard) (1991)

Phi Beta Kappa (1990)

Harvard Distinction in Teaching Award (1990)

Goldwater Fellowship (1989-91)

Professional Activities

Editorships and Boards:

Advisory Board, Hariri Institute at Boston University (2018-2021)

Editorial Board, Leibniz International Proceedings in Informatics (2009-2019)

Editorial Board, Communications of the ACM (2013-2018)

Science Board, Santa Fe Institute (2013-2019)

Guest Editor, SIAM Journal of Computing special issue for STOC 2009

SIAM Journal on Computing, Editor (2006-2010)

Internet Mathematics, Managing Editor

Guest Editor, Theory of Computing System special issue for SPAA 2002

Journal of Interconnection Networks, Editor

Organizing Committees:

Sublinear Algorithms Workshop Organizing Committee, FODSI (2022)
 Algorithms with Predictions Workshop Organizing Committee, STOC (2022)
 Machine Learning for Algorithms Workshop Organizing Committee, FODSI (2021)
 SIGACT Chair (2015-2018)
 Dagstuhl Seminar: Theory and Applications of Hashing (2017)
 EADS Summer School on Hashing: Theory and Applications (2014)
 ICERM Workshop on Stochastic Graph Models (2014)
 WAW 2013 Organizing Committee
 SIGACT Vice-Chair (2009-2012)
 General Chair, STOC (2010)
 Workshop on Randomized Algorithms and Random Graphs (2010)
 DARPA Information Science and Technology Study Groups (2007-2008)
 SIGACT Committee for the Advancement of Theoretical Computer Science (2005-2008)
 MSRI Workshop on Models of Real-World Random Networks (2005)
 FOCS 2003 Local Arrangements Chair
 Second Workshop on Randomized Algorithms and Random Graphs (2003)
 Workshop on Algorithms and Models for the Web Graph (2002)
 First Workshop on Randomized Algorithms and Random Graphs (2002)
 DIMACS Workshop on Quality of Service Issues in the Internet (2001)
 BU/NSF Workshop on Internet Measurement, Instrumentation, and Characterization (1999)
 FOCS 1998 Local Arrangements Chair

Program Committees:

NSDI 2025 Program Committee
 HotNets 2024 Program Committee
 WADS 2023 Program Committee
 ACM CoNEXT 2023 Program Committee
 NeurIPS 2022 Program Committee
 ACM CoNEXT 2022 Program Committee
 ICML 2022 Program Committee
 SIGCOMM 2022 Program Committee
 NeurIPS 2021 Program Committee
 ICML 2021 Program Committee
 ACDA 2021 Program Committee
 ICLR 2021 Program Committee
 APoCS 2021 Program Committee
 ACM CoNEXT 2020 Program Committee
 NeurIPS 2020 Program Committee
 ICML 2020 Program Committee
 APoCS 2020 Program Committee
 NeurIPS 2019 Program Committee
 SIGCOMM 2019 Program Committee
 SOSA 2019 Program Committee (Co-Chair)
 ACM CoNEXT 2018 Program Committee
 ANALCO 2018 Program Committee
 SIGCOMM 2017 Program Committee
 NSDI 2017 Program Committee
 ACM CoNEXT 2016 Program Committee
 SIGCOMM 2016 Program Committee
 ALENEX 2016 Program Committee (Co-Chair)
 ICALP 2016 Program Committee (Track C PC Chair)

ACM CoNEXT 2015 Program Committee
ICALP 2015 Program Committee
ITCS 2015 Program Committee
SIGCOMM 2014 Program Committee
ACM CoNEXT 2014 Program Committee
WSDM 2014 Senior Program Committee
WAW 2013 Program Committee (Co-Chair)
STOC 2013 Executive Committee
WSDM 2013 Senior Program Committee
SIGCOMM 2012 Program Committee
ISIT 2012 Program Committee
NSDI 2012 Program Committee
WAW 2011 Program Committee
ACM CoNEXT 2010 Program Committee
SIGCOMM 2010 Program Committee
NSDI 2010 Program Committee
FUN 2010 Program Committee
LATIN 2010 Program Committee
NetSciCom 2009 Program Committee
SIGCOMM 2009 Program Committee
STOC 2009 Program Committee (PC Chair)
NSDI 2009 Program Committee
ICALP 2008 Program Committee
SPAA 2008 Program Committee
WAW 2007 Program Committee
Analysis of Algorithms 2007 Program Committee
ANALCO 2007 Program Committee
ICALP 2007 Program Committee
RANDOM 2007 Program Committee
NCA 2006 Program Committee
STOC 2006 Program Committee
PODC 2005 Program Committee
ANALCO 2005 Program Committee
PODC 2004 Program Committee
SIGCOMM 2004 Program Committee
IPTPS 2004 Program Committee
Data Compression Conference 2004 Program Committee
ESA 2004 Program Committee
WAW 2003 Program Committee
FOCS 2003 Program Committee
Data Compression Conference 2003 Program Committee
ESA 2002 Program Committee
PODC 2002 Program Committee
SPAA 2002 Program Committee
STOC 2002 Program Committee
Data Compression Conference 2002 Program Committee
ALENEX 2002 Program Committee
HiPC 2001 Program Committee
ISAAC 2000 Program Committee
RANDOM 2000 Program Committee
STACS 2000 Program Committee

FOCS 99 Program Committee
RANDOM 98 Program Committee

Reviewer for several conferences, journals, and grant panels.

Books

Probability and Computing: Randomized Algorithms and Probabilistic Analysis, by Michael Mitzenmacher and Eli Upfal. Published by Cambridge University Press in 2005. (2nd edition available in 2017.) This is a textbook meant for an advanced undergraduate or beginning graduate class. The textbook has been used in courses at Brown, Harvard, U. C. Berkeley, Univ. of Victoria, Tufts, Univ. of Mass. at Amherst, Purdue, U. Penn., and several other universities.

**Conference
and Journal
Publications**

“Don’t Stop Me Now: Embedding Based Scheduling For LLMs,” with R. Shahout, E. Malach, C. Liu, W. Jiang, and M. Yu. To appear in *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

“Parallel Peeling of Invertible Bloom Lookup Tables in a Constant Number of Rounds,” with M. Goodrich and R. Kitagawa. To appear in *Proceedings of the 50th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, 2025.

“Efficient d -ary Cuckoo Hashing at High Load Factors by Bubbling Up,” with W. Kuszmaul. To appear in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2025.

“Leveraging Parameterized Chernoff Bounds for Simplified Algorithm Analyses,” with M. Dillencourt and M. Goodrich. *Information Processing Letters*, 187:106516, 2025.

“Optimal and Approximate Adaptive Stochastic Quantization,” with R. Basat, Y. Ben-Itzhak, and S. Vargaftik. In *Proceedings of the 38th Conference on Neural Information Processing Systems*, 2024.

“SkipPredict: When to Invest in Predictions for Scheduling,” with R. Shahout. In *Proceedings of the 38th Conference on Neural Information Processing Systems*, 2024.

“Beyond Throughput and Compression Ratios: Towards High End-to-end Utility of Gradient Compression,” with W. Han, S. Vargaftik, B. Karp, and R. Basat. In *Proceedings of the Twenty-Third ACM Workshop on Hot Topics in Networks*, pp. 186-194, 2024.

“Learning-Augmented Frequency Estimation in Sliding Windows,” with R. Shahout and I. Sabek. In *Proceedings of the Workshop on Intelligent Classification of High-Speed Network Traffic (at ICNP ’24)*.

“Learning-Based Heavy Hitters and Flow Frequency Estimation in Streams,” with R. Shahout. In *Proceedings of the 32nd IEEE International Conference on Network Protocols*, 2024.

“Tabula: Efficiently Computing Nonlinear Activation Functions for Secure Neural Network Inference,” with M. Lam, V. J. Reddi, G. Wei, and D. Brooks. In *Transactions on Machine Learning Research*, 2024.

“Accelerating Federated Learning with Quick Distributed Mean Estimation,” with S. Vargaftik, R. Basat, A. Portnoy, G. Mendelson, and Y. Ben-Itzhak. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 3410-3442, 2024.

- “THC: Accelerating Distributed Deep Learning Using Tensor Homomorphic Compression,” with M. Li, R. Basat, S. Vargaftik, C. Lao, K. Xu, X. Tang, and, M. Yu. In *Proceedings of NSDI*, pp. 1191-1211, 2024.
- “Separating k -Player from t -Player One-Way Communication, with Applications to Data Streams,” with E. Du, D. Woodruff, and G. Yang. In *Theory of Computing*, vol. 19, Article 10, pp. 1-44, 2023.
- “Direct Telemetry Access,” with J. Langlet, R. Basat, G. Oliaro, M. Yu, and G. Antichi. In *Proceedings of ACM SIGCOMM*, pp. 832-849, 2023.
- “Can Learned Models Replace Hash Functions?” with I. Sabek, K. Vaidya, D. Horn, A. Kipf, and T. Kraska. *Proceedings of the VLDB Endowment*, vol. 16, no. 3, pp. 532-545, 2022.
- “Algorithmic Tools for Mining the Motif Structure of Networks,” with B. Matejek, T. Chen, and C. Tsourakakis. To appear in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2022.
- “Edge-colored directed subgraph enumeration on the connectome,” with B. Matejek, D. Wei, T. Chen, C. Tsourakakis, and H. Pfister. *Scientific Reports*, 2022.
- “Algorithms with Predictions,” with S. Vassilvitskii. *Communications of the ACM*, Volume 65 Issue 7, pp. 33-35, 2022.
- “EDEN: Communication-Efficient and Robust Distributed Mean Estimation for Federated Learning,” with S. Vargaftik, R. Basat, A. Portnoy, G. Mendelson, and Y. Ben-Itzhak. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 21984-22014, 2022.
- “SNARF: A Learning-Enhanced Range Filter,” with K. Vaidya, S. Chatterjee, E. Knorr, S. Idreos, and T. Kraska. In *Proceedings of the VLDB Endowment*, 15(8):1632-1664, 2022.
- “Proteus: A Self-Designing Range Filter,” with E. Knorr, B. Lemaire, A. Lim, S. Luo, H. Zhang, and S. Idreos. In *Proceedings of the SIGMOD International Conference on Management of Data*, pp. 1670-1684, 2022.
- “Designing Sensitive Viral Diagnostics with Machine Learning,” with H. Metsky, N. Welch, P. Pillai, N. Haradhvala, L. Rumker, S. Mantena, Y. Zhang, D. Yang, C. Ackerman, J. Weller, P. Blainey, C. Myhrvold, and P. Sabeti. *Nature Biotechnology*, 2022.
- “The Supermarket Model with Known and Predicted Service Times,” with M. Dell’Amico. *IEEE Transactions on Parallel and Distributed Systems*, 2022.
- “Uniform Bounds for Scheduling with Job Size Estimates,” with Z. Scully and I. Grosz. In *Proceedings of 13th Innovations in Theoretical Computer Science Conference*, 114:1-114:30, 2022.
- “DRIVE: One-bit Distributed Mean Estimation,” with S. Vargaftik, R. Basat, A. Portnoy, G. Mendelson, and Y. Ben-Itzhak. In *Proceedings of the 34th Conference on Neural Information Processing Systems*, pp. 362-377, 2021.
- “Zero-CPU Collection with Direct Telemetry Access,” with J. Langlet, R. Basat, S.S. Ramanathan, G. Oliaro, M. Yu, and G. Antichi. In *Proceedings of the Twentieth ACM Workshop on Hot Topics in Networks*, pp. 108-115, 2021.

“Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix,” with M. Lam, V. J. Reddi, D. Brooks, and G. Wei. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 5959-5968, 2021.

“Putting the ‘Learning’ into Learning-Augmented Algorithms for Frequency Estimation,” with E. Du and F. Wang. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 2860-2869, 2021.

“Queues with Small Advice.” In *Proceedings of the SIAM Conference on Applied and Computational Discrete Algorithms*, pp. 1-12, 2021.

“How to Send a Real Number Using a Single Bit (and Some Shared Randomness),” with R. Basat and S. Vargaftik. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming*, 25:1-25:20, 2021.

“SALSA: Self-Adjusting Lean Streaming Analytics,” with R. Basat, G. Einzinger, and S. Vargaftik. In *The 37th IEEE International Conference on Data Engineering*, pp. 864-875, 2021.

“Partitioned Learned Bloom Filters,” with K. Vaidya, E. Knorr, and T. Kraska. In *The Ninth International Conference on Learning Representations (ICLR)*, 2021.

“Improved Sublinear Time Algorithm for Longest Increasing Subsequence,” with S. Seddighin. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 1934-1947, 2021.

“DISCOVering the heavy hitters with disaggregated sketches,” with V. Bruschi, R. Basat, Z. Liu, G. Antichi, and G. Bianchi. In *Proceedings of the 16th International Conference on Emerging Networks Experiments and Technologies (CoNEXT)*, pp. 536-537, 2020.

“Detecting Routing Loops in the Data Plane,” with J. Kucera, R. Basat, M. Kuka, G. Antichi, and M. Yu. In *Proceedings of the 16th International Conference on Emerging Networks Experiments and Technologies (CoNEXT)*, pp. 466-473, 2020.

“Optimal Learning of Joint Alignments with a Faulty Oracle,” with K. Larsen and C. Tsourakakis. In *Proceedings of ISIT*, pp.2492-2497, 2020.

“PINT: Probabilistic In-band Network Telemetry,” with R. Basat, S. Ramanathan, Y. Li, G. Antichi, and M. Yu. In *Proceedings of ACM SIGCOMM*, pp. 662-680, 2020.

“Dynamic Algorithms for LIS and Distance to Monotonicity,” with S. Seddighin. In *Proceedings of the 52nd Annual Symposium on Theory of Computing*, pp. 671-684, 2020.

“Clustering with a Faulty Oracle,” with K. Larsen and C. Tsourakakis. In *Proceedings of the Web Conference*, pp. 2831-2834, 2020.

“Prophets, Secretaries, and Maximizing the Probability of Choosing the Best,” with H. Esfandiari, M. Hajiaghayi, and B. Lucier. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, pp. 3717-3727, 2020.

“Faster and More Accurate Measurement through Additive-Error Counters,” with R. Basat, G. Einzinger, and S. Vargaftik. In *Proceedings of IEEE INFOCOM*, pp. 1251-1260, 2020. o “Scheduling with Predictions and the Price of Misprediction.” In *Proceedings of 11th Innovations in Theoretical Computer Science Conference*, 14:1-14:18, 2020.

“Equitability, Interval Estimation, and Statistical Power,” with Y. Reshef, D. Reshef, and P. Sabeti. *Statistical Science*, (35)2:202-217, 2020.

“A Digital Fountain Retrospective,” with J. Byers and M. Luby. *ACM SIGCOMM Computer Communication Review*, 49(5), pp. 82-85, 2019.

“Robust Set Reconciliation via Locality Sensitive Hashing,” with T. Morgan. In *Proceedings of the 38th Symposium on Principles of Database Systems*, pp. 164-181, 2019.

“Online Pandora’s Boxes and Bandits,” with H. Esfandiari, M. Hajiaghayi, and B. Lucier. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp. 1885-1892, 2019.

“Arithmetic Progression Hypergraphs: Examining the Second Moment Method.” In *Proceedings of ANALCO*, pp. 127-134, 2019.

“A Bayesian Nonparametric View on Count-Min Sketch,” with R. Adams and D. Cai. In *Proceedings of the 31st Conference on Neural Information Processing Systems*, pp. 8782-8791, 2018.

“A Model for Learned Bloom Filters and Optimizing by Sandwiching.” In *Proceedings of the 31st Conference on Neural Information Processing Systems*, pp. 462-471, 2018.

“Metric Sublinear Algorithms via Linear Sampling,” with H. Esfandiari. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science*, pp. 11-22, 2018.

“EMOMA: Exact Match in One Memory Access,” with S. Pontarelli and P. Reviriego. *IEEE Transactions on Knowledge and Data Engineering*, 30(11):2120-2133, 2018.

“Weightless: Lossy Weight Encoding for Deep Neural Network Compression,” with B. Reagen, U. Gupta, R. Adolf, A. Rush, G. Wei, and D. Brooks. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 4321-4330, 2018.

“Load Thresholds for Cuckoo Hashing with Double Hashing,” with K. Panagiotou and S. Walzer. In *Proceedings of the 16th Scandinavian Symposium and Workshops on Algorithm Theory*, pp. 29:1-29:9, 2018.

“Joint Alignment From Pairwise Differences with a Noisy Oracle,” with C. Tsourakakis. In *Proceedings of the 15th Workshop on Algorithms and Models for the Web Graph*, pp. 59-69, 2018. Journal version: *Internet Mathematics*, at <https://doi.org/10.24166/im.06.2019>, December 2020.

“Reconciling Graphs and Sets of Sets,” with T. Morgan. In *Proceedings of the 37th Symposium on Principles of Database Systems*, pp. 33-47, 2018.

“Simulated Annealing for JPEG Quantization,” with M. Hopkins and S. Wagner-Carena. In *Proceedings of the 2018 IEEE Data Compression Conference*, p. 414.

“Adaptive Cuckoo Filters,” with S. Pontarelli and P. Reviriego. In *Proceedings of the ALENEX Conference*, pp. 36-47, 2018. Journal version: *ACM Journal of Experimental Algorithmics*, Article 1.1, March 2020.

“An Empirical Comparison of the Maximal and Total Information Coefficients to Leading Measures of Dependence,” with D. Reshef, Y. Reshef, and P. Sabeti. *Annals of Applied Statistics*, 12(1):123-155, 2018.

- “Simple Multi-Party Set Reconciliation,” with R. Pagh. *Distributed Computing* 31(6):441-453, 2018.
- “Technical Perspective: Building a Better Hash Function.” *Communications of the ACM*, 60(7), p. 93, 2017.
- “Compresso: Efficient Compression of Segmentation Data For Connectomics,” with D. Haehn, F. Lekschas, B. Matejek, and H. Pfister. In *Proceedings of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 781-788, 2017.
- “Auditable Data Structures,” with M. Goodrich, E. Kornaropoulos, and R. Tamassia. In *Proceedings of the European Symposium on Security and Privacy*, pp. 285-300, 2017.
- “2-3 Cuckoo Filters for Faster Triangle Listing and Set Intersection,” with D. Eppstein, M. Goodrich, and M. Torres. In *Proceedings of the 36th Symposium on Principles of Database Systems*, pp. 247-260, 2017.
- “Scalable motif-aware graph clustering,” with C. Tsourakakis and J. Pachocki. In *Proceedings of the 26th International Conference on the World Wide Web*, pp. 1451-1460, 2017.
- “Measuring Dependence Powerfully and Equitably,” with Y. Reshef, D. Reshef, H. Finucane, and P. Sabeti. *Journal of Machine Learning Research*, 17(212):1-63, 2016.
- “Quantized Random Projections and Non-Linear Estimation of Cosine Similarity,” with P. Li and M. Slawski. In *Advances In Neural Information Processing Systems 29*, pp.2748-2756, 2016.
- “Analyzing Distributed Join-Idle-Queue: A Fluid Limit Approach.” In *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, pp. 312-318, 2016.
- “Hardness of Peeling with Stashes,” with V. Nathan. *Information Processing Letters*, 116(11), pp. 682-688, 2016.
- “More Practical and Secure History-Independent Hash Tables,” with M. Goodrich, E. Kornaropoulos, and R. Tamassia. In *European Symposium on Research in Computer Security*, pp. 20-38, 2016.
- “Models and Algorithms for Graph Watermarking,” with D. Eppstein, M. Goodrich, J. Lam, N. Mamano, and M. Torres. In *Proceedings of the Information Security Conference*, pp. 283-301, 2016.
- “Better Bounds for Coalescing-Branching Random Walks on Graphs,” with R. Rajaraman and S. Roche. In *Proceedings of the 28th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pp. 313-323, 2016. Journal version: *ACM Transactions on Parallel Computing*, 5(1), 2:1-2:23, 2018.
- “OMASS: One Memory Access Set Separation,” with S. Pontarelli and P. Reviriego. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1940-1943, 2016.
- “Voronoi Choice Games”, with M. Boppana, R. Hod, and T. Morgan. In *Proceedings of 43rd International Colloquium on Automata, Languages, and Programming*, 23:1-23:13, 2016.

“Space Lower Bounds for Itemset Frequency Sketches,” with E. Liberty, J. Thaler, and J. Ullman. In *Proceedings of the 35th Symposium on Principles of Database Systems*, pp. 441-454, 2016.

“Technical Perspective: Catching Lies (and Mistakes) in Offloaded Computation,” with J. Thaler. *Communications of the ACM*, 59(2), p. 102, 2016.

“A New Approach to Analyzing Robin Hood Hashing.” In *Proceedings of ANALCO*, pp. 10-24, 2016.

“More Analysis of Double Hashing for Balanced Allocations.” In *Proceedings of ANALCO*, pp. 1-9, 2016.

“Theory Without Experiments: Have We Gone Too Far?” *Communications of the ACM*, 58(9), pp. 40-42, 2015.

“Scaling Up Clustered Network Appliances with XBricks,” with D. Zhou, B. Fan, H. Lim, R. Wang, M. Kaminsky, D. Andersen, and A. Singh. In *Proceedings of ACM SIGCOMM*, 2015.

“Scalable Large Near-Clique Detection in Large-Scale Networks via Sampling,” with J. Pachocki, R. Peng, C. Tsourakakis, and S. Chen Xu. In *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2015.

“Repeated Deletion Channels,” with B. Haeupler. In *Proceedings of the IEEE Information Theory Workshop*, pp. 152-156, 2014.

“Multi-Party Set Reconciliation Using Characteristic Polynomials,” with A. Boral. In *Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing*, pp. 1182-1187, 2014.

“Cleaning Up the Record on the Maximal Information Coefficient and Equitability,” with D. Reshef, Y. Reshef, and P. Sabeti. *Proceedings of the National Academy of Sciences* 111(33):E3362-3, 2014.

“Cuckoo Filter: Practically Better than Bloom,” with B. Fan, D.G. Andersen, and M. Kaminsky. In *Proceedings of the 10th International Conference on Emerging Networks Experiments and Technologies (CoNEXT)*, pp. 75-88, 2014.

“Coding for Random Projections,” with P. Li and A. Shrivastava. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 676-684, 2014.

“Balanced Allocations and Double Hashing.” In *Proceedings of the 26th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pp. 331-342, 2014.

“Parallel Peeling Algorithms,” with J. Jiang and J. Thaler. In *Proceedings of the 26th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pp. 319-330, 2014. Journal version: *ACM Transactions on Parallel Computing*, 3(1), 7, 2016.

“Wear Minimization for Cuckoo Hashing: How Not to Throw a Lot of Eggs into One Basket”, with D. Eppstein, M. Goodrich, and P. Pszona. In *Proceedings of the 13th International Symposium on Experimental Algorithms*, pp. 162-173, 2014.

- “Efficient Estimation for High Similarities using Odd Sketches,” with R. Pagh and N. Pham. In *Proceedings of the 23rd International Conference on the World Wide Web*, pp. 109-118, 2014.
- “Improving the Performance of Invertible Bloom Lookup Tables,” with S. Pontarelli and P. Reviriego. *Information Processing Letters*, 114(4), pp. 185-191, 2014.
- “Special Section on the Forty-First Annual ACM Symposium on Theory of Computing (STOC 2009),” with N. Immerlica, J. Katz, R. Servedio, and C. Umans. *SIAM Journal on Computing*, 41(6):1591-1592, 2012.
- “The Daily Deals Marketplace: Empirical Observations and Managerial Implications,” with J. Byers and G. Zervas. *AC SIGecom Exchanges*, 11(2):29-31, 2012.
- “Peeling Arguments and Double Hashing,” with J. Thaler. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1118-1125, 2012.
- “The Complexity of Object Reconciliation, and Open Problems Related to Set Difference and Coding,” with G. Varghese. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1126-1132, 2012.
- “Cache-Oblivious Dictionaries and Multimaps with Negligible Failure Probability,” with M. Goodrich, D. Hirschberg, and J. Thaler. In *Proceedings of the First Mediterranean Conference on Algorithms*, pp. 203-218, 2012.
- “An Economic Analysis of User-Privacy Options in Ad-Supported Services,” with J. Feigenbaum and G. Zervas. In *Proceedings of the 8th Workshop on Internet and Network Economics*, pp. 30-43, 2012.
- “Verifiable Computation with Massively Parallel Interactive Proofs,” with H. Pfister, M. Roberts, and J. Thaler. In *Proceedings of the USENIX HotCloud Workshop*, 2012.
- “Biff (Bloom Filter) Codes: Fast Error Correction for Large Data Sets,” with G. Varghese. In *Proceedings of the International Symposium on Information Theory*, pp. 483-487, 2012.
- “Continuous Time Channels with Interference,” with I. Ivan, J. Thaler, and H. Yuen. In *Proceedings of the International Symposium on Information Theory*, pp. 860-864, 2012.
- “The Groupon Effect on Yelp Ratings: a Root Cause Analysis,” with J. Byers and G. Zervas. In *Proceedings of the ACM Conference on Electronic Commerce*, pp. 248-265, 2012.
- “Anonymous Card Shuffling and Its Applications to Parallel Mixnets,” with M. Goodrich. In *Proceedings of 39th International Colloquium on Automata, Languages, and Programming*, pp. 549-560, 2012.
- “Chernoff-Hoeffding Bounds for Markov Chains: Generalized and Simplified,” with K. Chung, H. Lam, and Z. Liu. In *Proceedings of the 29th International Symposium on Theoretical Aspects of Computer Science*, pp. 124-135, 2012.
- “Practical oblivious storage,” with M. Goodrich, O. Ohrimenko and R. Tamassia. In *Proceedings of the Second ACM Conference on Data and Application Security and Privacy*, pp. 13-24, 2012.

- “Hierarchical Heavy Hitters with the Space Saving Algorithm,” T. Steinke and J. Thaler. In *Proceedings of 2012 Meeting on Algorithm Engineering and Experiments*, pp. 160-174.
- “Daily Deals: Prediction, Social Diffusion, and Reputational Ramifications,” with J. Byers and G. Zervas. In *Proceedings of the 5th International Conference on Web Search and Web Data Mining*, pp. 543-552, 2012.
- “Practical Verified Computation with Streaming Interactive Proofs,” with G. Cormode and J. Thaler. In *Proceedings of Innovations in Theoretical Computer Science*, pp. 90-112, 2012.
- “Information Dissemination via Random Walks in d -Dimensional Space,” with H. Lam, Z. Liu, X. Sun, and Y. Wang. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1612-1622, 2012.
- “Privacy-Preserving Group Data Access via Stateless Oblivious RAM Simulation,” with M. Goodrich and R. Tamassia. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 157-167, 2012.
- “An Efficient Rigorous Approach for Identifying Statistically Significant Frequent Itemsets,” with A. Kirsch, A. Pietracaprina, G. Pucci, E. Upfal, and F. Vandin. *Journal of the ACM*, 59(3), 2012.
- “Detecting Novel Associations in Large Datasets,” with D. Reshef, Y. Reshef, H. Finucane, S. Grossman, G. McVean, P. Turnbaugh, E. Lander, and P. Sabeti. *Science*, Vol. 3004, No. 6602, pp. 1518-1524, December 2011.
- “External-Memory Multimaps,” with E. Angelino, M. Goodrich, and J. Thaler. In *Proceedings of the 22nd International Symposium on Algorithms and Computation*, pp. 384-394, 2011. Journal version: *Algorithmica*, 67(1), pp. 23-48, 2013.
- “Brief announcement: large-scale multimaps,” with M. Goodrich. In *Proceedings of the 23rd ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pp. 259-260, 2011.
- “Oblivious RAM Simulation with Efficient Worst-Case Access Overhead,” with M. Goodrich, O. Ohrimenko, and R. Tamassia. In *Proceedings of 3rd ACM Cloud Computing Workshop*, pp. 95-100.
- “Invertible Bloom Lookup Tables,” with M. Goodrich. In *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 792-799, 2011.
- “Cuckoo Hashing with Pages,” with M. Dietzfelbinger and M. Rink. In *Proceedings of the 19th Annual European Symposium on Algorithms*, pp. 615-627, 2011.
- “Network Coding Meets TCP: Theory and Implementation,” with J.K. Sundararajan, D. Shah, M. Medard, S. Jakubczak, and J. Barros. *Proceedings of the IEEE*, 99(3), pp. 490-510, 2011.
- “Privacy-Preserving Access of Outsourced Data via Oblivious RAM Simulation,” with M. Goodrich. In *Proceedings of 38th International Colloquium on Automata, Languages, and Programming*, pp. 576-587, 2011.

- “Graption: A Graph-based P2P Traffic Classification Framework for the Internet Backbone,” with M. Iliofotou, H. Kim, M. Faloutsos, P. Pappu, and G. Varghese. *Computer Networks*, 55(8), pp. 1909-1920, 2011.
- “On the Zero-Error Capacity Threshold for Deletion Channels,” with I. Kash, J. Thaler, J. Ullman. In *Proceedings of the Information Theory and Applications Workshop*, pp. 1-5, 2011.
- “Heapable Sequences and Subsequences,” with J. Byers, B. Heeringa, and G. Zervas. In *Proceedings of ANALCO*, pp. 33-44, 2011.
- “An Introduction to Human Guided Search.” *ACM Crossroads*, 17(2):34-35, 2010.
- “Popularity is Everything: A New Approach to Protecting Passwords from Statistical Guessing Attacks,” with S. Schechter and C. Herley. In *Proceedings of the 5th Usenix Workshop on Hot Topics in Security*, pp. 1-8, 2010.
- “Streaming Graph Computations with a Helpful Advisor,” with G. Cormode and J. Thaler. In *Proceedings of the European Symposium on Algorithms*, pp. 231-242, 2010. Journal version: *Algorithmica*, 65(2), pp. 409-442, 2013.
- “Tight Thresholds for Cuckoo Hashing via XORSAT,” with M. Dietzfelbinger, A. Goerdt, A. Montanari, R. Pagh, and M. Rink. In *Proceedings of 37th International Colloquium on Automata, Languages, and Programming*, pp. 213-225, 2010.
- “Local cluster aggregation models of explosive percolation,” with R.M. D’Souza. *Physical Review Letters*, 14(19), p. 104-107, 2010.
- “Tight Asymptotic Bounds for the Deletion Channel with Small Deletion Probabilities,” with A. Kalai and M. Sudan. In *Proceedings of the International Symposium on Information Theory*, pp. 997-1001, 2010.
- “An Improved Analysis of the Lossy Difference Aggregator,” with H. Finucane. *Computer Communication Review*, 40(2), pp. 4-11, 2010.
- “Information Asymmetries in Pay-Per-Bid Auctions: How Swoopo Makes Bank,” with J. Byers and G. Zervas. In *Proceedings of the ACM Conference on Electronic Commerce*, pp. 1-12, 2010.
- “Carousel: Scalable Logging for Intrusion Prevention Systems”, with T. Lam and G. Varghese. In *Proceedings of NSDI 2010*, pp. 361-376.
- “AMS Without 4-Wise Independence on Product Domains,” with V. Braverman, K. Chung, Z. Liu, and R. Ostrovsky. In *Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science*, pp. 119-130, 2010.
- “Adaptive Weighing Designs for Keyword Value Computation,” with J. Byers and G. Zervas. In *Proceedings of the Third International Conference on Web Search and Web Data Mining* pp. 331-340, 2010.
- “Human-guided search,” with G. Klau, N. Lesh, and J. Marks. *Journal of Heuristics*, 16(3), pp. 289-310, 2010.

“Real-Time Parallel Hashing on the GPU,” with D. Alcantara, A. Sharf, F. Abbasinejad, S. Sengupta, J. Owens, and N. Amenta. *ACM Transactions on Graphics*, 28(5), p. 154, 2009.

“Exploiting Dynamicity in Graph-based Traffic Analysis: Techniques and Applications,” with M. Iliofotou and M. Faloutsos. In *Proceedings of ACM Co-NEXT*, pp. 241-252, 2009.

“An Analysis of Random-Walk Cuckoo Hashing,” with A. Frieze and P. Mellsted. In *Proceedings of 2009 APPROX-RANDOM*, pp. 490-503, 2009. Journal version: *SIAM Journal on Computing*, 40(2), pp. 291-308, 2011.

“Some Open Questions Related to Cuckoo Hashing.” In *Proceedings of the 17th Annual European Symposium on Algorithms*, pp. 1-10, 2009.

“On Compressing Social Networks,” with F. Chierichetti, R. Kumar, S. Lattanzi, A. Panconesi, and P. Raghavan. In *Proceedings of the 15th ACM SIGKDD Int’l Conference on Knowledge Discovery and Data Mining*, pp. 219-228, 2009.

“An Efficient Rigorous Approach for Identifying Statistically Significant Frequent Itemsets,” with A. Kirsch, A. Pietracaprina, G. Pucci, E. Upfal, and F. Vandin. In *Proceedings of the 28th Symposium on Principles of Database Systems*, pp. 117-126, 2009.

“Using the Power of Two Choices to Improve Bloom Filters,” with Steven Lumetta. *Internet Mathematics*, vol. 4. no. 1, pp. 17-33, 2009.

“Network coding meets TCP,” with J. Sundararajan, D. Shah, M. Medard, and J. Barros. In *Proceedings of IEEE INFOCOM*, pp. 280-288, 2009.

“An Economically Principled Generative Model of AS Graph Connectivity,” with J. Corbo, S. Jain, and D. Parkes. In *Proceedings of IEEE INFOCOM (Mini-conference)*, pp. 2941-2945, 2009.

“Designing floating codes for expected performance,” with H. Finucane and Z. Liu. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, 2008. Journal version (with F. Chierichetti, H. Finucane, and Z. Liu): *IEEE Transactions on Information Theory*, vol 56, Issue 3, pp. 968-978, 2010.

“On the performance of multiple choice hash tables with moves on deletions and inserts,” with A. Kirsch. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1284-1290, 2008.

“More Robust Hashing: Cuckoo Hashing with a Stash,” with A. Kirsch and U. Wieder. In *Proceedings of the 16th Annual European Symposium on Algorithms*, pp. 611-622, 2008. Journal version: *SIAM Journal on Computing*, 39(4), pp. 1543-1561, 2009.

“A survey of results for deletion channels and related synchronization channels.” In *Proceedings of the 2008 Scandinavian Workshop on Algorithm Theory*, pp. 1-3, 2008. Journal version: “New Results and Open Problem for Channels with Synchronization.” *Probability Surveys*, 2009.

“The Power of One Move: Hashing Schemes for Hardware,” with A. Kirsch. In *Proceedings of IEEE INFOCOM*, pp. 106-110, 2008. Journal version: *IEEE/ACM Transactions on Networking*, 18(6):1752-1765, 2010.

“Distributed Beamforming with Binary Signaling,” with M. Johnson and K. Ramchandran. In *Proceedings of the 2008 Int’l Symposium on Information Theory (ISIT)*, pp. 890-894, 2008.

“The Hiring Problem and Lake Wobegon Strategies,” with A. Broder, A. Kirsch, R. Kumar, E. Upfal, and S. Vassilvitskii. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1184-1193, 2008. Journal version: *SIAM Journal on Computing*, 39(4), pp. 1233-1255, 2009.

“Why Simple Hash Functions Work: Exploiting the Entropy in a Data Stream,” with S. Vadhan. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 746-755, 2008. Journal version: with K. Chung and S. Vadhan. *Theory of Computing*, 9, pp. 897-945-1255, 2013.

“Trace reconstruction with constant deletion probability and related results,” with T. Holenstein, R. Panigrahy, and U. Wieder. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 389-398, 2008.

“Capacity bounds for sticky channels.” *IEEE Transactions on Information Theory*, vol 54, Issue 1, pp. 72-77, 2008.

“Network Monitoring Using Traffic Dispersion Graphs,” with M. Iliofotou, P. Pappu, G. Varghese, M. Faloutsos, and S. Singh. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 315-320, 2007.

“HEXA: Compact Data Structures for Faster Packet Processing,” with S. Kumar, J. Turner, and P. Crowley. In *Proceedings of the IEEE International Conference on Network Protocols*, pp. 246-255, 2007.

“Using a Queue to De-amortize Cuckoo Hashing in Hardware,” with A. Kirsch and M. Mitzenmacher. In *Proceedings of the 45th Annual Allerton Conference on Communication, Control, and Computing*, 2007.

“Improved lower bounds for the capacity of i.i.d. deletion and duplication channels, with E. Drinea. *IEEE Transactions on Information Theory*, vol 53, Issue 8, pp. 2693-2714, 2007.

“Wired Geometric Routing,” with J. Ledlie, P. Pietzuch, and M. Seltzer. In *Proceedings of the 6th International Workshop on Peer-to-Peer Systems (IPTPS)*, 2007.

“Capacity Upper Bounds for Deletion Channels,” with S. Diggavi and H. Pfister. In *Proceedings of the 2007 Int’l Symposium on Information Theory (ISIT)*, pp. 1716-1720, 2007.

“Codes for Deletion and Insertion Channels with Segmented Errors,” with Z. Liu. In *Proceedings of the 2007 Int’l Symposium on Information Theory (ISIT)*, pp. 846-850, 2007. Journal version: *IEEE Transactions on Information Theory*, vol 56, Issue 1, pp. 224-232, 2010.

“Towards a theory of networked computation : Executive Summary,” with J. Feigenbaum. *ACM SIGACT News*, (4):22-26, 2006.

“Bloom Filters via d-Left Hashing and Dynamic Bit Reassignment,” with F. Bonomi, R. Panigrahy, S. Singh, and G. Varghese. In *Proceedings of the Allerton Conference*, pp. 877-883, 2006.

- “Beyond Bloom Filters: From Approximate Membership Checks to Approximate State Machines,” with F. Bonomi, R. Panigrahy, S. Singh, and G. Varghese. In *Proceedings of ACM SIGCOMM*, pp. 315-326, 2006.
- “An Improved Construction for Counting Bloom Filters,” with F. Bonomi, R. Panigrahy, S. Singh, and G. Varghese. In *Proceedings of the European Symposium on Algorithms*, pp. 684-695, 2006.
- “Less Hashing, Same Performance: Building A Better Bloom Filter,” with A. Kirsch. In *Proceedings of the European Symposium on Algorithms*, pp. 456-467, 2006. Journal version: *Random Structures and Algorithms*, 33(2), pp. 187-218, 2008.
- “Stochastic Shortest Paths via Quasi-Convex Maximization,” with M. Brand, J. Kelner, and E. Nikolova. In *Proceedings of the European Symposium on Algorithms*, pp. 552-563, 2006.
- “Polynomial Time Low-Density Parity-Check Codes with Rates Very Close to the Capacity of the q -ary Random Deletion Channel for Large q .” *IEEE Transactions on Information Theory*, vol 52, Issue 12, pp. 5496-5501, 2006.
- “A Simple Lower Bound for the Capacity of the Deletion Channel,” with E. Drinea. *IEEE Transactions on Information Theory*, vol 52, Issue 10, pp. 4657-4660, 2006.
- “On Lower Bounds for the Capacity of Deletion Channels,” with E. Drinea. *IEEE Transactions on Information Theory*, vol 52, Issue 10, pp. 4648-4657, 2006.
- “On the Theory and Practice of Data Recovery with Multiple Versions.” In *Proceedings of the International Symposium on Information Theory*, pp. 982-986, 2006.
- “Fine-Grained Layered Multicast with STAIR,” with J. Byers, Gu-In Kwon, and M. Luby. *IEEE/ACM Transactions on Networking*, 14:1, pp. 81-93, 2006.
- “Distance-Sensitive Bloom Filters,” with A. Kirsch. In *Proceedings of the ALENEX Conference*, pp. 41-50, 2006.
- “BubbleSearch: A Simple Heuristic for Improving Priority-Based Greedy Algorithms,” with N. Lesh. *Information Processing Letters*, 97(4), pp. 161-169, 2006.
- “Network-Aware Overlays with Network Coordinates,” with P. Pietzuch, J. Ledlie, and M. Seltzer. In *Proceedings of the International Workshop on Dynamic Distributed Systems*, 2006.
- “Editorial: The Future of Power Law Research.” *Internet Mathematics*, vol. 2. no. 4, pp. 525-534, 2006.
- “X-Tolerant Test Response Compaction,” with S. Mitra, S. Lumetta, and N. Patil. *IEEE Design and Test of Computers*, vol. 22, no. 6, pp. 566-574, Nov/Dec 2005.
- “Privacy preserving keyword searches on remote encrypted data,” with Y.-C. Chang. In *Proceedings of the Third Intl. Conference on Applied Cryptography and Network Security*, pp. 442-455, 2005.
- “Simple Summaries for Multiple Choice Hashing,” with A. Kirsch. In *Proceedings of the 43rd Annual Allerton Conference on Communication, Control, and Computing*, 2005. Journal version: *IEEE/ACM Transactions on Networking*, 16(1), pp. 218-231, 2008.

- “A case study in large-scale interactive optimization,” with M. Chimani, N. Lesh, and C. Sidner. In *Proceedings of Artificial Intelligence and Applications 2005*, pp. 24-29.
- “New heuristic and interactive approaches to 2D rectangular strip packing,” with N. Lesh, A. McMahon, and J. Marks. *ACM Journal of Experimental Algorithms*, 10, 2005.
- “Multidimensional Balanced Allocations,” with A. Broder. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 195-196, 2005.
- “Digital Fountains: A Survey and Look Forward.” 2004 Information Theory Workshop.
- “Improved Lower Bounds on the Capacity of I.I.D. Deletion Channels,” with E. Drinea. In *Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.
- “Improved Lower Bounds on the Capacity of Deletion Channels,” with E. Drinea. In *Proceedings of the 2004 IEEE International Symposium on Information Theory*.
- “Geometric Generalizations of the Power of Two Choices,” with J. Byers and J. Conside. In *Proceedings of the 16th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pp. 54-63, 2004.
- “Interactive Data Summarization: An Example Application,” with N. Lesh. In *Proceedings of the 2004 Workshop on Advanced Visual Interfaces*, pp. 183-187.
- “On Power Law Distributions for Monkeys Typing Randomly: The Case of Unequal Probabilities,” with B. Conrad. *IEEE Transactions on Information Theory*, 50:7, pp. 1403-1414, 2004.
- “Dynamic Models for File Sizes and Double Pareto Distributions.” *Internet Mathematics*, 1(3):305-334, 2003.
- “A Scaling Result for Explosive Processes,” with R. Oliveira and J. Spencer. *Electronic Journal of Combinatorics*, 11(1), R31, 2004.
- “An Exhaustive Approach to 2D Strip Packing,” with N. Lesh, A. McMahon, and J. Marks. *Information Processing Letters*, 90(1), pp. 7-14, 2004.
- “New Models and Methods for File Size Distributions,” with B. Twestzky. In *Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing*, pp. 603-612, 2003.
- “Capacity Approaching Signal Constellations for Channels with Memory,” with A. Kavčić, and X. Ma. *IEEE Transactions on Information Theory*, 49:7, pp. 1636-1652, 2003.
- “The Power Spectra of Good Codes for Partial Response Channels,” with A. Kavcic and X. Ma. In *Proceedings of the 2003 IEEE International Symposium on Information Theory*, p. 45.
- “Concatenated Codes for Deletion Channels,” with J. Chen, C. Ng, and N. Varnica. In *Proceedings of the 2003 IEEE International Symposium on Information Theory*, p. 218.
- “Verification Codes for Deletion Channels.” In *Proceedings of the 2003 IEEE International Symposium on Information Theory*, p. 217.

- “Estimating and Comparing Entropies Across Written Natural Languages Using PPM Compression,” with F. Behr, V. Fossum, and D. Xiao. In *Proceedings of the 2003 IEEE Data Compression Conference*, p. 416.
- “Simple Load Balancing for Distributed Hash Tables,” with J. Byers, and J. Considine. Appears in the 2nd International Workshop on Peer-to-Peer Systems (IPTPS), pp. 80-87, 2003.
- “A Complete and Effective Move Set for Simplified Protein Folding,” with N. Lesh, and S. Whitesides. In *Proceedings of the 7th Annual International Conference on Research in Computational Molecular Biology*, pp 188-195, 2003.
- “Network Applications of Bloom Filters: A Survey,” with A. Broder. In *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, pp. 636-646, 2002. Journal version: *Internet Mathematics*, 1(4):485-509, 2003.
- “Verification Codes,” with M. Luby. In *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, pp. 38-47, 2002. Journal version: *IEEE Transactions on Information Theory*: 51(1), pp. 120-127, 2005.
- “Balls and Bins with Memory,” with B. Prabhakar and D. Shah. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pp. 799-808, 2002.
- “Informed Content Delivery Across Adaptive Overlay Networks,” with J. Byers, J. Considine, and S. Rost. In *Proceedings of ACM SIGCOMM '02*, available as *Computer Communications Review*, vol. 32:4, pp. 47-60, 2002. Journal version: *IEEE/ACM Transactions on Networking*, 12:5, pp. 767-780, 2004.
- “Optimal Plans for Aggregation,” with A. Broder. In *Proceedings of the 21st Annual ACM Symposium on Principles of Distributed Computing*, pp. 144-152, 2002.
- “Human-Guided Tabu Search,” with G. Klau, N. Lesh, and J. Marks. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pp. 41-47, 2002.
- “The HuGS Platform: A Toolkit for Interactive Optimization,” with G. Klau, N. Lesh, J. Marks, and G. Schafer. In *Proceedings of the Workshop on Advanced Visual Interfaces*, pp. 324-330, 2002.
- “Exact Sampling of TCP Window States,” with A. Goel. In *Proceedings of IEEE INFOCOM*, pp. 259-265, 2002.
- “Balls and Bins Models with Feedback,” with E. Drinea and A. Frieze. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 308-315, 2002.
- “A Brief History of Lognormal and Power Law Distributions.” In *Proceedings of the 39th Annual Allerton Conference on Communication, Control, and Computing*, pp. 182-191, 2001. Journal version: *Internet Mathematics*, 1(2):226-251, 2003.
- “Capacity Approaching Signal Constellations for Channels with Memory,” with A. Kavčić, X. Ma, and N. Varnica. In *Proceedings of the 39th Annual Allerton Conference on Communication, Control, and Computing*, pp. 311-320, 2001.
- “Compressed Bloom Filters.” In *Proceedings of the 20th Annual ACM Symposium on Principles of Distributed Computing*, pp. 144-150, 2001. Journal version: *IEEE/ACM Transactions on Networks*, 10:5, pp. 613-620, October 2002.

“Towards More Completed Models of TCP Throughput,” with R. Rajaraman. *Journal of Supercomputing*, 20(2), pp. 137-160, September 2001.

“Deriving Performance Bounds for ISI Channels Using Gallager Codes,” with A. Kavčić, B. Marcus, and B. Wilson. In *Proceedings of the 2001 International Symposium on Information Theory*, p. 345, 2001.

“Toward Compressing Web Graphs,” with M. Adler. In *Proceedings of the 2001 Data Compression Conference*, pp. 203-212, 2001.

“On the Hardness of Finding Multiple Preset Dictionaries.” In *Proceedings of the 2001 Data Compression Conference*, pp. 411-418, 2001. Journal version: *IEEE Transactions on Information Theory*, 50:7, pp. 1536-1539, 2004.

“Fine-Grained Layered Multicast,” with J. Byers and M. Luby. In *Proceedings of IEEE INFOCOM 2001*, pp. 1143-1151, 2001. Journal version: “Fine-Grained Layered Multicast with STAIR,” with J. Byers, Gu-In Kwon, and M. Luby. To appear in *IEEE/ACM Transactions on Networking*.

“Using Multiple Hash Functions to Improve IP Lookups,” with A. Broder. In *Proceedings of IEEE INFOCOM 2001*, pp. 1454-1463, 2001.

“Improved Results for Route Planning in Stochastic Transportation Networks,” with J. Boyan. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 895-902, 2001.

“Estimating Resemblance of MIDI Documents,” with S. Owen. In *Proceedings of the 3rd Workshop on Algorithm Engineering and Experiments*, available as *Lecture Notes in Computer Science 2153*, pp. 78-90, 2001.

“Linear Waste of Best Fit Bin Packing on Skewed Distributions,” with Claire Kenyon. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pp. 582-589, 2000. Journal version: invited to special issue of *Random Structures and Algorithms*, 20(3), pp. 441-464, 2002.

“FLID/DL: Congestion Control for Layered Multicast,” with J. Byers, M. Frumin, G. Horn, M. Luby, A. Roetter, and W. Shaver. In *Proceedings of the 2nd International Workshop on Networked Group Communication (NGC 2000)*, pp. 71-81, 2000. Journal version: with J. Byers, G. Horn, M. Luby, and W. Shaver. *IEEE Journal on Selected Areas in Communications*, 20:8, pp. 1558-1570, October 2002.

“On Near-Uniform URL Sampling,” with M. Henzinger, A. Heydon, and M. Najork. In *Proceedings of the Ninth International World Wide Web Conference*, pp. 295-308, 2000.

“An Extension of Path Coupling and Its Application to the Glauber Dynamics for Path Coupling,” with M. Dyer, L. Goldberg, C. Greenhill, and M. Jerrum. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 616-623, 2000. Journal version: *SIAM Journal on Computing*, 30(6), pp. 1962-1975, 2001.

“Improved Classification via Connectivity Information,” with A. Broder and R. Krauthgamer. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 576-585, 2000.

“The Asymptotics of Selecting the Shortest of Two, Improved,” with B. Vöcking. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 326-327, 1999. Extended version appears as a chapter in *Analytic Methods in Applied Probability: In Memory of Fridrikh Karpelevich*, American Mathematical Society, pp. 165-176, 2002.

“Completeness and Robustness Properties of Min-Wise Independent Permutations,” with A. Broder. In *Proceedings of Random '99*, available as *Lecture Notes in Computer Science 1671*, pp. 1-10, 1999. Journal version: *Random Structures and Algorithms*, 18(1), pp. 18-30, January 2001.

“Measuring Index Quality Using Random Walks on the Web,” with M. Henzinger, A. Heydon, and M. Najork. In *Proceedings of the Eighth International World Wide Web Conference*, pp. 213-225, Elsevier Science, 1999.

“Analysis of Timing Based Mutual Exclusion with Random Times,” with E. Gafni. In *Proceedings of the 18th Annual ACM Symposium on Principles of Distributed Computing*, pp. 13-21, 1999. Journal version: *SIAM Journal on Computing*, 31:3, pp. 816-837, 2001.

“Accessing Multiple Mirror Sites in Parallel: Using Tornado Codes to Speed Up Downloads,” with J. Byers and M. Luby. In *Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, pp. 275-284, 1999.

“Unscrambling Address Lines (Abstract),” with A. Broder and L. Moll. In *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 870-871, 1999.

“A Derandomization Using Min-Wise Independent Permutations,” with A. Broder and M. Charikar. In *Proceedings of Random '98*, available as *Lecture Notes in Computer Science 1518*, pp. 15-24, October 1998. Journal version: *Journal of Discrete Algorithms*, 1:1, pp. 11-20, 2003.

“On Balls and Bins with Deletions,” with R. Cole, A. Frieze, B. M. Maggs, A. W. Richa, R. Sitaraman, and E. Upfal. In *Proceedings of Random '98*, available as *Lecture Notes in Computer Science 1518*, pp. 145-158, October 1998.

“Studying Balanced Allocations with Differential Equations.” *Combinatorics, Probability, and Computing*, vol. 8, pp. 473-482, 1999.

“Delayed Information and Action in On-Line Algorithms,” with S. Albers and M. Charikar. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS-98)*, pp. 71-80, November 1998. Journal version: *Information and Computation*, 170, pp. 135-152, 2001.

“A Digital Fountain Approach to the Reliable Distribution of Bulk Data,” with John Byers, Michael Luby, and Ashu Rege. In *Proceedings of ACM SIGCOMM '98*, available as *Computer Communications Review*, vol. 28:4, pp. 56-67, 1998. Journal version: *IEEE Journal on Selected Areas in Communication*, 20:8, pp. 1528-1540, October 2002.

“Analyses of Load Stealing Models Using Differential Equations.” In *Proceedings of the 10th Annual ACM Symposium on Parallel Algorithms and Architectures*, pp. 212-221, 1998. Journal version: *Theory of Computing Systems*, 34:1, pp. 77-98, January 2001.

“Load Adaptive Routing on Butterfly Networks,” with R. Cole, B. Maggs, F. Meyer auf der Heide, A. Richa, K. Schröder, R. Sitaraman, and B. Vöcking. In *Proceedings of the*

30th Annual ACM Symposium on Theory of Computing (STOC-98), pp. 378-388, ACM Press, May 23-26, 1998.

“Min-Wise Independent Permutations,” with Andrei Broder, Moses Charikar, and Alan Frieze. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC-98)*, pp. 327-336, ACM Press, May 23-26, 1998. Journal version: *Journal of Computer and System Sciences*, 60(3), pp. 630-659, 2000.

“Analysis of Low Density Codes and Improved Designs Using Irregular Graphs,” with Michael Luby, Amin Shokrollahi, and Dan Spielman. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC-98)*, pp. 249-258, ACM Press, May 23-26, 1998.

“Improved Low Density Parity Check Codes Using Irregular Graphs and Belief Propagation,” with Michael Luby, Amin Shokrollahi, and Dan Spielman. In *Proceedings of the 1998 International Symposium on Information Theory*. Extended version available as Compaq Systems Research Center Technical Note 1998-009. Journal version: “Improved Low Density Parity Check Codes Using Irregular Graphs,” *IEEE Transactions on Information Theory*, 47(2), pp. 585-598, 2001.

“Analysis of Random Processes via And-Or Tree Evaluation,” with Michael Luby and Amin Shokrollahi. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 364-373, 1998.

“Average Case Analyses of First Fit and Random Fit Bin Packing,” with Susanne Albers. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 290-299, 1998. Journal version: *Random Structures and Algorithms*, 16(3), pp. 240-259, 2000.

“How Useful is Old Information?” In *Proceedings of the 16th ACM Symposium on Principles of Distributed Computing*, pp. 83-91, 1997. Journal version: *IEEE Transactions on Parallel and Distributed Systems*, 11:1, pp. 6-20, January 2000.

“Revisiting the COUNTER Algorithms for List Update,” with Susanne Albers. *Inf. Proc. Letters*, 64, pp. 155-160, 1997.

“Practical Loss-Resilient Codes,” with Michael Luby, Amin Shokrollahi, Dan Spielman, and Volker Stemann. In *Proceedings of the 29th ACM Symposium on the Theory of Computing*, pp. 150-159, 1997. Journal version: “Efficient Erasure Correcting Codes,” with Michael Luby, Amin Shokrollahi, and Dan Spielman. *IEEE Transactions on Information Theory*, 47(2), pp 569-584, 2001.

“On the Analysis of Randomized Load Balancing Schemes.” In *Proceedings of the 9th ACM Symposium on Parallel Algorithms and Architectures*, pp. 292-301, 1997. Journal version: special conference issue of *Theory of Computing Systems*, vol. 32, pp. 361-386, 1999.

“Load Balancing and Density Dependent Jump Markov Processes.” In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pp. 213-222, 1996. Journal version: “The Power of Two Choices in Randomized Load Balancing.” *IEEE Transactions on Parallel and Distributed Computing*, 12(10), pp. 1094-1104, 2001.

“Designing Stimulating Programming Assignments for an Algorithms Course: A Collection of Problems Based on Random Graphs.” In *SIGCSE Bulletin*, 28:3, pp. 29-36, September 1996.

“Constant Time per Edge is Optimal in Rooted Tree Networks.” In *Proceedings of the 8th ACM Symposium on Parallel Algorithms and Architectures*, pp. 162-169, 1996. Journal version: *Distributed Computing* vol. 10:4, 1997, pp. 189-197.

“Average Case Analyses of List Update Algorithms, with Applications to Data Compression,” with Susanne Albers. In *Proceedings of the 1996 International Colloquium on Automata, Languages, and Programming*, pp. 514-525. Journal version: *Algorithmica*, vol. 21:3, 1998, pp. 312-329.

“Pattern-based Compression of Text Images,” with Andrei Broder. In *1996 Data Compression Conference*, pp. 300-309.

“Parallel Randomized Load Balancing,” with Micah Adler, Soumen Chakrabarti, and Lars Rasmussen. In *Proceedings of the 27th ACM Symposium on the Theory of Computing*, pp. 238-247, 1995. Journal version: *Random Structures and Algorithms*, vol. 13:2, September 1998, pp. 159-188.

“Bounds on the Greedy Routing Algorithm in Array Networks.” In *Proceedings of the 6th ACM Symposium on Parallel Algorithms and Architectures*, pp. 346-353, 1994. Journal version: special conference issue of the *Journal of Computer and System Sciences* vol. 53:3, December 1996, pp. 317-327.

“Computational Complexity of Loss Networks,” with Graham Louth and Frank Kelly. *Theoretical Computer Science*, vol. 125, May 1994, pp. 45-59.

Other Publications

“Lyra: An Efficient and Expressive Subquadratic Architecture for Modeling Biological Sequences,” with Krithik Ramesh, Sameed Siddiqui, Albert Gu, and Pardis Sabeti. Available at <https://arxiv.org/abs/2503.16351>.

“Juniper: Reconstructing Transmission Events from Next-Generation Sequencing Data at Scale,” with Ivan Specht, Gage Moreno, Taylor Brock-Fisher, Lydia Kraslinkova, Brittany Petros, Jonathan Pekar, Mark Schifferli, Ben Fry, Catherine Brown, Lawrence Madoff, Megan Burns, Stephen Schaffner, Daniel Park, Bronwn MacInnis, Al Ozonoff, Patrick Varilly, and Pardis Sabeti. Available at <https://www.medrxiv.org/content/10.1101/2025.03.02.25323192v1>.

“Delphy: scalable, near-real-time Bayesian phylogenetics for outbreaks,” with Patrik Varilly, Mark Schifferli, Katherine Yang, Tim Burcham, Paul Cronan, Olivia Glennon, Olivia Jacks, Ellory Laning, Libby Marrs, Kyle Oba, Shannon Yeung, Edyth Parker, Ifeanyi Omah, Jonathan E. Pekar, Laura Luebbert, Kristian G. Andersen, Daniel J. Park, Stephen F. Schaffner, Bronwyn L. MacInnis, Christian Happi, Jacob E. Lemieux, Al Ozonoff, Michael D. Mitzenmacher, Ben Fry, and Pardis C. Sabeti. Available at <https://www.biorxiv.org/content/10.1101/2025.03.25.645253v1>.

“Use of popularity information to reduce risk posed by guessing attacks,” with Stuart Schechter and Cormac Herley. US Patent 08959644. Issued 2/17/2015.

“Method and apparatus providing network coding based flow control,” with Jay Kumar Sundararajan, Devavrat Shah, Muriel Medard, Joao Barros, and Szymon Jakubczak. US Patent 08526451. Issued 9/3/2013.

“Method and apparatus providing network coding based flow control,” with Jay Kumar Sundararajan, Devavrat Shah, Muriel Medard, Joao Barros, and Szymon Jakubczak. US Patent 08130776. Issued 3/6/2012.

“Method for finding optimal paths using a stochastic network model,” with Evdokia Nikolova and Matthew Brand. US Patent 07573866. Issued 8/11/2009.

“Ranking search engine results,” with Monika Henzinger. US Patent 07541388. Issued 11/11/2008.

“Distributed, compressed Bloom filter Web cache server.” US Patent 06920477. Issued 7/19/2005.

“Method for packing rectangular strips,” with Neal Lesh and Joe Marks. US Patent 06832129. Issued 12/14/2004.

“System and method for near-uniform sampling of web page addresses,” with Monika Henzinger, Allan Heydon, and Marc Najork. US Patent 06594694. Issued 7/15/2003.

“On demand encoding with a window,” with Michael Luby, Gavin Horn, Jeffrey Persch, John Byers, and Armin Haken. US Patent 06486803. Issued 11/26/2002.

“Generating high weight encoding symbols using a basis,” with Armin Haken, Michael Luby, Gavin Horn, Diane Hernek, and John Byers. US Patent 06411223. Issued 6/25/2002.

“Space-efficient multi-cycle barrel shifter circuit,” with Laurent Moll. US Patent 06314156. Issued 11/6/2001.

“Method for determining a random permutation of variables by applying a test function,” with Andrei Broder, Laurent Moll, and Mark Shand. US Patent 06292762. Issued 9/18/2001.

“Loss resilient code with double heavy tailed series of redundant layers,” with Michael Luby. US Patent 06195777. Issued 2/27/2001.

“Message encoding with irregular graphing,” with Michael Luby, Amin Shokrollahi, Dan Spielman, and Volker Stemann. US Patent 06163870. Issued 12/19/2000.

“Compression of grey scale images of text,” with Andrei Broder. US Patent 06088309. Issued 7/11/2000.

“Irregularly graphed encoding technique,” with Michael Luby, Amin Shokrollahi, Dan Spielman, and Volker Stemann. US Patent 06081909. Issued 6/27/2000.

“Loss resilient decoding technique,” with Michael Luby, Amin Shokrollahi, Dan Spielman, and Volker Stemann. US Patent 06073250. Issued 6/6/2000.

“Task processing optimization in a multiprocessor system,” with Andrei Broder. US Patent 05991808. Issued 11/23/1999.

“Compression protocol with multiple preset dictionaries,” with Andrei Broder and Jeff Mogul. US Patent 05953503. Issued 9/14/1999.

“When Simple Hash Functions Suffice,” with K. Chung and S. Vadhan. Appears as a chapter in *Beyond Worst Case Analysis*, edited by Tim Roughgarden, Cambridge University Press.

“Algorithms with Predictions,” with S. Vassilvitskii. Appears as a chapter in *Beyond Worst Case Analysis*, edited by Tim Roughgarden, Cambridge University Press.

“Some Practical Randomized Algorithms and Data Structures,” with E. Upfal. Appears as a chapter in *Computing Handbook, Third Edition: Computer Science and Software Engineering*, edited by T. Gonzalez, J. Diaz-Herrera, and A. Tucker. Chapman and Hall, 2014.

“Building an Efficient Hash Table on the GPU,” with D. Alcantara, V. Volkov, S. Sengupta, J. Owens, and N. Amenta. Appears as a chapter in *GPU Computing Gems: Jade Edition*, edited by Wen-mei W. Hwu. Morgan Kaufmann, 2011.

“Codes – Protecting Data Against Loss and Errors.” Appears as a chapter in *Algorithms Unplugged*, edited by B. Vöcking, M. Dietzfelbinger, R. Reischuk, and H. Vollmer. Springer, 2010.

“Hash-Based Techniques for High-Speed Packet Processing”, with A. Kirsch and G. Varghese. Appears as a chapter in *Algorithms for Next Generation Networks*, edited by G. Cormode and M. Thottan. Springer, 2010.

“Bloom Filters.” Encyclopedia of Database Systems, 2009.

“Unbiasing Random Bits.” *Dr. Dobbs Journal*, No. 335, pp. 101-104, April 2002.

“Good Hash Tables & Multiple Hash Functions.” *Dr. Dobbs Journal*, No. 336, May 2002.

“Challenging Students with Creative Assignments.” *SIGACT News*, 32(1):70-73, 2001.

“An Experimental Assignment on Random Processes.” *SIGACT News*, 32(1):74-78, 2001.

“The Power of Two Random Choices: A Survey of Techniques and Results,” with A. Richa and R. Sitaraman. Appears as a chapter in *Handbook of Randomized Computing: Volume 1*, edited by P. Pardalos, S. Rajasekaran, and J. Rolim, pp. 255-312.

“Tornado Codes for Forward Error Correction,” with Michael Luby, in *Digital Forefront*.

“Bounds and Improvements for BiBa Signature Schemes,” with A. Perrig. Harvard Computer Science Group Technical Report TR-02-02.

“A Note on Low Density Parity Check Codes for Erasures and Errors.” Digital SRC Technical Note 1998-017.

“Tight Thresholds for the Pure Literal Rule.” Digital SRC Technical Note 1997-011.

“The Power of Two Choices in Randomized Load Balancing.” Ph.D. thesis, U.C. Berkeley, August 1996.

“*Calculus: A Solution Manual*,” with others, Harcourt Brace Jovanovich, 1993.

“*Elliptic Curves in Computer Science: Primality Testing, Factoring, and Cryptography*,” senior honors thesis, Harvard University, 1991. Published as a Hoopes Prize winner.

Conference Presentations “Queueing, Predictions, and LLMs: Challenges and Open Problems.” Tutorial, SIGMETRICS. June 2025.

“SkipPredict: When to Invest in Predictions for Scheduling. Presented at the TTIC Workshop on Learning-Augmented Algorithms, August 2024.

“Algorithms with Predictions.” Invited talk presented at the INFORMS 2023 Annual Meeting, October 2023.

“Algorithms with Predictions.” Presented at the 2nd Vienna Workshop on Computational Optimization. March 2023.

“Algorithms with Predictions.” Presented at the Sublinear Algorithms Workshop, FODSI/MIT. August 2022.

“Algorithms with Predictions: State of the Area.” Presented at the Algorithms with Predictions Workshop at the 54th Annual ACM Symposium on Theory of Computing. June 2022. Also panel lead.

“Efficient Randomized Network Telemetry Algorithms.” Presented at the Data plane algorithms in programmable networks tutorial at ACM SIGMETRICS 2022.

“Scheduling with Predictions and the Price of Misprediction.” Presented at the 11th Annual Innovations in Theoretical Computer Science Conference. January 2020.

“Learning-Augmented Algorithms: How ML Can Lead to Provably Better Algorithms.” ALGO Keynote Talk, September 2019.

“Robust Set Reconciliation via Locality Sensitive Hashing.” Presented at the 38th Symposium on Principles of Database Systems, July 2019.

“Arithmetic Progression Hypergraphs: Examining the Second Moment Method.” Presented at the 2019 ANALCO Workshop, January 2019.

“A Model for Learned Bloom Filters and Optimizing by Sandwiching.” Presented at the NeurIPS 2018 Poster Session.

“Adaptive Cuckoo Filters.” Presented at the 2018 ALENEX Conference, January 2018.

“Analyzing Distributed Join-Idle-Queue: A Fluid Limit Approach.” Presented at the 54th Annual Allerton Conference on Communication, Control, and Computing, October 2016.

“Voronoi Choice Games.” Presented at the 2016 International Colloquium on Automata, Languages, and Programming, July 2016.

“A New Approach to Analyzing Robin Hood Hashing.” Presented at the 2016 ANALCO Workshop, January 2016.

“More Analysis of Double Hashing for Balanced Allocations.” Presented at the 2016 ANALCO Workshop, January 2016.

“Data exploration: equitability and the maximal information coefficient.” Presented at the Workshop for Data-driven Algorithmics, September 2015.

“Multi-Party Set Reconciliation Using Characteristic Polynomials.” Presented at the 2014 Allerton Conference, October 2014.

“Parallel Peeling Algorithms.” Presented at the 2014 Symposium on Parallelism in Algorithms and Architectures, June 2014.

“Balanced Allocations and Double Hashing.” Presented at the 2014 Symposium on Parallelism in Algorithms and Architectures, June 2014.

“Peeling Arguments and Double Hashing.” Presented at the 2012 Allerton Conference, October 2012.

“The Complexity of Object Reconciliation and Open Problems Related to Set Difference and Coding.” Presented at the 2012 Allerton conference, October 2012.

“The Ecology of Communication, Control, and Computing Research and Education (Panel).” Presented at the 2012 Allerton conference, October 2012.

“Peeling Arguments: Invertible Bloom Lookup Tables and Biff Codes.” Presented at the 2012 Workshop on Algorithms for Modern Massive Data Sets (MMDS), July 2012.

“Invertible Bloom Lookup Tables.” Presented at the 2011 Allerton conference, September 2011.

“Brief Announcement: Large-Scale Multimaps.” Presented at the 2011 Symposium on Parallel Algorithms and Architectures, June 2011.

“On the Zero-Error Capacity Threshold for Deletion Channels.” Presented at the 2011 ITA Workshop, February 2011.

“Some Recent Uses of Hashing for Network Applications.” Presented at the 2010 Workshop on Research and Analysis of Tail Phenomena, August 2010.

“Some Recent Uses of Hashing for Network Applications.” Presented at the 2010 Workshop on Random Graphs and Algorithms, June 2010.

“Tight Asymptotic Bounds for the Deletion Channel with Small Deletion Probabilities.” Presented at the 2010 International Symposium on Information Theory, June 2010.

“Some Open Questions on Cuckoo Hashing.” Presented at the 2009 European Symposium on Algorithms, August 2009.

“On the Performance of Multiple Choice Hash Tables with Moves on Deletions and Inserts.” Presented at the 2008 Allerton conference, September 2008.

“Designing Floating Codes for Expected Performance.” Presented at the 2008 Allerton conference, September 2008.

“A Survey of Results for Deletion Channels and Related Synchronization Channels.” Presented at the 11th Scandinavian Workshop on Algorithm Theory, July 2008.

“Why Simple Hash Functions Work: Exploiting the Entropy in a Data Stream.” Presented at the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, January 2008.

“Using a Queue to De-amortize Cuckoo Hashing in Hardware.” Presented at the 2007 Allerton conference, September 2007.

“Capacity Upper Bounds for Deletion Channels.” Presented at the 2007 Int’l Symposium on Information Theory (ISIT), June 2007.

“Codes for Deletion and Insertion Channels with Segmented Errors.” Presented at the 2007 Int’l Symposium on Information Theory (ISIT), June 2007.

“Digital Fountains: Applications and Related Issues.” Presented at the INFORMS Annual Meeting, November 2006.

“New Directions for Power Law Research.” Presented at the Radcliffe Exploratory Seminar on Dynamic Networks, October 2006.

“Bloom Filters via d -left Hashing and Dynamic Bit Reassignment.” Presented at the 2006 Allerton conference, September 2006.

“Beyond Bloom Filters: From Approximate Membership Checks to Approximate State Machines.” Presented at the 2006 SIGCOMM conference, September 2006.

“An Improved Construction for Counting Bloom Filters.” Presented at the European Symposium on Algorithms, September 2006.

“Building a Better Bloom Filter.” Presented at the European Symposium on Algorithms, September 2006.

“On the Theory and Practice of Data Recovery with Multiple Versions.” Presented at the 2006 International Symposium on Information Theory, July 2006.

“New Results for an Old Data Structure: Bloom Filters and Beyond.” Presented at the 2006 Stochastic Networks conference, June 2006.

“New Results and Open Problems for Deletion Channels.” Presented at ANALCO 2006, January 2006 (plenary talk).

“Digital Fountains and their Application to Informed Content Delivery over Adaptive Overlay Networks.” Presented at 19th International Symposium on Distributed Computing, September 2005 (plenary talk).

“Simple Summaries for Multiple Choice Hashing.” Presented at the 2005 Allerton Conference.

“New Directions in Power Law Research.” MSRI Workshop on Models of Real-World Random Networks, April 2005.

“Multi-dimensional Balanced Allocations.” Presented at the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, January 2005.

“Digital Fountains: A Survey and Look Forward.” Presented at the 2004 Information Theory Workshop. October 2004.

“Improved Lower Bounds on the Capacity of I.I.D. Deletion Channels.” Presented at the 2004 Allerton Conference.

“Geometric Generalizations of the Power of Two Choices.” Presented at the 16th ACM Symposium on Parallel Algorithms and Architectures, June 2004.

“New Models and Methods for File Size Distributions.” Presented at the 40th Annual Allerton Conference on Communication, Control, and Computing, October 2003.

“Verification Codes for Deletion Channels.” Presented at the International Symposium on Information Theory, July 2003.

“New Exhaustive, Heuristic, and Interactive Algorithms for 2-Dimensional Bin-Packing.” Presented at the 2nd Annual Workshop on Random Graphs and Algorithms, June 2003.

“Network Applications of Bloom Filters: A Survey.” Presented at the 40th Annual Allerton Conference on Communication, Control, and Computing, October 2002.

“Verification Codes.” Presented at the 40th Annual Allerton Conference on Communication, Control, and Computing, October 2002.

“New Directions in Balls and Bins Problems.” Invited talk for the 6th International Workshop on Randomization and Approximation Methods in Computer Science. September 2002.

“Optimal Plans for Aggregation.” Presented at the 21st Annual ACM Symposium on Principles of Distributed Computing, July 2002.

“How to Get Across Town by Bus.” Presented at the INFORMS Annual Meeting, November 2001.

“A Brief History of Lognormal and Power Law Distributions.” Presented at the 39th Annual Allerton Conference on Communication, Control, and Computing, October 2001.

“Compressed Bloom Filters.” Presented at the 20th Annual ACM Symposium on Principles of Distributed Computing, August 2001.

“How Useful Is Old Information?” Presented at the INFORMS Applied Probability Conference, July 2001.

“Using Multiple Hash Functions to Improve IP Lookups.” Presented at the INFOCOM, April 2001.

“On the Hardness of Finding Multiple Preset Dictionaries.” Presented at the 2001 Data Compression Conference, March 2001.

“Improved Results for Route Planning in Stochastic Transportation Networks.” Presented at the 12th Annual ACM-SIAM Symposium on Discrete Algorithms, January 2001.

“Estimating Resemblance of MIDI Documents.” Presented at the 3rd Workshop on Algorithm Engineering and Experiments, January 2001.

“Linear Waste of Best Fit Bin Packing on Skewed Distributions.” Presented at the 41st Annual IEEE Symposium on Foundations of Computer Science, November 2000.

“Improved Classification via Connectivity Information.” Presented at the 11th Annual ACM Symposium on Discrete Algorithms, January 2000.

“The Asymptotics of Selecting the Shortest of Two, Improved.” Presented at the 37th Annual Allerton Conference on Communication, Control, and Computing, August 1999.

“Analysis of Timing Based Mutual Exclusion with Random Times.” Presented at the 18th Annual ACM Symposium on Principles of Distributed Computing, May 1999.

“Accessing Multiple Mirror Sites in Parallel: Using Tornado Codes to Speed Up Downloads.” Presented at INFOCOM '99, March 1999.

“Delayed Information and Action in On-Line Algorithms.” Presented at the 39th Annual IEEE Symposium on Foundations of Computer Science, November 1998.

“A Derandomization Using Min-Wise Independent Permutations.” Presented at Random ’98, October 1998.

“On Balls and Bins with Deletions.” Presented at Random ’98, October 1998.

“The Power of Two Choices and Other Examples of Using Differential Equations to Analyze Algorithms.” Invited presentation at the Warwick Randomized Algorithms and Stochastic Simulation Tutorial and Workshop. July 1988.

“Analysis of Low Density Codes and Improved Designs Using Irregular Graphs.” Presented at the 30th Annual ACM Symposium on Theory of Computing, May 1998.

“Average Case Analyses of First Fit and Random Fit Bin Packing.” Presented at the 9th Annual ACM Symposium on Discrete Algorithms, January 1998.

“How Useful Is Old Information?” Presented at the 16th Annual ACM Symposium on Principle of Distributed Computing, August 1997.

“On the Analysis of Randomized Load Balancing Schemes.” Presented at the 9th ACM Symposium on Parallel Algorithms and Architectures, Newport, Rhode Island, June 1997.

“Load Balancing and Density Dependent Jump Markov Processes.” Presented at the 37th Annual Symposium on Foundations of Computer Science, Burlington, Vermont, October 1996.

“Constant Time per Edge is Optimal in Rooted Tree Networks.” Presented at the 8th ACM Symposium on Parallel Algorithms and Architectures, Padua, Italy, June 1996.

“GRID: Pattern-based Compression of Text Images.” Presented at the 1996 Data Compression Conference, Snowbird, Utah, April 1996.

“Parallel Randomized Load Balancing.” Presented at the 27th ACM Symposium on the Theory of Computing, Las Vegas, Nevada, May 1995. Extended version presented at the Bay Area Theory Seminar, July 1995.

“Bounds on the Greedy Routing Algorithm on Array Networks.” Presented at the 6th ACM Symposium on Parallel Algorithms and Architectures, Cape May, New Jersey, June 1994.

Other Presentations

“Algorithms with Predictions.” Georgia Tech. 2023.

“Modern Algorithms Lecture Series.” Lipari Summer School. 2022.

“Algorithms with Predictions.” Carnegie Mellon, UC Santa Barbara, Technical University of Munich. 2022.

“Locality Sensitive Hashing.” Broad Institute (Models, Inference, and Algorithms).

“Algorithms with Predictions.” Open Data Science Conference East 2020, Open Data Science Conference Europe 2020, Foundations of Data Science Institute Kickoff Workshop 2020.

“Learning-Augmented Algorithms: How ML Can Lead to Provably Better Algorithms.” Rice University. 2019.

“Metric Sublinear Algorithms via Linear Sampling.” Sublinear Algorithms Workshop (MIT), Baidu Research. 2018.

“Bloom Filters, Cuckoo Hashing, Cuckoo Filters, and Adaptive Cuckoo Filters.” Northeastern University, Univ. of Illinois at Urbana-Champaign, Georgetown University, University of California Irvine, Baidu Research, Georgia Tech. 2018.

“Hashing Lecture Series.” Swedish Summer School in Computer Science. 2016.

“More Analysis of Double Hashing for Balanced Allocations.” U.C. Irvine. 2016.

“Algorithm Engineering with Hashing,” U.C. Irvine, AMD. 2016.

“Bloom Filters and Such,” EADS Summer School on Hashing: Theory and Applications, July 2014.

“Cuckoo Hashing, Balanced Allocations, and Such,” EADS Summer School on Hashing: Theory and Applications, July 2014.

“Peeling Arguments: Invertible Bloom Lookup Tables and Biff Codes,” ICERM Workshop on Stochastic Graph Models, March 2014.

“Peeling Arguments: Invertible Bloom Lookup Tables and Biff Codes,” University of Copenhagen. 2013.

“Some Hash-Based Data Structures and Algorithms Everyone Should Know.” UC Davis, 2013.

“Peeling Algorithms,” Simons Institute. 2013.

“Verifying Computations in the Cloud (and Elsewhere),” TTI/Vanguard : Ginormous Systems. Microsoft Faculty Summit. 2013.

“The Groupon Effect on Yelp Ratings: a Root Cause Analysis,” Cornell, 2012.

“Peeling Arguments: Invertible Bloom Lookup Tables and Biff Codes,” Google. 2012.

“Detecting Novel Associations in Large Data Sets.” eHarmony (LA Machine Learning). 2012.

“Cuckoo Hashing: New Results and Open Questions.” University of Madison-Wisconsin. 2011.

“Practical Verified Computation with Streaming Interactive Proofs.” University of Madison-Wisconsin. 2011.

“The Power of Choice: Simple Hash-Based Data Structures Everyone Should Know.” Santa Fe Institute. 2011.

“Information Asymmetries in Pay-Per-Bid Auctions: How Swoopo Makes Bank.” Liverpool University, Microsoft UK. 2010.

“Some Open Questions on Cuckoo Hashing.” Stanford, Microsoft Research, Yale, Dartmouth, Princeton, Cambridge University, University College London . 2009-2010.

“An Efficient, Rigorous Approach for Identifying Statistically Significant Frequent Itemsets.” Google. 2009.

“Some Recent Results on Cuckoo Hashing.” AT&T. 2009.

“Some Results on Coding for Flash Memory.” University College London. 2009.

“Network Coding Meets TCP.” University College London. 2009.

“A Survey of Results for Deletion Channels and Related Synchronization Channels.” Cambridge, Georgia Tech, Microsoft Research, UCLA, 2009-2010.

“Bloom Filters, Related Data Structures, and their Applications.” Cambridge (Kuwait Lecture). 2009.

“A History of and New Directions for Power Law Research.” SUNY Buffalo. 2009.

“Why Simple Hash Functions Work.” Cisco. 2008.

“Cuckoo Hashing and CAMs.” Cisco, Google. 2008.

“New Results in Hash-Based Data Structures.” Cisco. 2007.

“The Hiring Problem and Lake Wobegon Strategies.” Microsoft, U.C. Berkeley, Princeton, Google, Yahoo. 2007-2008.

“A Brief History of Lognormal and Power Law Distributions.” Harvard Statistics. 2006.

“New Results and Open Problems for Deletion Channels and Related Synchronization Problems.” CMU, Brown, Harvard. 2006.

“Some New Results on an Old Data Structure, Bloom Filters .” Stanford, Cisco, Microsoft. 2005.

“Improved Lower Bounds for Deletion Channels.” MIT, EPFL (Lausanne). 2005.

“Building a Better Bloom Filter.” Google, Tufts, EPFL (Lausanne), MIT. 2004-05.

“Geometric Generalizations of the Power of Two Choices.” Columbia, Brown, MIT, Georgia Tech. 2004-05.

“New Lower Bounds on the Capacity of the Binary Deletion Channel.” MIT, 2004.

“Digital Fountains, and their Use for Informed Content Delivery across Overlay Networks.” Cornell, 2003.

“Verification Codes for Error and Deletion Channels.” U.C. Berkeley, 2003.

“New Exhaustive, Heuristic, and Interactive Algorithms for 2D Strip Packing.” U.C. Berkeley, 2003.

“Dynamic Models for File Sizes and Double Pareto Distributions.” IBM Research, Microsoft Silicon Valley, Stanford, Tufts, Boston University, Random Graphs and Algorithms Workshop. 2002.

“Verification Codes.” DIMACS workshop. 2001.

“Getting Across Town by Bus, Aggregation, and Reliability.” MIT (LIDS), Mitsubishi Electronic Research Labs. 2001.

“New Ideas for Layered Multicast Congestion Control.” International Computer Science Institute. 2001.

“Compressed Bloom Filters, and Towards Compressing Web Graphs.” U. of Washington, AT&T Labs, U. C. Berkeley, MIT, Brandeis, Carnegie Mellon University, Stanford, DIMACS workshop on compression, U.C. San Diego. 2001.

“Toward Compressing Web Graphs.” Compaq Systems Research Center, Altavista Corporation. 2000.

“Improved Results for Route Planning in Stochastic Transportation Networks.” Compaq Systems Research Center. 2000.

“Choosing the Shortest of Two, Improved, with Applications to IP Routing.” U. Washington, Princeton, Yale, U.C. Berkeley, Boston University, Stanford, M.I.T., Northeastern, Compaq Systems Research Center. 2000.

“Tornado Codes, with Applications to Reliable Multicast.” Harvard, CMU, Duke, AT&T Labs, Univ. of Illinois at Urbana-Champaign, University of Rhode Island, Boston University. 1998-2000.

“How Useful Is Old Information?” MIT, U.C. Berkeley, Stanford, Microsoft. 1997.

“Practical Loss-Resilient Codes.” U.C. Berkeley, Harvard. 1997.

“An Average Case Analysis for First-Fit Bin Packing.” Stanford, DIMACS workshop. 1997.

“The Power of Two Choices.” U.C. Berkeley, Brown, Harvard, Boston University, IBM Almaden, Digital Systems Research Center. 1997.

“The Power of Two Choices and Other Examples of Using Differential Equations to Analyze Algorithms.” U. Mass.