I know what you will do next summer

Balachander Krishnamurthy
AT&T Labs–Research
http://www.research.att.com/~bala/papers

This article is an editorial note submitted to CCR. It has NOT been peer reviewed. The author takes full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

This is a brief journey across the Internet privacy landscape. After trying to convince you about the importance of the problem I will try to present questions of interest and how you might be able to apply your expertise to them.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Privacy

General Terms

Measurement, Security, Economics, Legal Aspects

Keywords

Privacy, Identity, Anonymization, Online Social Networks

1. INTRODUCTION: WHAT IS PRIVACY?

Privacy has become a hot topic recently although work in this area has been ongoing for close to two decades. The networking community has of course largely tended to ignore this topic, although that appears to be changing. I am not going to try and frighten you about how your unique government issued identification number may now be in the wrong hands or how the shady pictures you deleted from your Online Social Network account remain available with additional tags identifying you by your "secret" nickname. Instead, I will try to shed light on a few key issues: What is privacy? Should we care about privacy leakage? How bad is the current situation? What are the technical questions and how could you contribute to solving the problems? Beyond technology, can anything be done? Not too surprisingly my thoughts mirror my work in this area but I also include what I have learned from my interactions while giving talks in universities, industrial research laboratories, conferences, and some, er, less well-known government agencies in various countries.

Let me begin with an anecdote. A few years back, a PC chair of SIGCOMM invited the PC members (and a guest!) to a snooty dinner (*not* paid for by SIGCOMM; I know, I know, shocking). The restaurant was elegant, and one PC member was even turned away for not adhering to their dress code. Some pictures were taken during the dinner and a few days later some of them were posted on the Web. Email was sent to those in the pictures and upon protest by some (somewhat to the surprise of the poster) the pictures were taken down immediately.

The simple incident above depicts several interesting facets of privacy. One could argue that SIGCOMM PC members are somehow "public" (given the hue and cry about decisions made each year, one wonders how long the list would/should remain public). One could also argue that the poster at least notified the featured persons about the fact that their pictures were on the Web and took them down promptly upon hearing the objection. In many cases, pictures (and other information about users) routinely show up on the Web, often without the knowledge of the users. If some of the information is deemed private, how could one go about preventing such collateral damage (i.e., the information was made public by others, and not the users)?

In recent years, we have all become very familiar with individuals uploading a significant amount of information about themselves *voluntarily* on a variety of Online Social Networks (OSNs). Given the rate at which OSNs have grown (a third of all users who have access to the Internet are on some OSN), such sharing of personal information is now a *fait accompli*. Some people may have a libertarian view about privacy and argue that people should be allowed to post any information (factual or otherwise) about themselves and that legislative bodies should not make laws against such actions.

So what is privacy and how might privacy leakage affect *you*? This is a simple question but without simple answers. Like the former US Supreme Court Justice Potter Stewart's memorable quip about obscenity, many of us could probably identify a violation of privacy when we "see it". There is no rigid definition of personal privacy but I mentioned a couple in a recent book (Chapter 8 of Internet Measurement [8]): the European Union's Privacy Directive [1] defines an "identifiable person" as "one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". The World Wide Web Consortium's Platform for Privacy Preferences (P3P [13]) specification allows for the view that most information referring to an individual is "identifiable" in some way. Data that uniquely identifies a person is *identified* data. If the data can be combined with other data to identify a person, then such data is termed *identifiable* data.

A serious subset of private data—Personally Identifiable Information (PII)—has been formally defined [4] as referring to "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc." It is not hard to imagine how one's anonymous action in a different sphere of one's life could be linked with one's identity and exposed to others. For several years it has been possible to look up someone's home value (on `zillow.com` for example) although this might still surprise some. By the time of arrival of Spokeo, Pipl, 123people.com, Intelius, and other such social aggregators, the no-

[1] http://www.cdt.org/privacy/eudirective/EU_Directive.html

tion of expectation of privacy itself is being redefined. Beyond PII, a user's relationships, actions on the Web, travels, consumption habits, etc. are among the issues that raise strong privacy concerns to varying degrees. Some users voluntarily upload all of the above information, including recent credit card purchases (Blippy, Tribesmart).

Beyond issues of definition, individuals vary in their perceived need for privacy: some people "overshare" (by some people's estimation) information and are seemingly unconcerned even when it is brought to their attention. Ignorance is a strong component: most of us do not have any idea of what fraction of our personal identity is available on the Internet, and more importantly, who all have access to it. Not all personal bits of information are considered private by all users.

A dimension typically forgotten is the temporal one: what we might be willing to expose today, we might want to take back at a later date, presumably because we have changed our mind about our earlier decision to share. As we all know, the Internet does not forget; too much data is indexed and available in cached form even if the source is 'deleted'.

2. SHOULD WE CARE?

Should we care about privacy leakage? This is relatively straightforward to answer: you know the answer already to some extent. If you think you have a right to privacy (which by the way is guaranteed in many countries as a fundamental human right), then clearly you should get to decide *what* information about you is being recorded by *whom* and whether you think it is for a legitimate sanctioned purpose. The mantra is "informed consent", which translated from Sanskrit means that you, the owner of your private data, understands the nature of the information and consent to it being received, stored, processed, and analyzed with your knowledge for a specific purpose, and possibly for a specific duration. Unfortunately, laws vary from country to country and there are significant differences between even the European Union and the United States (e.g., in data retention durations).

In reality, informed consent is rarely available as an option. Even if you allow a Web site to track your movements via cookies or JavaScript, and even if the Web site offers to share with you what information they have currently stored about you, it is *not enough*. Why? Because, what you are able to see in the best of circumstances is partial information: the bits about you can be combined with past or future data to draw new inferences. The Wall Street Journal recently reported on a study[2] whereby companies, such as Lotame, indicated how raw information about users (age, gender, location etc.) are encoded in cookies.

Going back to the original 1997 definition of cookie (RFC 2109) as an opaque string (i.e., the content in a Cookie header is "of interest and relevance only to the origin server"), it has been long been suspected that *any* information can be encoded in such strings. "Deleted" cookies could be re-spawned if the trackers are able to establish that the host (or in some cases, the user) associated with the cookie string is very likely to be the same one in the previously deleted cookie string. After all, when a user deletes a cookie they are *only* deleting what is in their browser or hard disk—and *not* the associated data in the tracker's machine(s). Unless the data aggregators can demonstrate that a user's wishes to have their information deleted has been met to the fullest extent (i.e., any and all information about that user has been permanently deleted by the aggregators), users *should* care.

[2]http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html

But do you care? Maybe not, unless you are *explicitly* or *repeatedly* presented with some material harm to your data or identity.

Initially cookies were delivered by the sites you visited directly ("first-party cookies") and soon with the advent of third-party aggregators, third-party cookies (the ones sent by the Web sites that are automatically visited by your browser when you visit a first-party site, often without you being aware of it) became quite prevalent. Blocking of such third-party cookies (often used to track users) became an option in browsers and the aggregators resorted to delivering their third-party cookies disguised as first-party cookies! But cookies are just the simplest of tracking mechanisms. Hidden third-partiy cookies and increasingly sophisticated tracking through complex JavaScript code, multiple parallel strands of tracking etc., is the new normal on the Internet [6]. The tracking industry is a multi-billion dollar one and economic acquisitions of behavioral tracking companies purely for the purpose of older, longitudinal data has become commonplace. The number of large aggregators are few and their visibility into user's movements on the Internet is quite high [6]. If we are to follow the Watergate scandal adage, 'Follow the money', then clearly there is reason to be concerned.

To present a slightly more balanced picture: not all tracking is inherently evil. Many users do like targeted ads; at least they prefer relevant advertisements and appear to be willing to forgo some concerns in return for potentially useful information. A recent Washington Post interview quoted a privacy engineer as saying that "What's good for the consumer is good for the advertiser" [18]. The question is of course who gets to make the decision of what is "good". The current system is set up in such a way that all decisions are made by aggregators and the degree of control especially via a simple user interface to users is quite limited. It is also a fact that even when there are choices to control their privacy, most users are either reluctant to use them or feel that they are cumbersome. Additionally, the knowledge gap, among a large fraction of the users is alarmingly high.

3. HOW BAD IS THE SITUATION?

So how bad is the situation? I will discuss three aspects of privacy: user awareness, data collection at the aggregators end, and users' ability to control leakage.

From an awareness point of view, the situation is pretty bad. A vast majority of users are unaware of what information about them is being tracked, by whom, and for what purpose.

Over the last several years, my friend and longtime collaborator Craig Wills (of Worcester Polytechnic Institute) and I have been examining the issue of privacy from different perspectives. We are hardly alone: several studies (too many to go into in an informal writeup like this) have pointed out multiple vectors of privacy leakage in different facets of Internet usage. Our first study [5] in 2005 began by examining the cat and mouse game between users who are downloading content and advertisers. Since then, we have observed steadily increasing aggregation of Internet tracking data amongst just a handful of companies. This is further compounded by economic aggregation of these tracking entities in the hands of even fewer large companies. The same large companies are well represented in aggregation of data on the popular OSNs. We point out that a large fraction of user's Web activities are monitored by just a handful of third-party 'families' (large aggregator companies) who are getting larger via economic acquisitions over time (see Figure 7 and Table 1 in [6]). Suddenly there is a recognition that seemingly small third parties or ad networks are now part of much larger, well-known entities.

Some events routinely follow the publication of a paper about

privacy leakage: in most cases the publication is ignored both by the technical community and the wider population at large. Occasionally, a brief firestorm of publicity breaks out and some of the 'offending' parties will be asked to respond by some sections of the media (the rare blogs that cover technical papers are often ignored). The responses from the companies that are allegedly behind the leakage will vary from "We do no such thing" to "We have fixed the problem affecting a small number of users in a few rare cases". A few days or weeks later the firestorm dies down and everyone goes back to business as usual: users continue using the site(s), aggregators largely continue what they have been doing. There are counter-examples: recently when Google came out with its attempt at a pre-baked OSN ("Google Buzz"), where they had opted-in by default over 170 million Gmail users into their new attempted OSN, there was a firestorm of criticism and several of the most objectionable features were changed *within 72 hours*. The fact that some information about user's email accounts was made more public than before was of course a problem that could not be fixed: on the Internet, data once leaked is largely leaked forever.

Last year, in SIGCOMM WOSN 2009, we disclosed that personally identifiable information was being leaked *via* multiple popular OSNs [1]. We showed how a OSN user's unique identifier was being leaked via HTTP headers, external applications, and how raw bits of PII (name, age, gender etc.) were also being leaked. The aha! moment in awareness comes when the aggregators seen receiving personally identifiable information from popular OSNs belong to the *same* handful of families mentioned above.

The paper included actual examples of leakage and our prior direct notification about the results was ignored by the OSN companies that we contacted. Initially, the external reaction was somewhat muted. Recently, I was interviewed by the Wall Street Journal [15] and the Los Angeles Times [3]. The same facts were retold to a much wider audience and there was an eruption of attention (thousands of links from most every technical blog and media outlets followed within a day). A prominent OSN announced a fix within hours! The important role played by the WSJ in the overall business community was probably the reason for such widespread publicity; many other papers do not get anywhere this much attention. The WSJ article was followed a few days later by an op-ed response from the founder of Facebook in the Washington Post [20] about various privacy issues. To be honest, the problem of privacy had been building up and the timing of the two articles was coincidental. A few weeks later the issue died down. And in August 2010, the Wall Street Journal published a series of articles under the topic "What They Know"[3] bringing further attention to the topic—this was basically a re-doing of our work described in [6].

In terms of the data acquisition front, the overall appetite of aggregators has *grown*. Not a week goes by without some kind of horror story about leakage of some personal data from some Web site. Bruce Schneier has written about the illusion surrounding privacy[4] pointing out that many at the helm of companies that have access to information about users have a negative view about privacy and concludes that legislation may be the only solution.

In terms of ability to control, there has been some improvement due to a lot of pressure on the part of privacy advocacy organizations, the work of the privacy commissioners, co-operation from advertisers (such as the consumer education project National Advertising Initiative [12]), and the United States Federal Trade Commission's key role in trying to bring various affected parties together. Much remains to be done.

[3]http://wsj.com/wtk

[4]http://www.schneier.com/essay-311.html

4. WHAT ARE THE TECHNICAL ISSUES?

So far, I have tried to convey why the problem of privacy is important. I will now turn my attention to addressing how one could improve the situation. As students, academics, and researchers, we want to know what are the technical questions of interest that have been asked, what are people already working on, and what is hard? Given the broad set of capabilities of the readership here (protocols, systems, security, theory, algorithms, database, measurement, etc.), the hope is that my call to arms would allow you to match your talents to some of the problems in this space. I should stress that this is *not* a survey paper and so there aren't comprehensive answers here to any of these questions. I encourage you to read papers that appear in numerous venues (WOSN, PET, Usenix Security, IMC, HotSec, SOUPS, CFP, etc.) to get a better idea. Maybe one of you will even be inspired to write a good survey paper for CCR and then tweet about it.

How can privacy leakage be detected? A large fraction of the work has been devoted to identifying different manners of privacy leakage. Systems and measurements researchers can easily contribute here. What information is being leaked, to whom, and how—this is the typical set of questions covered in such work. The Personally Identifiable Information (PII) leakage work [1] mentioned in Section 3 was a relatively serendipitous discovery, but one that benefited from our earlier multi-year longitudinal work [6] on the role of data aggregators tracking visitors to popular Websites. Most of the leakage detection does not require sophisticated sniffing or code insertion techniques. However, the knowledge of HTTP combined with the ability to do some straightforward system work (in the form of writing Firefox extensions) and carrying out targeted measurements was all that was needed.

Although there have been a few toolset proposals for detecting most Web- and OSN-related privacy leakage discovery, browser extensions appear to be the most popular technique. Extensions are easy to deploy and much like iPhone apps, one could just wait for users to download. More sophisticated attempts, such as, modified browsers or external Javascript packages have also been made. The potential for good systems work in this area remains quite high, especially if you understand the underpinnings of the Web and are willing to do some patient measurements.

If the default setup on an OSN is not to share any more information than what is absolutely required, it would bound potential leakage. Often the defaults are permissive and there is no easy way to examine the current settings in an understandable fashion. Automatically mapping the user's desired degree of privacy and tightening the defaults would be quite useful. But a tool that does this for the universe of a user's interactions would be a good start. Companies like Blue Kai[5] and Lotame[6] have made an initial foray through their registries that disclose exactly what information they have about users and allow users to delete some or all of the acquired data.

Transitive closure issue: Most of the leakage identification work is *local*, i.e., limited to just the set of Websites explicitly visited. Recently, while researching the issue of PII leakage in *mobile* OSNs [7], we came across the transitive closure problem: a user's actions on one mobile OSN was, in many cases, being *automatically* gatewayed to one or more popular traditional OSN (like Facebook or Twitter). A user would update their location to alert their friends on a mobile OSN who are nearby. However, a setting on the mobile OSN would automatically translate this 'check-in' (at a specific latitude/longitude) into a Facebook or Twitter status update—one that

[5]http://tags.bluekai.com/registry

[6]http://www.lotame.com/preferences.html

is typically visible to a many people who are far away. Thus, the privacy settings on the mobile OSN alone do not control the visibility of a user's action–one has to examine the transitive closure of the impact of the user's action. This is a simple example: imagine having to go through all possible actions on all possible Internet outlets that a user might visit during the course of a day using multiple devices, access methodologies, and privacy settings. I am not aware of a comprehensive mechanism that examines this question in a holistic way: what information about a user is shared on the Internet without the user's knowledge and what actions and by whom led to this situation? The research questions here include automatic generation of destinations to which data is sent, the specific actions and settings that triggered the transmission, enumeration of leakage as a result of the universe of actions, and potential ways of blocking any or all of it.

Where to provide protection? The question of how and where to provide privacy protection is an interesting one. Should privacy protection be offered at the browser, at a gateway (possibly using lower layers of the protocol stack), or at some intermediary between the user and the Internet destination? The degree of per-user control and efficiency of a solution depends on where protection is provided. An organization might benefit if all users could be assured of a certain degree of privacy protection via a common mechanism. But the diversity in the set of destinations and the variance in the degree of tracking and comfort level of users hints at a split-solution: some higher degree of overall protection combined with more nuanced per-user protection. Simple protocol-level proposals have been made (e.g., to eliminate a common avenue of leakage, the HTTP Referer header) but privacy can leak both in headers and payload.

Architectural issues: An architectural question is where should a user's private data be stored? All the popular OSNs have a centralized data store. This has implications not just for performance but privacy as well. One way to circumvent dependence on a single central entity is distributing the private data—thus leading to a decentralized OSN. Two projects in this space are Lockr [17] and Vis-a-Vis [14]. Use of intermediaries such as a proxy and other anonymization techniques have also been proposed; although one has to now trust the intermediaries. Not too surprisingly, there is movement afoot to store user data in the cloud; the attendant privacy issues are beginning to be explored. Subtle issues arise in this context: legal and contractual issues may limit government agency from using cloud storage for official records; physical geographic laws may limit distributed storage. The World Privacy Forum has a good discussion on cloud privacy issues[7].

Identity management: The issue of managing a user's online identity is a broad open question and one that is being addressed via a growing body of work. What is the best way to ensure that a user's identity can be presented to various Web entities that preserves the privacy of the user, ensures that man-in-the-middle attacks don't take place, while preserving the potential of all traditional transactions. Considerable work has occurred in the IETF and elsewhere on distributed identity management. Ranging from the OAUTH open standard to the complementary work on OpenID (and the more recent PseudoID [2] proposal) there have been standards, proposals, protocols, code, and compliant implementations from several key players in the field.

Anonymization: There has been considerable work related to anonymization that may have implications for privacy in the database world. Beginning with Sweeney's seminal work on *k-anonymity* [16] setting up the basic parameters for publishing data, there has been a veritable alphabet soup of followup work (l-diversity, t-closeness etc.). These pieces of work concern themselves with ensuring that data can be safely published and the published data has utility for analysis. The interesting idea of differential privacy has seen applications in numerous fields such as protecting privacy in recommendation systems [11] and has been adapted recently in network data anonymization [10]. It would be interesting to see if it could also be used for providing personal privacy.

Security: On the security front, there are various cryptographic approaches to preserving privacy. There has been recent interesting work on protecting privacy while retrieving information by breaking user's queries into subqueries to reduce risk of reverse engineering of the user's intent [19]. A visit to the recent blackhat convention would have introduced you to various privacy issues ranging from being able to track RFID tags from significant distance, to the deployment status of anonymous darknets, and attacks on device privacy[8]. Tor has been a well studied tool and although its potential for user privacy has always been there, its widespread usage for this purpose still remains low.

Usable privacy: No matter what techniques we come up with for protecting privacy, a key factor is its *usability*. Not too surprisingly, the more sophisticated the technique, the fewer the takers. As it is, even a simple Firefox extension which requires three clicks to download and enable, results at best in only a few thousand takers out of a potential user population of several million. There are extensions and tools that have a large number of faithful users and with each new press story about privacy leakage, a few more adherents show up. For example, the aforementioned attempt to create the pre-baked Online Social Network "Google Buzz" led to a spike in attention to the age-old topic of opt-in vs. opt-out. Companies prefer opt-out: users are lethargic and may not notice (or care if they notice) that they have been opted-in forcibly. Opt-in, on the other hand, *requires* work on the part of the same lethargic user; thus it is harder to have a large number sign up for any new tool. A key item here is that both opt-in and opt-out have to be presented to the user in a *usable* manner. Just as a popular Firefox extension allows users to selectively turn off JavaScript execution but could result in partial rendering of Web pages, the use of a privacy protection tool should not result in inexplicable results. If users are not presented with a highly usable interface, they may not use it.

Who is going to pay for all this? To my academic colleagues and others who are concerned about research grants, multiple proposals have been funded recently by the National Science Foundation for work in the area of privacy. DARPA has had RFIs in this space. There are several EU projects as well.

5. BEYOND TECHNOLOGY, WHAT CAN BE DONE?

There are at least three different angles through which one could approach the problem of reducing privacy leakage: technical, legislative, and economic. Although I largely focused on the technical means in the previous section, it is useful to talk about the other two angles.

The Internet spans way too many countries to expect any single legislative solution to address even one of its myriad problems. Privacy leakage is no exception. However, there is a role that governments can and do play. Privacy commissioners of different nations have their own annual conference exchange information and try to coordinate possible solutions. The privacy commissioner of Canada, for example, has been quite active in addressing privacy leakage issues in OSNs. There is a new European Union law being

[7]http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf [8]http://www.blackhat.com/html/bh-us-10/bh-us-10-home.html

proposed that would require online publishers to receive consent before placing a cookie on a user's machine[9]. The Federal Trade Commission (FTC), in the United States, has organized several technical get-togethers and solicited white papers and comments from technologists, sociologists, and others. The FTC does work closely with several large companies and with advertisers, as well as with the US Congress on possible legislative remedies. Often the threat of legislation helps focus attention on a problem.

There are at least two different efforts currently pending in the U.S. Congress. Representative Ed Markey's Congressional Privacy Caucus has continued to hold several hearings on various privacy related topics[9]. Most recently, they have been examining potential legislation related to consumer tracking and exploring the potential of a "Do not Track" registry ostensibly similar to the earlier successful "Do Not Call" telecommunication registry. The analogy is not quite right—the one-to-one mapping of a telephone number to the customer does not have an exact parallel in a single user who may use different IP addresses to access the same site possibly via different devices. There are also plans by the U.S. Senate's Committee on Commerce to introduce legislation this year to give people more control over how their personal information is collected and distributed online. The role of privacy professionals has also grown over the years—a recent report by the 6,000 member International Association of Privacy Professionals (https://www.privacyassociation.org/) indicates the market for privacy-advice to be around $1 Billion.

The role of physical (i.e., human) security has not been discussed so far. Thanks to recent world events, an extraordinary amount of attention has been focused on improving security globally, both on- and off-line. In any battle between privacy and physical security, there is a high probability that privacy might be the loser: governments and other institutions will inevitably state that protecting the physical security of their population is paramount and any consequent loss of privacy should be an acceptable cost. Privacy advocates have generally been on the losing end in battles against omni-present cameras, tracking for security purposes etc. One hopes however, that it would be possible to come to some middle ground where users do not have to give up their privacy for security, especially security that in many cases may turn out to be illusory.

However, when it comes to private corporations, a new battle is joined: privacy vs. economics. Here, privacy is in a much stronger position. Users could express their unhappines with a company that was violating their privacy. Corporations do feel the heat of mass desertions or largescale threats of boycotts and are willing to amend their procedures, even in the absence of legal requirements. Increasingly, these economic questions have become important: what is the tradeoff for sharing personal information? Is there a tangible and quantifiable benefit to a user in return for sharing information with a specific site? Answers to these questions might turn out to be mutually beneficial: users will know who has what information and what they gain in return for it. It would thus be interesting to carry out an economic analysis and itemize the trade-offs of cost to a company vs. gains in tracking users (privacy advocates tend to use the phrase "value exchange"). Some start-ups have started offering bargains in return for users' data[10]. There is also work in trying to figure out the connection between behavioral economics and privacy and in trying to work out actual cost of breaches of privacy (see the publications of Alessandro Acquisti, for example).

[9]http://register.consilium.europa.eu/pdf/en/09/st03/st03674.en09.pdf
[10]http://www.nytimes.com/2010/05/31/business/media/31privacy.html

6. QUO VADIS?

My view is that any plausible long term solution for addressing the privacy leakage problem would necessarily involve raising awareness, offering several alternative technical solutions, involving various legislative bodies, and exploring smart use of economic analysis to bring aggregators into a modus vivendi. I hope that I have at least piqued your interest in looking into some of the privacy-related issues.

In my talks over the years on various aspects of privacy, virtually every one of the audience members (the ones that were awake) fell into one of three categories: shocked ("Wow, I didn't know this was going on!"), libertarian ("such tracking is needed for my Internet activities and I don't care *who* knows anything about me"), and resigned ("C'est la vie, we have no control over our data or our identity"). I hope some reader will write back and indicate their membership in a *new* category.

Acknowledgments

My thanks to numerous colleagues whose research have been instrumental in shaping the direction of my thinking in this area. Craig is my long-time (long-suffering) co-author of several pieces of research discussed here. Thanks to Sharon Goldberg for her suggestions (nay, demands) and generous colleagues—Steven Bellovin, Ramon Caceres, Jon Crowcroft, Krishna Gummadi, Greg Minshall, Flip Korn, Jeff Mogul, Dina Papagiannaki, Sherry Ramsey, Jennifer Rexford, and Craig Wills—for their incredibly prompt comments. Errors, however, are all still mine.

7. REFERENCES

[1] Balachander Krishnamurthy and Craig Wills. On the Leakage of Personally Identifiable Information via Online Social Networks. *ACM SIGCOMM Computer Communication Review*, Jan 2010. (SIGCOMM 2009: Best Workshop Papers) http://ccr.sigcomm.org/online/?q=node/577.

[2] A. Dey and S. Weis. PseudoID: Enhancing Privacy in Federated Login. In *Hot Topics in Privacy Enhancing Technologies*, December 2010. `http://research.google.com/pubs/archive/36553.pdf`.

[3] M. Hiltzik. Is your privacy secure online? `http://articles.latimes.com/2010/jun/06/business/la-fi-hiltzik-20100606/2`, June 6, 2010.

[4] C. Johnson III. Safeguarding against and responding to the breach of personally identifiable information, May 22 2007. Office of Management and Budget Memorandum. `http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf`.

[5] B. Krishnamurthy and C. E. Wills. Cat and Mouse: Content Delivery Tradeoffs in Web Access. In *WWW*, May 2006. `http://www.research.att.com/~bala/papers/cam.pdf`.

[6] B. Krishnamurthy and C. E. Wills. Privacy Diffusion on the Web: A Longitudinal Perspective. In *WWW*, April 2009. `http://www.research.att.com/~bala/papers/www09.pdf`.

[7] B. Krishnamurthy and C. E. Wills. Privacy Leakage in Mobile Online Social Networks. In *Proceedings of the Workshop on Online Social Networks*, June 2010. `http://www.research.att.com/~bala/papers/pmob.pdf`.

[8] Mark Crovella and Balachander Krishnamurthy. *Internet Measurement: Infrastructure, Traffic, and Applications*. John Wiley & Sons. 550pp, 2006.

[9] Congressman Ed Markey, Privacy. http://markey.house.gov/index.php?option=com_issues&task=view_issue&issue=13&Itemid=152.

[10] F. McSherry and R. Mahajan. Differentially-private network trace analysis. In *Proceedings of the ACM SIGCOMM Conference*, September 2010.

[11] F. McSherry and I. Mironov. Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders. In *Proceedings of the ACM KDD Conference*, June 2009. http://research.microsoft.com/pubs/80511/NetflixPrivacy.pdf.

[12] Network Advertising Initiative: Opt out of NAI member ad networks. http://networkadvertising.org/managing/opt_out.asp.

[13] The Platform for Privacy Preferences 1.1 (P3P1.1) specification. http://www.w3.org/TR/P3P11, July 2005.

[14] A. Shakimov, A. Varshavsky, L. Cox, and R. Caceres. Privacy, Cost, and Availability Tradeoffs in Decentralized OSNs. In *Proceedings of the Workshop on Online Social Networks*, August 2009. http://www.kiskeya.net/ramon/work/pubs/wosn09.pdf.

[15] E. Steel and J. E. Vascellaro. Facebook, MySpace Confront Privacy Loophole. http://online.wsj.com/article/SB10001424052748704513104575256701215465596.html, May 21, 2010.

[16] L. Sweeney. k-anonymity: A Model for Protecting Privacy. *International Journal of Uncertain. Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[17] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: Better Privacy for Social Networks. In *Proceedings of Co-Next*, December 2009. http://lockr.org/papers/lockr-conext2009.pdf.

[18] J. E. Vascellaro. Google Agonizes on Privacy as Ad World Vaults Ahead. http://online.wsj.com/article/SB10001424052748703309704575413553851854026.html, Aug 10, 2010.

[19] S. Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, 2010.

[20] M. Zuckerberg. From Facebook, answering privacy concerns with new settings. http://www.washingtonpost.com/wp-dyn/content/article/2010/05/23/AR2010052303828.html, May 24, 2010.