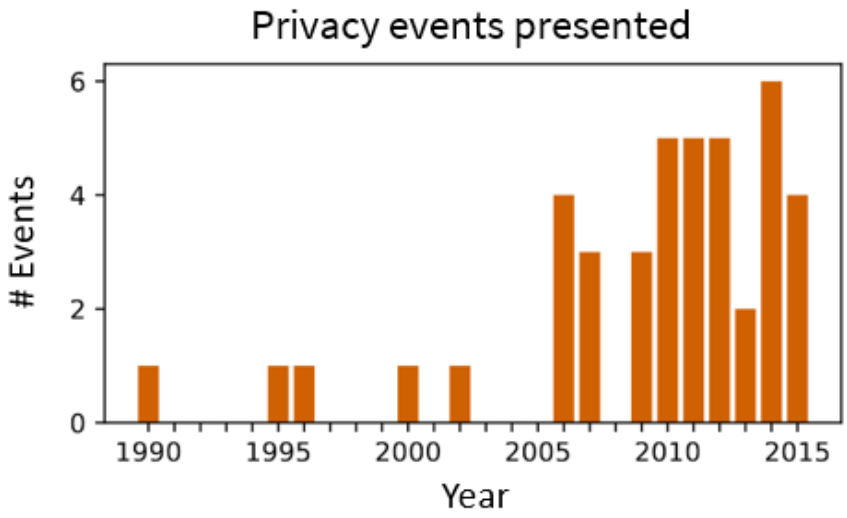
Garfinkel S and Theofanos M. Non-Breach Privacy Events. *Technology Science*. 2018100903. October 09, 2018. http://techscience.org/a/2018100903

Non-Breach Privacy Events

Simson L Garfinkel and Mary Theofanos

Highlights

- Presents a curated list of 44 historically noteworthy incidents in which individuals suffered privacy harms that were not the result of data breaches (theft of personal information).
- Shows application of Solove's Taxonomy of Privacy to recent privacy incidents.



Non-breach privacy events between 1990 and 2015

Abstract

While data breaches frequently create privacy concerns, other types of non-security-related events may also raise privacy concerns. The present study collected and characterized a corpus of non-security-related privacy events that we term "non-breach privacy events." In this article, we consider *non-breach privacy events,* which we define as incidents in which the action or inaction by an individual or organization resulted in a perceived privacy violation,

but where the action did not involve the theft of data as the result of a computer intrusion. Using a systematic search methodology, we identified 44 non-breach privacy events enabled by technology. We then organized these events according to the data flows using Solove's Taxonomy of Privacy and several characteristics that examined the range of these non-breach events.

Results summary: A curated dataset of 44 events that resulted in privacy harms. This dataset is a valuable tool for other researchers wanting to explore handling user data while respecting privacy. Also provided is a qualitative analysis of the nature of the privacy incidents, revealing several trends and lessons learned, the most significant being that just a few people operating within a large organization can create large-scale privacy events.

Introduction

Privacy is about more than data security. *Non-breach privacy events* are incidents in which the action or inaction by an individual or organization resulted in a perceived privacy violation, but where the action did not involve the theft of data as the result of a computer intrusion. Such privacy violations are typically the result of an organization's polices or procedures, not of outside intruders. The desire to prevent these kinds of events was one of the drivers for the European Union's recently implemented General Regulation on Data Protection (GDPR).

This article collects and characterizes non-breach privacy events so that they can remain as part of a common lexicon for discussing privacy related incidents as new people join the privacy field. We also identify trends and commonalities between the cases. To do that, we use a systematic search methodology to catalog 44 non-breach privacy events from 1990 through 2015. We classify the events according Solove's Taxonomy of Privacy, and rate each with respect to our own taxonomy of Scale, Purview, Awareness and Goal of Identifiability.

This article focuses on privacy events enabled by technology. When academics, policy makers, journalists and the general public talk about privacy events, there is a natural tendency to base those discussions on case studies. The cases that make up the modern digital privacy canon now span a quarter-century. Some of these incidents are widely known by those working in the field and are routinely discussed without proper attribution—a practice that makes it difficult for students and early-career researchers to come up to speed on complex privacy issues. Other cases are historical and have largely been forgotten, although they have teachings and precedents that may still be useful today.

We exclude data breaches from this corpus because our primary purpose is to help draw lessons for privacy researchers and professionals regarding the permissible collection and use of data, rather than the protection of computer systems using various information assurance approaches (which have been widely chronicled elsewhere). While a data breach may have an impact on privacy, the lessons drawn from those cases are typically lessons about information security, vulnerability management, and the impact of poor authentication practices. In contrast, the majority of the cases presented here involve

organizations collecting or handling information that they could obtain without an intrusion or computer "hacking."

Background

Other researchers are also trying to separately characterize privacy events, with the goal of identifying categories, creating taxonomies, and helping to advance an understanding of this area.

The Patient Privacy Rights Foundation prepared a collection of 74 *Medical Privacy Stories* [1] in which the privacy of an individual or group of people was violated, usually as the result of inappropriate data release by a healthcare institution. The collection divides the stories into nine categories:

- "Individuals Exposed," which involve releasing of medical information. Some of these resulted in specific harms to an individual, such as losing a job, while others are solely invasions of privacy.
- "Unauthorized Access," in which an individual accessed information to which they were not entitled.
- "Poor Security," in which poor security practices resulted in information being disclosed. Several of these incidents involved email being sent to the wrong recipients, or a single message being sent simultaneously to many individuals because all of their email addresses were listed in the same To: field.
- "Poor Disposal," in which computers or paper records were not properly destroyed.
- "Medical Information Used for Marketing," in which consumers received targeted advertisements based on their medical status.
- "Government Use of Records," in which medical records were inappropriately shared between different government agencies.
- "Researchers," in which medical records or information were used to recruit subjects for research studies.
- "Law Enforcement," in which law enforcement agents inappropriately used medical information in law enforcement or internal personnel matters.
- "Lawsuits," which are cases of medical privacy that were resolved by lawsuits.

The Department of Health & Human Services maintains a list of Resolution Agreements that detail violations of the HIPAA Privacy Rule [2]. Many involve the misuse or inappropriate distribution of data, rather than security incidents and resulting data breaches.

Researchers at North Carolina State University, UNC Charlotte and Clemson University maintain a database of privacy incidents to answer questions like "what is the most common cause of privacy incidents" and "how do privacy incidents vary by country" [3]. As of June 2018 the database had 408 incidents.

Researchers at RAND created "a unique dataset of over 12,000 cyber incidents recorded over the years 2004 and 2015"[4] assembled from mandatory disclosures other sources. Each incident is categorized as a data breach, a security incident, a privacy violation, or a phishing/skimming attack. According to RAND's analysis, roughly 60% of all incidents are caused by malicious actions, "as opposed to accidental or unintentional activities."

Khan has reviewed a year's worth of FTC "privacy enforcement actions" [5].

Unlike the prior work, this paper focuses on privacy events not caused by information security failings or malicious actors misappropriating confidential data from trusted custodians. Instead, this paper focuses on other ways that privacy can be violated. This is important for educators, policymakers and researchers to consider, because no amount of effort spent on improving data security can prevent these kinds of incidents.

This paper also presents a manageable, curated collection of 44 privacy events from the past three decades. The collection contains many events that received media attention at the time but that have not been included in other collections because they lack the spectacle frequently associated with high-profile data breaches. As such, this list is useful to educators, whose students were likely not following developments in privacy when these events took place. The collection can also be used to drive future research to determine the stage (collection, storage, or access) at which most privacy-related incidents are happening.

Finally, this paper introduces a system for classifying non-breach privacy events based on Solove's privacy taxonomy, the number of affected individuals, the number of individuals within an organization who knew about the incident, and whether or not the goal of the incident was identifying individuals.

Methods

Search Methodology

We found events using the following approaches:

- We performed searches on the Federal Trade Commission (FTC) website for privacy enforcement actions that that contained the keyword "privacy" and then manually reviewed each to remove the enforcement actions resulting from data breaches as we define the term [6], [7]. Because the FTC does not require the existence of a malicious actor in order to categorize an incident as a data breach, some incidents that the FTC classified as a data breach may appear in this list.

- We searched the Federal Communications Commission (FCC) website for press releases (keyword: "For immediate release") that featured the keyword "privacy" [8] and reviewed the results from 2000 through the present day. FCC actions that specifically mentioned "data breaches" were not included, but FCC actions that

featured data being exposed but not necessarily being exploited were included. We also reviewed FCC enforcement actions [9].

- We reviewed journalistic articles that featured lists of famous privacy mishaps by searching for the keyword *privacy* along with the search terms *flaps, snafus,* and *credit bureaus.*

- We asked colleagues to review earlier drafts of this article and provide us with events that we had missed.

We sought to include cases between 1990 through 2015 that were public, well publicized, well known, and legally settled (that is, no longer the subject of an appeal) .

Exclusion Criteria

Because this article is solely concerned with privacy events that result from the improper, authorized, and intentional use of data, we excluded the following kinds of events:

- Where data were stolen by an outsider due to a computer security configuration error or a vulnerability that was exploited. Such "breach events" have been widely reported elsewhere.

- Where data were stolen by an insider due to the insider's dishonesty or systems that allowed the insider to exceed his or her authorized access.

- Where data were released because of the failure on the part of a data custodian to properly destroy data on equipment prior to disposal [10].

- When attackers engaged in pretexting, identity theft, or identity fraud. (*Pretexting* is a form of social engineering, in which the perpetrator lies or provides false or misleading information to an information custodian in an attempt to obtain confidential information about a targeted individual. *Identity theft* is the theft of personal information that could be used to obtain credit or steal something of value; *identity fraud* is the use of the personal information for a fraudulent purpose.)

- When an individual was harmed because of a mismatch in a database (for example, a person being prohibited from boarding a flight because of a mismatch on a "no-fly" list).

- When employers legally accessed email, phone conversations, or the work space of their employees.

- Incidents of improper government surveillance, such as the incidents described by the Senate Select Committee to Study Governmental Operations with Respect to Intelligence Activities, 1975-1976 (Church Committee) [11].

- Data released by inadvertent inclusion in publicly accessible directories. Although such incidents are unfortunate, they typically are the result of poor usability, incorrect system configuration, or poor training, and do not reflect the actor's common business practices.

We also attempted to limit our collection to cases that had cultural or societal significance, as indicated by the number of people impacted, the involvement of government, or media attention.

Finally, we excluded events in which there was apparent wrongdoing but no finding from a government agency, or in which there was no statement of explanation or apology issued by the organization involved in the event.

Analysis Criteria

Solove's taxonomy views privacy as a series of information flows. Informational privacy involves information collected from an individual by surveillance of the individual or collected by data holders through interrogation of an individual. In Solove's taxonomy, the difference between surveillance and interrogation rests with the manner of data collection: "*Surveillance* is the watching, listening to, or recording of an individual's activities. *Interrogation* consists of various forms of questioning or probing for information." (p. 490)

Despite the outsized role that consent plays in today's privacy world, with many organizations requiring all manner of consent from consumers before the consumers can use their service, Solove's taxonomy makes little mention of consent. Solove notes that consent frequently determines the context of an activity and, as a result, whether or not a privacy violation has occurred: "Thus, if a couple invites another to watch them have sex, this observation would not constitute a privacy violation. Without consent, however, it most often would." (p. 484) But the word "consent" appears just 24 times on 12 pages of the 84-page article, mostly to emphasize that a particular privacy violation happened, in part, because an action was taken without an individual's informed consent.

Once a data holder has information about a data subject, the data holder can violate an individual's privacy by employing a variety of privacy-invading information processing techniques. Finally, the data holders may disseminate the personal information in several privacy-invading techniques. Separately, Solove considers privacy invasions that an individual may suffer. Thus, using the taxonomy, it is possible to decompose a single privacy-violating event into multiple kinds of privacy harms.

Solove's taxonomy does not directly address the scale, network, growth, and movement of data within the ecosystem of data holders that has grown drastically since the early 2000s when the taxonomy was formulated. Nevertheless, privacy harms in today's data economy can be readily categorized using the taxonomy. This strongly implies that while the modern

data economy has created more opportunities for privacy harms, it is not creating fundamentally new kinds of privacy harms.

Solove's taxonomy is summarized by its iconic diagram, which we reprint below as Figure 1, and display as a list in Figure 2:

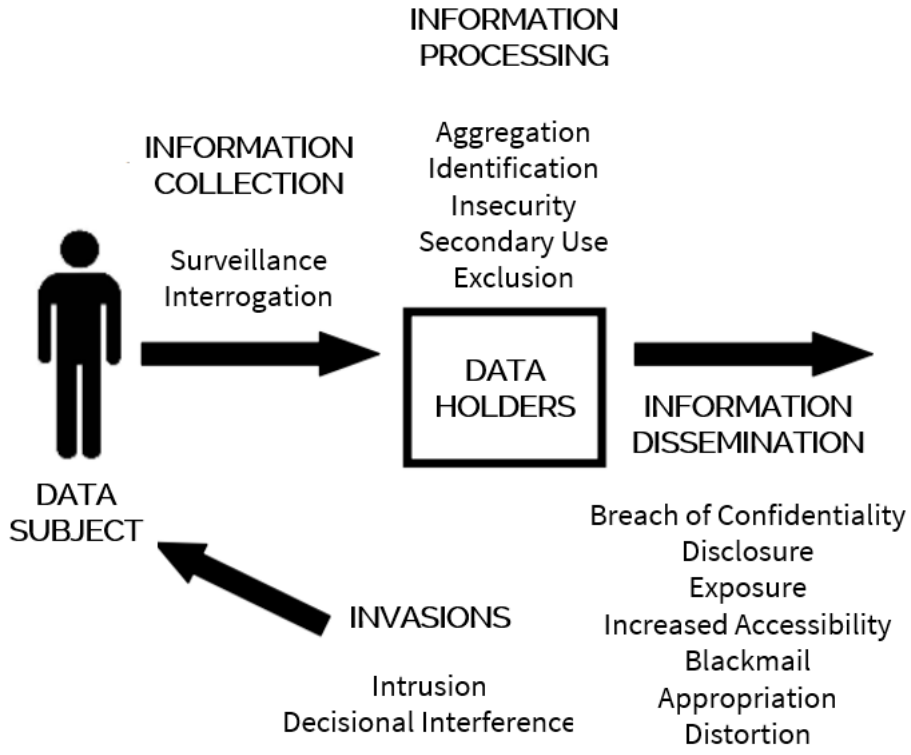


Figure 1: Solove's privacy taxonomy diagram.

Solove's Taxonomy of Privacy

A. Information Collection
1. **Surveillance** (watching, listening to, or recording)
2. **Interrogation** (questioning or probing for information)

B. Information Processing
1. **Aggregation** (gathering together information about a person)
2. **Identification** (connecting information to individuals)
3. **Insecurity** (regarding the way information is handled or protected)
4. **Secondary Use** (the use of data for purposes unrelated to the purpose for which the data was initially collected without the data subject's consent)
5. **Exclusion** (the failure of data holders to provide notice to individuals about the records or the ability to correct those records)

C. Information Dissemination

1. **Breach of Confidentiality** (the harm … is not simply that information has been disclosed, but that the victim has been betrayed.)
2. **Disclosure** (when certain true information about a person is revealed to others)
3. **Exposure** (exposing to others certain physical and emotional attributes about a person… [that] we have been socialized into concealing)
4. **Increased Accessibility (**information that is already available to the public is made easier to access … enhanc[ing] the risk of the harms of disclosure.)
5. **Blackmail** (coercing an individual by threatening to expose her personal secrets if she does not accede to the demands of the blackmailer)
6. **Appropriation** (the use of one's identity or personality for the purposes or goals of another)
7. **Distortion** (manipulation of the way a person is perceived and judged by others)

D. Invasion

1. **Intrusion** (invasions or incursions into one's life)
2. **Decisional Interference** (interference with people's decisions regarding certain matters of their lives… [such as] relating to sex and sexuality, [and] concerning the upbringing of one's children)

Figure 2. Solove's Taxonomy of Privacy [12]; text in parentheses is from Solove's descriptions.

In creating this compilation, we noted that many areas of concern to privacy researchers in recent years do not fit neatly into only one of these categories. For example, early criticisms of Google Street View service focused on the perceived privacy harms of **surveillance** (the recording of street-level photos), **aggregation** (the assemblage of photos from all over the world), **identification** (the matching of photos to places), and **disclosure** (the revealing of true facts—the photographs). Although Street View could not function without all of these aspects, we classify it under Information Collection/Surveillance because of the FCC's action against Google pertaining to Street View's collection of wireless network traffic.

In addition, we characterize each case according to the following characteristics, using terms that we adopted for this purpose:

- **Scale —** The number of people impacted, to the nearest order of magnitude. When discussing a privacy event, the number of people affected is a useful measure for putting the event into perspective. Scale is *not* a measure of the impact of the event on those people, as events that impact a small number of people typically have a larger impact on those people than events that impact thousands or millions.

- **Purview —** We reviewed official statements from regulatory agencies, letters from corporations, news reports, and other material that we reference to determine the number of individuals that had direct knowledge of the actions leading up to the

privacy event. We use the word *purview* to indicate knowledge or experience; we are silent on the issue of legal responsibility.

- **Awareness —** Because of our selection methodology, all of the events in this article involved some aspect of intentional data use. Beyond the intent to use the data, sometimes the actors were aware that their actions might create a privacy event, while other times the privacy event was an unexpected outcome. We use *yes* to indicate that the privacy incident was the result of deliberate, intentional decisions to engage in a particular practice, while *no* indicates that those with purview were not aware that their actions would result in a privacy event.

- **Goal of Identifiability —** If the technology, application, or focus of the privacy event was to single out individuals or equipment associated with individuals including identifying a person with a group or characteristic. We use *yes* to indicate that the privacy event was focused on singling out individuals, while *no* indicates that the event was not focused on singling out individuals, even though the event may ultimately have had that result.

We find these characterizations useful for understanding why organizations engage in practices that impact the privacy experiences of customers, individuals, and the public at large.

Results

In this section, we briefly describe each of the 44 non-breach privacy incidents in our corpus. For each, we provide a brief name and description and the year that it took place. We provide a category and sub-category for each, using Solove's taxonomy. For each we provide a scale, which is the base-10 logarithm of the number of people who were affected. To the best of our ability based on published information, we present the purview of the event, as well as whether the event had the goal of identifiability. All of this is presented in Table 1.

Paragraph	Incident	Year	Category	Sub-Category	Scale Power	Purview	Goal of Identifiability
1.1	Street View Wi-Fi	2007	Collection	Surveillance	8	Few	No
1.2	Lower Merion	2010	Collection	Surveillance	3	Few	Yes
1.3	KTVX(DT)	2012	Collection	Surveillance	0	Organization	Yes
1.4	Brightest Flashlight	2013	Collection	Surveillance	7	Organization	Yes
1.5	Yelp, TinyCo COPPA	2014	Collection	Interrogation	6	Few	Yes
1.6	Harvard Photography	2014	Collection	Surveillance	3	Several	Yes
1.7	Add This	2014	Collection	Interrogation	8	Organization	Yes
1.8	Perma-Cookie	2014	Collection	Surveillance	8	Organization	Yes
1.9	Nomi	2015	Collection	Surveillance	7	Several	Yes
1.10	Pearson Twitter	2015	Collection	Surveillance	7	Organization	Yes
1.11	Spying/Stalking	2015	Collection	Surveillance	6	Organization	Yes

2.1	Facebook News Feed	2006	Processing	Aggregation	7	Organization	Yes
2.2	Verizon Marketing	2006	Processing	Secondary Use	6	Organization	Yes
2.3	Facebook Beacon	2007	Processing	Secondary Use	7	Organization	Yes
2.4	MIT Gaydar	2009	Processing	Aggregation	3	Few	Yes
2.5	Target Pregnancy	2010	Processing	Aggregation	6	Few	Yes
2.6	Apple iPhone Geolocation	2011	Processing	Aggregation	7	None	No
2.7	Uber Overnight Data Analysis	2012	Processing	Secondary Use	5	Small Group	Yes
2.8	PaymentsMD	2012	Processing	Secondary Use	3	Organization	Yes
2.9	Facebook Year in Review	2014	Processing	Secondary Use	8	Organization	Yes
3.1	Lotus Marketplace	1990	Dissemination	Increased Accessibility	8	Organization	Yes
3.2	Massachusetts GIC	1996	Dissemination	Breach of Confidentiality	4	Organization	No
3.3	Lilly Prozac	2000	Dissemination	Breach of Confidentiality	2	Few (1)	No
3.4	jetBlue	2002	Dissemination	Breach of Confidentiality	6	Organization	Yes
3.5	AOL Search Logs	2006	Dissemination	Disclosure	5	Organization	No
3.6	Netflix Prize	2006	Dissemination	Disclosure	5	Organization	No
3.7	Google Street View	2007	Dissemination	Increased Accessibility	8	Organization	Yes
3.8	Jerk.com	2009	Dissemination	Blackmail/Appropriation	7	Few (1)	Yes
3.9	Facebook Like	2009	Dissemination	Disclosure	8	Small Group	Yes
3.10	CVS Dumpster	2010	Dissemination	Breach of Confidentiality	7	Small Group	No
3.11	Google Buzz	2010	Dissemination	Disclosure	8	Small Group	Yes
3.12	Snapchat	2011	Dissemination	Breach of Confidentiality	6	Small Group	Yes
3.13	Uber God View	2011	Dissemination	Breach of Confidentiality	5	Small Group	Yes
3.14	Location in Messenger	2012	Dissemination	Disclosure	8	Small Group	Yes
3.15	Washington Health	2013	Dissemination	Breach of Confidentiality	1	Organization	No
3.16	Revenge Porn	2015	Dissemination	Appropriation	3	Few	Yes
3.17	Healthcare.gov tracking	2014	Dissemination	Breach of Confidentiality	6	Small Group	Yes
4.1	Spam Email	1995	Invasion	Invasion	9	Organization	No
4.2	Facebook Vote	2010	Invasion	Decisional Interference	7	Organization	Yes
4.3	Dialing Services	2011	Invasion	Invasion	6	Organization	No
4.4	Sprint "Do not Call"	2011	Invasion	Invasion	6	None	No
4.5	Emotional Contagion	2012	Invasion	Decisional Interference	5	Small Group	No

Table 1. The compilation of incidents. "Paragraph" refers to the paragraph in this article where the incident is discussed. "Incident" is our title for the incident. "Year" is the year the incident took place. "Category" and "Sub-Category" refer to the Solove category we used to classify the incident. Scale refers to the order-of-magnitude of number of people impacted by the incident. "Purview" refers to the number of people in the organization who were aware of the incident before it became publicly known. "Awareness" indicates whether the organization responsible for the privacy incident was aware of the privacy impact. "Goal of identifiability" indicates whether the goal of the incident was to identify or single out individuals.

1. A — Information Collection

This section presents information processing incidents involving the information collection activities of data holders. Surveillance events typically involve passive collection, while interrogation events involve interaction between the data subject and the data holder.

1.1 Google Street View Wi-Fi Capture (2007) *A1 Surveillance*

Scale: 10^8; Purview: few [13],[14]; Awareness: yes; Goal of Identifiability: no

The Google Street View program involves driving cars on public roads, collecting photographs as the cars drive, and geolocating those photographs on Google's online map products. As part of its Street View program, Google captured the location and Wi-Fi MAC address of every wireless router that it could identify so that Google could deploy a Wi-Fi–based geolocation service, similar to the service pioneered by Skyhook Wireless in 2003. In 2010 Google conducted a technical review of Street View and determined that, in addition to photographs and Wi-Fi geolocation data, Google's cars also recorded and aggregated unencrypted Wi-Fi frames. Google commissioned an outside consulting firm to audit its practices and self-reported to multiple national regulatory agencies, which then conducted their own reviews. In the United States, the FCC concluded that the data Google captured included "names, addresses, telephone numbers, URLs, passwords, e-mail, text messages, medical records, video and audio files, and other information from Internet users in the United States" [15]. As a result of its investigation, the FCC assessed a $25,000 Notice of Apparent Liability (similar to a fine) against Google "for willfully and repeatedly violating an Enforcement Bureau directive to respond to a letter of inquiry" [16]. Separately, Google agreed to pay $7 million to 38 states and the District of Columbia to settle claims arising from the incident [17].

1.2 Lower Merion School District "spycam" (2010) *A1 Surveillance*

Scale: 10^3; Purview: few; Awareness: yes; Goal of Identifiability: yes

System administrators at the Lower Merion, Pennsylvania school district installed software on laptop computers provided by the school district to high school students that secretly snapped photographs every 15 minutes and transmitted those photographs to servers operated by the school system [18]. After a student was disciplined at school for conduct in his bedroom, two parents filed suit against the school district for invading the students' privacy rights. During the course of the suit, it was revealed that more than 66,000 images of students had been secretly snapped and recorded and that two school staffers knew that the images were being recorded. Several students later alleged that the photos included images in which they were nude or partially dressed, and filed suit against the school system [19], [20], [21]. In October 2010 the school district settled the primary lawsuit for $610,000.

1.3 Newport Television KTVX(DT) Telephone Disclosure (2012) A1 Surveillance; C2 Disclosure

Scale: 1; Purview: organization; Awareness: yes; Identifiability: yes

In August 2012, Newport Television LLC's KTVX(DT) in Salt Lake City recorded and broadcast "a consumer's telephone conversation as part of a news segment without first telling the person that the call was being recorded and would be broadcast" [22]. Newport Television LLC agreed to pay a $35,000 civil penalty for the violation of the FCC's Telephone Broadcast Rule in November 2014.

1.4 Brightest Flashlight (2013) A1 Surveillance

Scale: 10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

A popular Android operating system application called "Brightest Flashlight Free" was downloaded over 50 million times [23] by users and used to turn an Android phone into a flashlight. Unknown to users, the app also collected precise location and Device ID from the user's phone and transmitted this data to third parties for the purpose of improving advertising messages. The FTC took action against the makers of the app, Goldenshores Technologies, LLC, for not disclosing the collection of personal information to users [24]. The company was required to improve their notification of users, to provide users controls regarding the collection, use, and sharing of geolocation information, and to delete the data collected from users prior to the settlement.

1.5 BabyBus (2014) A1 Surveillance

Scale: 10^6; Purview: unknown; Awareness: yes; Goal of Identifiability: yes

BabyBus created popular mobile applications designed to teach letters, numbers, and shapes to young children. In 2014 the FTC sent a letter to BabyBus, a Chinese developer of apps for children, warning that the company might be in violation of the Children's Online Privacy Protection Act (COPPA) [25]. "Your apps, offered to users in nine languages, have been downloaded millions of times," the FTC wrote in its letter to BabyBus. "Several of your apps appear to collect precise geolocation information that is transmitted to third parties, including advertising networks and/or analytics companies. Under COPPA and its implementing Rule, 16 C.F.R. § 312 *et seq...,* developers of apps that are directed to children under 13—or that knowingly collect personal information from children under 13—are required to post accurate privacy policies, provide notice, and obtain verifiable parental consent *before* collecting, using, or disclosing any 'personal information' collected from children."

According to BabyBus, the geolocation information was collected by an "Android third-party statistics software plug-in" [26]. Google suspended the BabyBus apps from the PlayStore a week after the FTC's letter was publicized. The apps were later re-admitted to the app marketplace.

1.6 Yelp, TinyCo COPPA (2014) A2 Interrogation

Scale: 10^6; Purview: few; Awareness: yes; Goal of Identifiability: yes

The online review site Yelp, Inc., and its mobile application developer TinyCo, Inc., settled an FTC action involving the collection of children's information on mobile applications in violation of COPPA. The Yelp mobile application requested that users enter their date of birth, name, email address, and other personal information. In thousands of cases, the FTC alleged, children told Yelp's app that they were under 13 but the app continued to collect their personal information. The FTC alleged that this was a violation of COPPA, as Yelp did not have written permission from parents to collect that information. Under the terms of the settlements, Yelp agreed to pay a $450,000 civil penalty, while TinyCo agreed to pay a $300,000 civil penalty [27].

1.7 Harvard University Classroom Covert Photography (2014) *A1 Surveillance*

Scale:10^3; Purview: few; Awareness: yes; Goal of Identifiability: yes

As part of an experiment measuring course attendance and completion rates, digital cameras placed in classrooms at Harvard University photographed students in order to electronically determine classroom attendance using facial recognition. Neither the professors nor the students in the courses were told that video monitoring would be taking place. Harvard's Institutional Review Board (IRB), the body federally mandated to regulate human subjects research at the university, gave approval to the study on the grounds that the work "did not constitute human subjects research," and thus did not require consent of those being monitored [28]. Following the disclosure of the surveillance, Harvard's Vice Provost for Advances in Learning said that all of the collected images would be destroyed.

1.8 AddThis canvas fingerprinting (2014). *A2 Interrogation*

Scale: 10^8; Purview: organization; Awareness: yes; Goal of Identifiability: yes

AddThis developed free website tools including a "sharing button" and "follow buttons," making it easy for website operators to have buttons that allow users to post information from a website on social media such as Facebook and Twitter, and to "follow" the organization. The buttons are deployed on a website by including JavaScript code on the website that includes code from the AddThis website. Unknown to organizations using the technology, AddThis modified its code to include a technology called "canvas fingerprinting" [29] that allowed AddThis to uniquely identify and track every website visitor, irrespective of the use of "private browsing," cookie deleting, or other privacy-signaling mechanisms. Because AddThis was used by thousands of top websites, it allowed AddThis to correlate browsing activity across a large percentage of the Internet's users and properties [30]. Following the publicity of the tracking technique, some websites removed the AddThis technology.

1.9 Verizon "Perma-Cookie" (2014) *A1 Surveillance*

Scale:10^8 Purview: organization; Awareness: yes; Goal of Identifiability: yes

Wireless provider Verizon injected a new header ("X-UIDH") in unencrypted Hyper Text Transfer Protocol (HTTP) requests sent from Verizon cell phones to websites. The header, which was only sent for requests sent over the carrier's wireless network (as opposed to Wi-Fi), contained a device-specific header that did not change, allowing websites to correlate activity from individual cell phones as the cell phone moved from place to place [31]. Further testing revealed that AT&T also experimented with device-specific headers; AT&T stopped this practice in November 2014 [32]. In January 2015 Verizon announced that it would allow users to opt out of the UIDH advertising program [33]; as of June 2018, Verizon's customer support website indicated that the UIDH advertising program was still operational, but that the UIDH headers were only sent on a limited basis [34].

1.10 Nomi Technologies Wi-Fi Marketing (2015) *A1 Surveillance*

Scale:10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

Nomi Technologies developed technologies for tracking consumers entering stores based on identifiers transmitted by their mobile phones. Nomi posted a privacy statement on its website indicating that consumers could opt out of the tracking process on the Nomi website or in person at the stores; however, no opt-out process existed at the stores. FTC brought an action against Nomi and negotiated a settlement in which Nomi acknowledged misleading customers, promised that it would not mislead customers in the future, and agreed to FTC monitoring of its public statements and consumer complaints relating to the FTC action for a period of five years [35].

1.11 Pearson Twitter (2015) *A1 Surveillance*

Scale:10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

Pearson is a British-owned US company that publishes educational materials and assessment tests. In March 2015 the company informed the superintendent of a New Jersey high school district that one of the school district's students posted information about the Partnership for Assessment of Readiness for College and Careers (PARCC) test to Twitter [36]. The company issued a statement stating that it was "contractually required by states to monitor public conversations on social media to ensure that no assessment information (text, photos, etc.) that is secure and not public is improperly disclosed" [37]. The American Federation of Teachers issued a statement criticizing Pearson for not signing the Student Privacy Pledge "designed to limit the collection, maintenance and use of student personal information" [38].

2. B — Information Processing

This section presents incidents involving the information processing activities of data holders. In these cases, the incident resulted not from the collection of the data, but from its inappropriate use. We characterize these incidents using Solove's taxonomy, creating five potential harms: aggregation, identification, insecurity, secondary use, and exclusion.

2.1 Facebook News Feed (2006) *B1 Aggregation*

Scale:10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

Two years after its founding, Facebook launched News Feed, a new service within Facebook that aggregates status updates and changes in one's "friends" and places them on each user's Facebook home page. Previously users needed to check on each of their friends' "walls" to see what they were doing. News Feed automatically aggregated all of this information. News Feed had the result of making information evident that was previously accessible but not prominently featured. For example, parents received details of their children's lives that they previously had to seek out, potentially revealing more information than desired or expected. Facebook founder and CEO Mark Zuckerberg issued an apology for not including sufficient privacy controls into News Feed, saying "we really messed this one up" [39]. Facebook kept the News Feed as one of the primary ways that users interact with the website and attempted to address the privacy issues by adding a steadily growing and changing number of end-user controls [40].

2.2 Verizon Marketing with Consumer Information without Opt-Out (2006), *B4 Secondary Use.*

Scale:10^6; Purview: organization; Awareness: yes; Goal of Identifiability: yes

The FCC holds that the Communications Act requires approval from consumers before a carrier can use consumer information for marketing purposes. However, between 2006 and 2008 Verizon used customer proprietary network information (CPNI) for marketing without first allowing the consumers to opt out or opt in. Verizon discovered the privacy event in September 2012 and reported it to the FCC on January 18, 2013 (126 days later). Verizon settled with the FCC, agreeing to pay $7.4 million, to create an internal compliance program, and "to notify consumers of their opt-out rights on every bill for the next three years" [41], [42].

2.3 Facebook Beacon (2007) *B4 Secondary Use*

Scale:10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

Facebook Beacon was an advertising tracking system that monitored what a Facebook user purchased on a non-Facebook website and then reported purchases in the news feeds of the user's friends.[43] For example, Beacon could report to a user's friends when the user rented a

movie at Blockbuster, purchased a movie ticket at Fandango, or purchased an engagement ring. An investigation by Computer Associates found that Facebook received information from partner websites even when the Facebook user had logged out of Facebook.[44] Facebook Founder and CEO Mark Zuckerberg apologized: "We've made a lot of mistakes building this feature, but we've made even more with how we've handled them" [45].

Facebook terminated Beacon in September 2009 and paid $9.5 million to resolve a class-action lawsuit resulting from the introduction of the service [46].

2.4 MIT Gaydar: Facebook friendships expose sexual orientation (2009) B1 Aggregation

Scale:10^3; Purview: few; Awareness: yes; Goal of Identifiability: yes

Students at the Massachusetts Institute of Technology developed a statistical model based on data from the Facebook social network graph that could accurately predict the sexual orientation of MIT community members. This was significant, as the model could predict the orientation even in cases where the individual chose not to make that information public. The model required a valid Facebook account within the MIT network in order to access the complete list of an individual's Facebook "friends" [47].

The students trained the computer program on the social networks of 1,544 men whose Facebook profile indicated they were straight, 21 whose profile said they were bisexual, and 33 whose profiles claimed to be gay. They then tested the program on 947 men who did not report their sexuality on Facebook. The students reviewed 10 people in the sample whom they knew to be gay, and the program identified all 10 as being gay [48]. The project was heralded as an example of the power of social network analysis, and the students' faculty advisors reported on numerous occasions that leaders in the MIT gay community confirmed that the program could identify people who did not make their sexual orientation public.

2.5 Target Pregnancy Forecasting (2010) B1 Aggregation

Scale:10^6; Purview: few; Awareness: yes; Goal of Identifiability: yes

At a talk at Predictive Analytics World, Andy Pole, a statistician at Target, explained how the company could infer whether some customers were pregnant by a sudden change in their buying habits [49]. By discovering that women started purchasing unscented hand lotions and some vitamins, Pole said that Target could proactively send the women coupons for baby items. The story was largely unnoticed at the time but received significant attention after an article appeared two years later in *The New York Times* [50] and an article about the *Times* article appeared in *Forbes* [51]. According to the *Times* article, women establish new buying habits when pregnant, and these new habits may last for 10 or more years, so

companies like Target are highly motivated to influence those habits to the company's advantage.

According to the *Times* article, a teenage girl's father in Minnesota received coupons and called Target to complain about inappropriate coupons, after which the father called back to apologize when he learned that his daughter was actually pregnant. Other customers were similarly spooked, reported *The Times*, so Target started balancing the targeted advertisements with advertisements for lawn mowers so that the targeted women would think that their receiving maternity-themed coupons was a sheer coincidence.

Target refused to meet with the *Times* reporter while he was working on his story, and some commentators alleged that the Minnesota story is apocryphal, since the *Times* did not name the Target executive who provided the apocryphal anecdote [52].

2.6 Apple iPhone Tracking Location (2011) *B1 Aggregation*

Scale:10^7; Purview: no one; Awareness: no; Goal of Identifiability: no

A programming error on Apple's iPhone operating system caused the phone to remember the time and date of every Wi-Fi hotspot and cell phone tower it encountered. (The operating system collected this information and reported it back to Apple to assist in geolocation.) Users discovered that the iPhone's database was copied when the phone was backed up to a desktop, and from there the database could be accessed by others, providing a database of where the user had been. After the bug was publicly disclosed, Apple acknowledged the error and issued a software update so that the iPhone (and the iPhone backups) would not retain more than seven days of data [53].

2.7 Uber "Rides of Glory" (2012) *B4 Secondary Use*

Scale:10^5; Purview: small group; Awareness: yes; Goal of Identifiability: yes

In a blog post, Uber's data science team showed how data from the ride-hailing service could be used to find customers who spent the night at a place other than their primary residence (with implicit sexual overtones). The original blog post was removed after negative publicity regarding the misuse of transactional data [54], [55].

2.8 PaymentsMD Improper Collection (2012) *B4 Secondary Use*

Scale:10^3; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In 2014 the FTC filed a complaint against PaymentsMD, alleging that the company's consumer-facing payment processing website solicited consent from consumers to obtain their complete medical record, which PaymentsMD then used to build an electronic heath record (EHR) for a new business opportunity that the company was pursuing. The FTC alleged that consumers were deceived and misled into providing consent, even though consent was

not required for the purpose of bill presentation and payment. Under the terms of the 2014 settlement between PaymentsMD and its former CEO Michael C. Hughes, the company was forced to destroy the healthcare information that it collected and was prohibited from engaging in similar practices in the future [56].

2.9 Facebook "Year in Review" (2014) B1 Aggregation; B4 Secondary Use

Scale:10^8; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In 2014, Facebook launched a "Year in Review" feature on its website that automatically evaluated photographs from each user's photo history and prepared a collage customized for each user with the default tagline "It's been a great year. Thanks for being a part of it." These collages were prepared for all users, without opting in, and there was no way to opt out. In some cases, the photos produced pain and suffering as they reminded Facebook users of tragic events [57]. One example commonly cited was that of Eric Meyer of Cleveland Heights, Ohio, whose daughter had died from brain cancer during 2014: the daughter's photo was prominently featured in the automatically generated collage with the tag line "See Your Year" [58], [59]. A Facebook product manager who oversaw the project apologized to Meyer and said that the company would do better in the future [60].

Facebook continued the practice of preparing "year in review" slideshows. Today it is common for photo management system from Facebook, Google, and Apple to show users images from a few years earlier with the hope of triggering pleasant memories and increasing user interaction with the system.

3. C — Information Dissemination

Privacy incidents can result when an organization legitimately entitled to holding personal data releases that data in an inappropriate manner. Solove identifies seven categories of information dissemination harms: breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, and distortion.

3.1 Lotus Marketplace (1990*) C4 Increased Accessibility*

Scale:10^8; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In April 1990 the Lotus Development Corporation [61] announced a partnership with Equifax to create Lotus Marketplace:Households, a targeted marketing platform. Delivered on CDROM, the system contained a compact disc database with data on 150 million individuals, each categorized by name, address, age, gender, marital status, household income, 50 lifestyle categories, and buying propensity for 100 specific products [62]. The $695 product include a software meter that would allow users to search on any field but only generate name and address reports of 5000 individuals.

Privacy activists expressed grave concern with Marketplace:Households shortly after it was announced. They noted that even though these kinds of data had long been collected and used by service providers to create targeted prospect lists for mailing and telemarketing, this would be the first time that the entire database would be put directly in the hands of users, allowing them to perform searches and generate lists without any oversight.

Lotus countered, saying that it considered privacy issues in producing the product, planned to limit purchasers to corporations, and required a license agreement that prohibited specific uses of the data. Lotus also claimed that protection mechanisms built into the software prevented a user from simply extracting the entire database—a claim that activists disputed. Privacy activists also said that consumers would be unable to opt out once each quarterly CDROM had been produced. The controversy ignited a grassroots campaign against the product. Lotus received more than 30,000 email messages from consumers demanding that their names be removed from the database. Soon afterward, Lotus terminated the project in October 1990, without ever releasing the product.

3.2 Massachusetts GIC (Group Insurance Commission) (1996), *C1 Breach of Confidentiality*

Scale:10^4; Purview: organization; Awareness: yes; Goal of Identifiability: no

In 1996 the Massachusetts Group Insurance Commission released a dataset to healthcare researchers of records belonging to Massachusetts state employees who had been hospitalized. Then-governor of Massachusetts William Weld championed the release. “He said privacy would be protected because all identifiers had been eliminated from the records” [63]. Specifically, the state de-Identified the records by removing each employee’s name and address, but the employee’s date of birth, ZIP code, and sex remained to allow for statistical analysis. Latanya Sweeney, then an MIT graduate student, obtained a copy of the GIC data and decided to look for the medical records of Governor Weld, who was hospitalized after collapsing during a graduation ceremony at Bentley College on May 18, 1996. Knowing that Weld lived in Cambridge and was almost certainly a registered voter, Sweeney purchased the city of Cambridge’s voter rolls for $20, used them to learn Weld’s birthday and ZIP code, and then used this information to find the corresponding medical records in the GIC data set [64], [65]. Partly as a result of this study, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule later established a de-identification standard requiring suppression of 18 different data fields, including days and months, and generalization of ZIP codes to the first three digits [66].

3.3 Eli Lilly Prozac Mailing List (2001) *C1 Breach of Confidentiality*

Scale:10^2; Purview: few (1), Awareness: no; Goal of Identifiability: no

Between March 15, 2000 and June 22, 2001, Eli Lilly and Company, a major US pharmaceutical company, operated an e-mail reminder service that allowed patients to sign

up on Prozac.com to receive messages reminding them to refill their prescriptions for the antidepressant. On June 27, 2001, a Lilly employee sent an email message to all of the service's 669 subscribers informing them that the service was discontinued. Unfortunately, the email message listed all of the subscribers in the "To:" field of the email message, effectively providing each subscriber with a complete list of the other service subscribers. The FTC negotiated a settlement with Lilly in which the company agreed to establish an information security program to protect personal data [67].

3.4 JetBlue Releases Customer Data to DHS (2002) *C1 Breach of Confidentiality*

Scale:10^6; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In September 2002, jetBlue Airways provided 5 million "passenger name records" to Torch Concepts, a defense contractor developing a counterterrorism tool based on "data pattern analysis" [68].

Following the terrorist attacks of September 11, 2001, Torch Concepts of Huntsville, AL, approached the Department of Defense (DoD) with a proposal to use "data pattern analysis" to evaluate the risk posed by visitors to DoD installations. Briefly, the approach was to combine consumer reporting and demographic information with travel information to create risk analysis models. DoD added Torch Concepts as a subcontractor to an existing contract in March 2002 to perform a "limited initial test" of the technology. Torch made numerous approaches to federal agencies to obtain information but was unsuccessful. After unsuccessfully approaching American Airlines and Delta Airlines, Torch contacted the Department of Transportation (DOT) and the Transportation Security Agency (TSA). Finally, after it was approached by "a relatively new" TSA employee, jetBlue agreed to provide Torch with assistance. Torch engaged the data aggregation firm Acxiom to handle aspects of its data processing. In September 2002 jetBlue provided Acxiom with five million records, representing 1.5 million passengers. In October 2002, Torch purchased additional information from Acxiom.

Based on its analytics, Torch prepared a presentation concluding that "several distinctive travel patterns were identified" in the data and that "known airline terrorists appear readily distinguishable from the normal jetBlue passenger patterns" [69]. This presentation was eventually discovered on the Internet by members of the public and the media. As a result, the Department of Homeland Security (DHS) Privacy Office investigated. The DHS Privacy Office concluded that while TSA employees were involved in the data transfer and "acted without appropriate regard for individual privacy interests or the spirit of the Privacy Act of 1974," no actual Privacy Act violation had taken place, since the data were transferred directly from jetBlue to Acxiom.

3.5 Release of "de-identified" AOL search logs for research (2006) *C2 Disclosure*

Scale:10^5; Purview: group; Awareness: yes; Goal of Identifiability: no

A group of researchers at the consumer Internet provider America Online (AOL) released a series of search queries made by AOL subscribers using the AOL search engine to assist academic researchers working in the area of Internet search and text retrieval. Prior to release, AOL removed the users' identifying information and replaced it with a randomly generated pseudonym so that subsequent searches by the same individual could be correlated. Journalists were able to identify several users from their search terms and contacted the users to verify the re-identification. "There are also many thousands of sexual queries, along with searches about 'child porno' and 'how to kill oneself by natural gas' that raise questions about what legal authorities can and should do with such information," read an article in *The New York Times* [70], [71]. Although AOL apologized for the release [72], researchers noted that other Internet search engines had released de-identified user search histories in 1999 and 2001[73].

Following the release, a class action lawsuit, *Landwehr v. AOL Inc.,* alleged that AOL violated specific privacy and consumer protection laws by publicly releasing some of its users' search queries. On May 28, 2013, a federal court approved a settlement of up to $5 million in the case in which AOL admitted no wrongdoing. A settlement fund allowed affected AOL users to claim up to $100 each [74], [75].

3.6 Netflix Prize (2006) *C2 Disclosure*

Scale:10^5; Purview: organization; Awareness: yes; Goal of Identifiability: no

To spur academic research in data mining, the video rental firm Netflix released customer video rental histories for roughly 480,000 Netflix customers that included the rental date and the customer's rating. Netflix tried to protect customer privacy by replacing customer names with a unique number. The dataset included no other direct identifiers. Netflix offered a prize of $1 million to the winning team that could develop a recommendation algorithm that performed better than the internet Netflix algorithm. Arvind Narayanan and Vitaly Shmatikov, two graduate students at the University of Texas, developed an approach for identifying some of the records in the Netflix Prize dataset by correlating the video ratings with ratings in IMDb, a publicly available dataset. Unlike the Netflix set, the IMDb dataset also included the names or other identifiers of the individuals who performed the ratings. By using the IMDb dataset, the researchers showed that they could discover additional movies that a Netflix subscriber might have watched but not publicly rated on IMDb. As a result of the release of the Netflix data and the company's announcement of a second contest, the FTC sent a letter of inquiry to Netflix [76], and a class action lawsuit accused Netflix of violating fair-trade laws and the Video Privacy Protection Act [77]. Four months later, Netflix announced that it had settled the lawsuit and canceled the second contest [78].

3.7 Google Street View Photography Capture (2007) *C4 Increased Accessibility*

Scale:10^8; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In 2007, Google released Street View, street-level photographs of streets, houses and businesses taken from vehicles that Google had driven in major US cities. (The program was later expanded worldwide.) Privacy activists criticized Street View for collecting these photographs without permission. Although the photographs had been taken from public streets, Street View made it possible for people around the world to see and share imagery that was previously difficult to collect. Because Street View was based on automated collection and processing of the geolocation-tagged images, many potentially embarrassing images were found and publicized by the general public before they could be reviewed and removed by Google. Privacy activists also raised concerns about the fact that Google's vehicles took photographs inside of houses if the windows were open. Google responded by blurring faces and license plates and creating a mechanism for individuals to request that images be removed [79], [80]. The government of Italy levied a €1 million ($1.4 million) fine against Google for taking Street View photographs from cars not clearly marked as belonging to Google, and for the interception of unencrypted Wi-Fi signals [81].

3.8 Jerk.com (2009-2015) C.5 Blackmail; *C6 Appropriation*

Scale:10^7; Purview: small organization; Awareness: yes; Goal of Identifiability: yes

In a 2014 enforcement action, the FTC found that John Fanning created the website Jerk.com, downloaded up to 85 million individual user profiles from Facebook, labeled some of the people as a "Jerk" or "not a Jerk," and then offered users $30 "memberships" to his website. The memberships allegedly gave Jerk.com users the ability to "manage your reputation" and to "dispute" the information posted online [82]. The FTC ruled against Jerk LLC in March 2015. Fanning appealed the FTC decision to the First Circuit Court of Appeals, which affirmed the Commission's summary decision except for the provision regarding ongoing monitoring of Fanning.

3.9 Facebook "Like" (2009) *C1 Disclosure*

Scale:10^8; Purview: organization; Awareness: yes; Goal of Identifiability: yes

In 2009 Facebook created a "like" button which allowed users of the Facebook platform to click "like" on items in their Facebook News Feed to indicate that they approved of a posting or message. Code on the Facebook platform collected all of the Facebook users who "liked" an item and displayed the total numbers to Facebook users, as well as the names of any of their friends who might like something. The Facebook "like" button also allowed third parties to place "like" buttons on their own web properties. As with the "like" button in the newsfeed, Facebook users visiting those third-party websites would see how many other users "liked" the web property, and who among those likers were their Facebook "friends." As of this article's publication, the Facebook "like" button was still operational.

3.10 CVS Caremark and Rite Aid Improper Disposal of Sensitive Documents (2009, 2010) *C1 Breach of Confidentiality*

Scale:10^7; Purview: small groups; Awareness: yes; Goal of Identifiability: no

Investigations by the FTC and the US Department of Health and Human Services found that both CVS Caremark [83] and Rite Aid [84] improperly disposed of pharmacy-related information in open dumpsters behind their stores. In the case of CVS, "media reports from around the country [indicated] that its pharmacies were throwing trash into open dumpsters that contained pill bottles with patient names, addresses, prescribing physicians' names, medication and dosages; medication instruction sheets with personal information; computer order information from the pharmacies, including consumers' personal information; employment applications, including social security numbers; payroll information; and credit card and insurance card information, including, in some cases, account numbers and driver's license numbers." In the case of Rite Aid, the personal information included "pharmacy labels and job applications."

In both cases the pharmacy chains were penalized by the FTC for deceptive trade practices, both having publicly claimed to respect consumer privacy and to properly safeguard protected health information. CVS, the largest pharmacy chain in the US, "agreed to pay $2.25 million and implement a Corrective Action Plan to ensure that it will appropriately dispose of protected health information such as labels from prescription bottles and old prescriptions" [85]. Rite Aid and 40 affiliated entities agreed to pay $1 million, as well as "to take corrective action to improve policies and procedures to safeguard the privacy of its customers when disposing of identifying information on pill bottle labels and other health information" [86].

3.11 Google Buzz (2010) *C1 Disclosure*

Scale:10^8; Purview: small group; Awareness: yes; Goal of Identifiability: yes

In 2010 Google launched a social networking service called Google Buzz as a complement to its Gmail e-mail platform. When Google users logged into Gmail on the day Buzz launched, they were encouraged to automatically sign up for Buzz. Users who signed up for Buzz were automatically configured to "follow" the Gmail users that they "email and chat with the most," and this list of followers became publicly available, violating Gmail's privacy policy. Information in some users' Buzz public profiles was augmented with information from other Google products, including Picasa (photo sharing) and Reader (news reading). In February 2011 the Electronic Privacy Information Center filed a complaint before the FTC requesting an investigation against Google. Google and the FTC reached a preliminary agreement March 2011 and a final agreement in October 2011, in which Google agreed to establish a comprehensive privacy program and be subject to regular, independent privacy audits for 20 years [87].

3.12 Snapchat (2011) *C1 Breach of Confidentiality*

Scale:10^6; Purview: small group; Awareness: yes; Goal of Identifiability: yes

In 2011, mobile application developer Snapchat launched a service that allowed users to send "disappearing" photos to each other. By default, the photos were visible for 10 seconds, after which time the company promised the photos would be deleted. In its FAQ the company promised that "snaps disappear after the timer runs out" and stated they could not be recovered. In fact, the snaps were not removed from the consumer phones, only made invisible. Furthermore, the snaps remained accessible on the company's servers through an API. Snapchat further promised that it did not "ask for, track, or access any location-specific information from your device at any time while you are using the Snapchat application," when it fact it had integrated an Android analytics tracking service into its application.

The FTC filed a complaint against SnapChat. As a result of the complaint, SnapChat agreed to establish a comprehensive privacy program and submit to third-party monitoring of its privacy practices for a period of 20 years, and direct monitoring by the FTC for a period of five years [88].

3.13 Uber "God View" (2011) *C1 Breach of Confidentiality*

Scale:10^5; Purview: small group or organization; Awareness: yes; Goal of Identifiability: yes

Uber's "God View" is a tool that showed the present location of every Uber vehicle in a geographical area. The company developed the tool to allow it to see operations and direct vehicles to areas that lacked service. Reportedly the tool allowed the operator to see the names of individuals in Uber vehicles, or to monitor a rider's history. In 2011 the tool was demonstrated at a party [89]. In 2013 a person who interviewed for a job at Uber's Washington office was given access to the "God View" application for a full day following the person's interview, during which time he was able to review travel records of people that he knew [90].

In a letter to Senator Al Franken [91], Uber confirmed that it had created a tool that allowed individuals in Uber's operations department to view the location of every car, and admitted that one of its employees had looked at real-time information on a journalist.

The Federal Trade Commission conducted several investigations of Uber as a result of the company's business practices and a 2014 data breach. On January 19, 2017, Uber agreed "to pay $20 million to settle FTC changes that it recruited prospective drivers with exaggerated earnings claims" [92]. The settlement was amended in August 2017 to include 20 years of privacy assessments by a third party [93]. The settlement was amended yet again in April 2018 to cover an unrelated cyberattack that took place in 2016 that was "strikingly similar [to the] 2014 breach" [94].

3.14 Location Sharing in Facebook Messenger (2012). *C2 Disclosure*

Scale:10^8; Purview: small group; Awareness: yes; Goal of Identifiability: yes

Facebook Messenger is an instant messaging service that allows users to communicate with each other using a "chat" interface. From 2011 until May 2015, Facebook tagged every message sent by Facebook's Android mobile app with the location of the sender. These locations were shared with all users in a group chat, irrespective of the users' relationships or privacy settings. The location sharing was discovered by Aran Khanna, a Harvard University undergraduate, who hypothesized that there was no public outcry regarding the casual sharing of location information because users were either not aware of the sharing or were not concerned about the collection and visibility of their locational data. Khanna developed a tool that allowed users to display a map of all of the location data they shared with other users through Facebook Messenger chats—information that was already available through the Facebook user interface, but in a more aggregated form. The tool was downloaded more than 85,000 times, and more than 170 global news publications wrote about the article. Nine days after the tool's release, Facebook made location sharing an opt-in feature, demonstrating that "sufficient public attention may be necessary for redress of reported privacy concerns" [95].

3.15 Release by Washington State of de-identified Patient Health Records (2013). *C1 Breach of Confidentiality*

Scale:10^1; Purview: organization; Awareness: yes; Goal of Identifiability: no

Acting under state law, Washington State released de-identified hospital discharge records to assist in healthcare policy analysis. Researchers demonstrated that discharge records for hospitalizations resulting from accidents could occasionally be re-identified manually by correlating information in the discharge records with newspaper articles describing the accident that caused the hospitalization [96].

3.16 Revenge Porn (2015) *C3 Exposure, C6 Appropriation*

Scale:10^3; Purview: few (1); Awareness: yes; Goal of Identifiability: yes

Website operator Craig Brittain had been operating a so-called "revenge porn" site that solicited nude photographs of women and posted the photographs with the women's names. Brittain also operated other websites, "Takedown Hammer" and "Takedown Lawyer," which accepted money from victims and caused the photos to be taken down. Brittain agreed to refrain from posting nude photographs or videos of people without their affirmative consent. Brittain also agreed to 10 years of monitoring by the FTC of any new business he started or employment that he took [97].

3.17 Healthcare.gov ad tracking (2015) *C1 Breach of Confidentiality*

Scale:10^6; Purview: small group; Awareness: yes; Goal of Identifiability: yes

Analytics software deployed on the US Government's Healthcare.gov website transmitted personal data, including age, smoking status, pregnancy status, parental status, zip code, state, and income to at least 14 third-party analytics and marketing firms [98]. Following Congressional hearings, the website's operators responded by adding a privacy control panel that allows web visitors to disable tracking from their computer [99].

4. D — Invasion

Invasion is a fundamentally different kind of privacy offense than those examined above. While Collection, Processing, and Dissemination all involve information that's *taken from* a data subject, invasion involves *doing something to* the data subject. Invasion directly impact the subject and forces a reaction. Note that there is a significant difference between invasion and interrogation: although both are the result of an interaction between the data subject and the perpetrator, in invasion the harm is caused by the interaction, while in interrogation the purpose of the interaction is the extraction of personal information.

4.1 Commercial Spam Email (1995-) *D1 Invasion*

Scale:10^8; Purview: organization (small); Awareness: yes; Goal of Identifiability: no

Large-scale commercial use of unsolicited email started in 1994 [100] and grew rapidly. Spam mail was highly intrusive from approximately 1995 to 2005, until security and filtering techniques largely prevented it from reaching the inboxes of victims. Today spam mail is widely sent but is more frequently an annoyance, as filtering occasionally causes legitimate mail to be missed. In 2010, studies found that 88 percent of the worldwide email traffic was spam, amounting to roughly 90 billion email messages sent to valid email addresses each day [101]. A study published in 2012 estimated that the cost of spam to American firms and consumers was almost $20 billion annually, while spammers and spam-advertised merchants received less than $200 million per year as a result of their efforts. "Thus, the 'externality ratio' of external costs to internal benefits for spam is around 100:1" [102].

4.2 Facebook Get Out the Vote experiment (2010) *D2 Decisional Interference*

Scale:10^7; Purview: organization; Awareness: yes; Goal of Identifiability: yes

Researchers from the University of California, San Diego and Facebook conducted a randomized controlled trial on 61 million Facebook users during the 2010 US congressional elections to see if they could motivate individuals to vote. Some users saw messages in their newsfeed allowing them to post that they voted to their newsfeeds and showing them the names and faces of their Facebook friends who voted. The results indicated that those who saw the message were 0.39% more likely to vote than those who received no messages at all.

"First and foremost, online political mobilization works. It induces political self-expression, but it also induces information gathering and real, validated voter turnout," the authors of the study noted. "Furthermore, as many elections are competitive, these changes could affect electoral outcomes. For example, in the 2000 US presidential election, George Bush beat Al Gore in Florida by 537 votes (less than 0.01% of votes cast in Florida). Had Gore won Florida, he would have won the election" [103]. The implication is that, by determining the political leanings of an individual and then targeting specific individuals with "get out the vote" messages, major social media providers might be able to influence the outcome of closely contested elections.

4.3 Dialing Services, LLC automated calls to cellphones (2011) *D1 Invasion*

Scale:10^6; Purview: organization; Awareness: yes; Goal of Identifiability: no

The FCC alleged that Dialing Services, LLC placed automated phone calls with "artificial or prerecorded" messages to millions of wireless phones without authorization, a violation of the Communications Act and the Commission's rules that prohibit "robocalls" and "autodialed calls" to wireless phones when not made for emergency purposes or with prior express consent [104], [105], [106].

In 2012 the FTC announced a series of contests and challenges to spur inventors to develop technical solutions for fighting automated callers [107]. The grand prize went to Daniel Klein and Dean Jackson for a system called Nomorobo, which used simultaneous call technology to suppress robot calls from phone numbers that appeared on a blacklist [108]. However, within a few years robot calls were once again a major problem, largely the result of technology allowing robot callers to spoof caller-ID technology [109].

4.4 Sprint "Do Not Call" violations (2011, 2014) *D1 Invasion*

Scale:10^6; Purview: no one; Awareness no; Goal of Identifiability: no

Sprint Corporation placed unwanted marketing calls and texts to consumers who requested to be placed on the company's "do not call" list. In 2011 Sprint paid the FCC a $400,000 fine following the negotiation of a consent decree. However, Sprint continued to place phone and text messages to consumers, a violation of the Telephone Consumer's Privacy Act. Three years later, Sprint settled with FCC, agreeing to pay an additional $7.5 million and to implement "a two-year plan to ensure compliance with FCC requirements designed to protect consumer privacy and prevent consumers from receiving unwanted telemarketing calls" [110].

4.5 Facebook Emotional Contagion Experiment (2012) *D2 Decisional Interference*

Scale:10^5; Purview: small group; Awareness: yes; Goal of Identifiability: no

Facebook intentionally manipulated the news feeds of 689,003 Facebook users to determine if it could change their emotions by controlling the information they saw. In an article published in the *Proceedings of the National Academy of Sciences of the United States of America*, the authors concluded: "We show, via a massive (*N* = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues" [111].

Following the publication, there was considerable outcry in both the news media and the academic community that the researchers experimented on Facebook users without their permission and without giving the users the ability to opt out. Furthermore, even though two of the study's authors were affiliated with Cornell University, the Cornell Institutional Review Board, the organization at Cornell that reviews human subjects research, did not approve the study. Following the outcry, the editor of *PNAS* published an "Editorial Expression of Concern and Correction." The editorial noted that the original paper stated that the research "was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research." However, Facebook only added the term "research" to its data policy four months after the study took place [112]. Furthermore, the authors reported to the journal that "[b]ecause this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell's Human Research Protection Program" [113].

4.6 Spying/Stalking Apps on Mobile Phones (2015) *D1 Intrusion*

Scale:10^6; Purview: organization; Awareness: yes; Goal of Identifiability: yes

There is growing attention to apps on mobile phones that covertly collect geolocation, application use, screen displays, and user interaction and send this information to third parties. Such apps are reportedly used by men to spy on their ex-girlfriends and by employers to spy on their employees [114]. The FTC publishes consumer information for victims of violence and stalking [115].

Analysis of Events

An aggregate-level examination of the characteristics of the privacy events reveals several trends. First, while the timeline includes 32 events from 1990 through 2015, 17 of those events occurred between 2010 and 2015. The number of privacy events per year increased as did the number of years per decade that experienced events. This trend of increasing non-breach privacy events is shown in Figure 3.

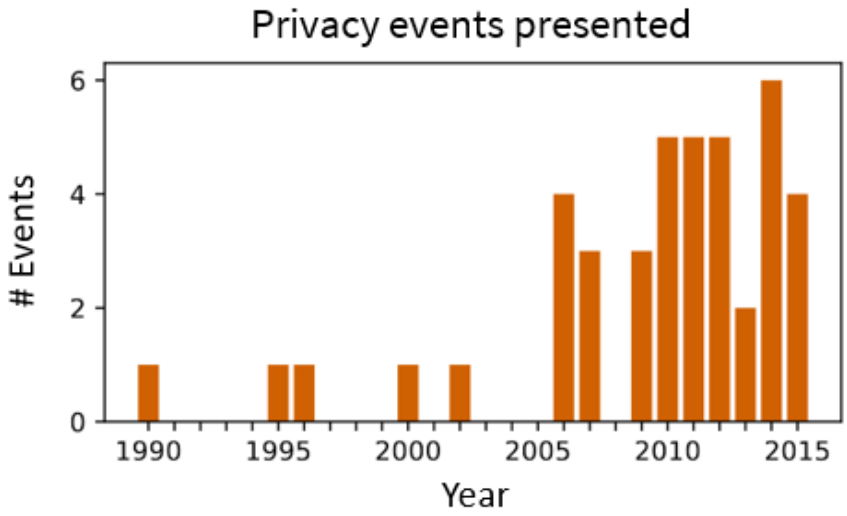


Figure 3: Privacy events between 1990 and 2015.

Discussion

Some of the major lessons from these examples include:

- **Events with significant scale can result from relatively small purview.** Both Google Street View Wi-Fi and Apple iPhone Tracking show that a few engineers can have an outsized impact on the use of personal information by an organization, demonstrating the need for internal controls. It may also indicate that organizations need to develop technology that can detect, monitor and measure data flows, as data collection and misuse may not be detected merely by the review of documentation and discussions with engineers.

- **Technology allows small organizations to have disproportionate impacts.** The AddThis and Jerk.com incidents show that relatively small organizations can affect millions of people. This suggests the need for controls on the movement of data and code between an organization, its suppliers, and its customers. Because data and code can change in an instant, organizations may choose to develop technology and management controls that supplement the legal agreements typically used to broker these relationships.

- **Although human error and poor information practices are frequently the root cause of data breaches, they are not significant causes of the events presented here.** This is in part a result of our selection criteria. Although many media reports of privacy problems focus on the impact of hackers, data breaches, and misconfigured systems, this catalog of incidents shows that poor privacy outcomes can be the result of intentional acts by individuals or organizations. Some of these acts had

unintentional or unforeseen results, but others were deliberate acts undertaken for a specific purpose.

- **University Institutional Review Boards (IRBs) did not prevent the privacy harms described here.** Several of the incidents that we presented, including MIT Gaydar, Harvard Covert Photography, and Facebook Emotional Contagion, involved researchers at U.S. universities where an Institutional Review Board oversees human subjects research. Although legally the U.S. Common Rule only applies to federally funded human subjects research, most U.S. universities apply the Common Rule to all human subjects research, and oversee that research by an IRB. The Common Rule defines a *Human Subject* as "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information" [116]. Although this would seem to provide some amount of privacy protection, in the cases we examined, the relevant IRBs did not prevent the privacy harms from taking place. For example, the Facebook emotional contagion study was not reviewed by the Cornell IRB because the Cornell researchers were only provided with de-identified data not traceable back to a given individual, and such data are typically excluded by the Common Rule.

- **Context is creating more privacy spheres than "public" and "private."** The traditional notion that information is either "public" or "private" frequently is insufficient to convey both individual expectations and the effective reality of information processing. Several of our incidents, such as the Facebook News Feed and Harvard Covert Photography, show that there is increasingly a middle ground between the two. Public information that is hard to obtain is effectively private or restricted in many circumstances; aggregating this information and making it widely available has a practical impact on individual privacy.

- **Websites that governments use to communicate with their citizens can inadvertently become conduits for collecting private information from citizens and providing it to commercial entities.** The complex nature of web technology combined with the desire of government agencies to publicize their work and monitor the effectiveness of those publicity efforts have resulted in personal information provided to government agencies becoming part of private information ecosystems. This happened without direct knowledge of the government agencies in the case of the Facebook Like button and the Healthcare.gov tracking. Also, because of the authoritative face of government, users are incentivized/forced to provide the most accurate information. This creates true self-reported data that the private sector covets. As a result, governments must be reactive. For example, in 2011 the German Data Protection Commissioner's Office (Independent Centre for Privacy Protection – ULD) in Schleswig-Holstein issued a ruling that required all official German government websites in the German state of Schleswig-Holstein to "shut down their

fan pages on Facebook and remove social plug-ins such as the 'like'-button from their websites," because the systems transferred personal information to the US in violation of the German Telemedia Act (TMG) and the Federal Data Protection Act (BDSG) [117].

- **Publicly revealed information can be used to infer traits that individuals consider private.** Predictive algorithms are increasingly able to infer information that individuals wish to keep private based on information or behaviors that the individuals do not consider private. Joint research by the University of Cambridge and Microsoft Research has shown that a wide variety of human traits and attributes can be predicted from an analysis of likes, including sexual orientation, race, and political affiliation, with more likes allowing predictions to be made more accurately. This was made evident in both the MIT Gaydar and the Facebook Like incidents. For example, a single Like allowed the model to predict a user's age with 25% accuracy, but with 20 likes the number rose to more than 50% [118].

The breadth of our compilation allows for qualitative analysis of the nature of privacy incidents over the past three decades. However, our list is not comprehensive, nor the result of random sampling, and thus a detailed statistical analysis will have limited value. Our list is also biased in favor of English media reports in the United States, and as such, the list may systematically miss privacy incidents more likely to occur in other countries.

Future analysis of this dataset might include more in-depth qualitative analysis using grounded theory techniques of open, axial, and selective coding. Applying qualitative research techniques such as coding, memoing and grounded theory will provide an analysis approach that identifies relationships in the data and among the codes. It would also be useful to expand this corpus with more internationally representative samples. Current efforts on privacy engineering and "privacy by design" focus on approaches for building privacy into design and development practices. This list shows that many privacy events are the result of deliberate decisions made by individuals or their organizations. As such, any "privacy by design" effort will need to incorporate a mechanism to assist organizations in identifying and measuring privacy risk, so that these potential events can be eliminated before they occur.

We presented 44 recent and historical high-profile cases involving privacy in the digital age, avoiding cases in which the privacy event was the result of a data breach. We hope that this list will prove useful to those who seek to understand the range of challenges facing organizations in the attempt to handle user data while respecting privacy.

References

1. Medical Privacy Stories, Health Privacy Project, August 2013. http://patientprivacyrights.org/wp-content/uploads/2013/08/True_Stories1.pdf

2. Resolution Agreements and Civil Money Penalties, US Department of Health & Human Services, Last Accessed February 10, 2016. http://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/agreements/

3. Privacy Incidents Database, Jessica Staddon, Linda Vue, Yuxu Yang, Heather Lipford and Bart Knijnenburg, 2015. https://sites.google.com/site/privacyincidentsdatabase/

4. Examining the costs and causes of cyber incidents, Sasha Romanosky, FTC PrivacyCon, January 14, 2016. https://www.ftc.gov/system/files/documents/public_comments/2015/10/00027-97671.pdf

5. Survey of Recent FTC Privacy Enforcement Actions and Developments, Fatima Nadine Kahn, The Business Lawyer, 68:1 (November 2012), pp. 225-232. http://www.jstor.org/stable/23527087

6. Press Releases, https://www.ftc.gov/news-events/press-releases

7. Enforcing Privacy Promises, Federal Tarde Commission, https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/enforcing-privacy-promises.

8. Search Results, https://www.fcc.gov/search/#q=%22for%20immediate%20release%22%20privacy&o=new

9. Enforcement Actions (By Date), https://transition.fcc.gov/eb/tcd/eabydate.html

10. A list of incidents resulting from improper sanitization of data on equipment that was later sold or disposed of can be found at http://forensicswiki.org/wiki/Residual_Data_on_Used_Equipment.

11. Foreign and Military Intelligence, Book 1: Final Report of the Select Committee to Study Governmental Operations with Respect To Intelligence Activities, United State Senate, April 26, 1976. Government Printing Office, Washington, DC. http://www.intelligence.senate.gov/sites/default/files/94755_I.pdf

12. A Taxonomy of Privacy, Daniel J. Solove, University of Pennsylvania Law Review, 154:3, pp. 447-560, January 2006, https://www.law.upenn.edu/journals/lawreview/articles/volume154/issue3/Solove154U.Pa.L.Rev.477(2006).pdf

13. Can We Sniff Wi-Fi: Implications of Joffe v. Google, S. Garfinkel and M. McCarrin, IEEE Security & Privacy, 12:4, July-August 2014, pp. 22-28. https://ieeexplore.ieee.org/document/6876254/

14. Notice of Apparent Liability for Forfeiture, In the Matter of Google, Inc., File No. EB-10-IH-4055, Chief, Enforcement Bureau, Federal Communications Commission, April 13, 2012. Full, unreacted report available at https://www.scribd.com/fullscreen/91652398.

15. DA 12-592: Notice of Apparent Liability for Forfeiture, Federal Communications Commission, 13 Apr. 2012; http://transition.fcc.gov/Daily_Releases/

16. "Enforcement Bureau issues $25,000 NAL to Google Inc.," US Federal Communications Commission, April 13, 2012. https://www.fcc.gov/document/enforcement-bureau-issues-25000-nal-google-inc

17. "Google pays $7 million to settle 38-state Wi-Fi investigation," Alexei Oreskovic, Reuters, March 12, 2013. https://www.reuters.com/article/us-google-wifi-fine-idUSBRE92B0VX20130312

18. Lower Merion School District and Blake Robbins Reach a Settlement in Spycamgate, Kasmir Hill, Forbes. October 11, 2010. http://www.forbes.com/sites/kashmirhill/2010/10/11/lower-merion-school-district-and-blake-robbins-reach-a-settlement-in-spycamgate/

19. Lower Merion District's Laptop Saga Ends with $610,000 Settlement, John P. Martin, Philly.com, October 12, 2010. http://articles.philly.com/2010-10-12/news/24981536_1_laptop-students-district-several-million-dollars

20. Lower Merion School District faces another webcam suit, Ashley Nguyen, Philly.com, December 8, 2011. http://www.philly.com/philly/blogs/neighbors/Lower-Merion-School-District-faces-another-webcam-suit.html

21. Sister of Lower Merion Teen in 2010 Spycam Case Files Her Own Lawsuit, Brad Segall, CBS Philly, December 9, 2011. http://philadelphia.cbslocal.com/2011/12/09/sister-of-lower-merion-teen-in-2010-spycam-case-files-her-own-lawsuit/

22. Newport Television to Pay $35,000 fine for Broadcasting Private Telephone Conversation, FCC News, November 28, 2014. https://apps.fcc.gov/edocs_public/attachmatch/DOC-330722A1.pdf

23. Millions of Android users 'deceived' by flashlight app that shares location with advertisers, Tom Warren, The Verge, Dec. 6, 2013. https://www.theverge.com/2013/12/6/5181472/brightest-flashlight-free-ftc-location-data-settlement

24. Android Flashlight App Developer Settles FTC Charges it Deceived Consumers, Federal Trade Commission, December 5, 2013. https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived

25. FTC Warns Children's App Maker BabyBus About Potential COPPA Violations, Federal Trade Commission, December 22, 2014. https://www.ftc.gov/news-events/press-releases/2014/12/ftc-warns-childrens-app-maker-babybus-about-potential-coppa

26. Google Suspends BabyBus Apps After FTC Warns of Privacy Violations, Wendy Davis, MediaPost, December 29, 2014. http://www.mediapost.com/publications/article/240860/google-suspends-babybus-apps-after-ftc-warns-of-pr.html?edition=

27. Yelp, TinyCo Settle FTC Charges Their Apps Improperly Collected Children's Personal Information, Federal Trade Commission, September 17, 2014. https://www.ftc.gov/news-events/press-releases/2014/09/yelp-tinyco-settle-ftc-charges-their-apps-improperly-collected

28. Bol Authorized Study that Photographed Faculty, Students in Class Without Notice, Dev A. Patel and Steven R. Watros, The Harvard Crimson, November 5, 2014. http://www.thecrimson.com/article/2014/11/5/bol-photograph-courses-hilt/

29. Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14). ACM, New York, NY, USA, 674-689. DOI=http://dx.doi.org/10.1145/2660267.2660347

30. Meet the Online Tracking Device That is Virtually Impossible to Block, Julian Angwin, ProPublica, January 21, 2014. https://www.propublica.org/article/meet-the-online-tracking-device-that-is-virtually-impossible-to-block

31. Find Out Whether This Unkillable Tracker is On Your Smartphone, Kashmir Hill, Forbes, October 28, 2014. http://www.forbes.com/sites/kashmirhill/2014/10/28/find-out-whether-this-privacy-killing-super-cookie-is-on-your-phone/

32. AT&T Stops Using Undeletable Phone Tracking IDs, Julia Angwin, ProPublica, November 14, 2014. https://www.propublica.org/article/att-stops-using-undeletable-phone-tracking-ids

33. Verizon will allow users to opt out of its 'permacookie,' Russell Brandom, The Verge, January 30, 2015. https://www.theverge.com/2015/1/30/7952233/verizon-permacookie-opt-out-tracking-privacy

34. Verizon Wireless' use of a Unique Identifier Header (UIDH), Verizon Wireless, last accessed June 24, 2018. https://www.verizonwireless.com/support/unique-identifier-header-faqs/

35. Nomi Technologies, Inc., In the Matter of, Docket No. C-4538, Decision and Order, United States of America Before the Federal Trade Commission, August 28, 2015, https://www.ftc.gov/system/files/documents/cases/150902nomitechdo.pdf,

36. Pearson monitoring social media for security breaches during PARCC testing, Valerie Strauss, March 14, 2015. The Washington Post. https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/14/pearson-monitoring-social-media-for-security-breaches-during-parcc-testing/

37. Pearson admits to monitoring students' social media use during its online tests, Nicky Woolf, The Guardian, March 18, 2015. https://www.theguardian.com/education/2015/mar/18/pearson-monitoring-students-social-media-tests

38. The power of Pearson threatens academic integrity, American Federation of Teachers, Summer 2015. https://www.aft.org/periodical/aft-campus/summer-2015/power-pearson-threatens-academic-integrity

39. An Open Letter from Mark Zuckerberg, Facebook, September 8, 2006. https://www.facebook.com/notes/facebook/an-open-letter-from-mark-zuckerberg/2208562130/

40. Strategies and Struggles with Privacy in an Online Social Networking Community, Katherine Strater & Heather Richter Lipford, Proceedings of HCI 2008: People and Computers XXII, The 22nd British HCI Group Annaul Conference, Liverpool John Moores University, UK. 1-5 September 2008; Understanding Privacy Settings in Facebook with an Audience View, Heather Richter Lipford, Andrew Besmer, Jason Watson, UPSEC 2008.

41. Verizon to pay $7.4 million to settle consumer privacy investigation, FCC News, September 3, 2014. https://apps.fcc.gov/edocs_public/attachmatch/DOC-329127A1.pdf

42. In the Matter of Verizon, Compliance with the Commission's Rules and Regulations Governing Customer Proprietary Network Information, Adopting Order, Federal Communications Commission, September 2, 2014. https://apps.fcc.gov/edocs_public/attachmatch/DA-14-1251A1.pdf

43. Facebook's Beacon More Intrusive Than Previously Thought, Juan Carlos Perez, PCWorld via IDG News Service, December 1, 2007. http://www.pcworld.com/article/140182/article.html

44. Facebook's Misrepresentation of Beacon's Threat to Privacy: Tracking users who opt out or are not logged in, S. Berteau, Computer Associates, November 29, 2007.

http://community.ca.com/blogs/securityadvisor/archive/2007/11/29/facebook-smisrepresentation-of-beacon-s-threat-to-privacy-tracking-users-who-opt-out-or-are-notlogged-in.aspx.

45. Thoughts on Beacon, Mark Zuckerberg, Facebook, December 5, 2007. https://www.facebook.com/notes/facebook/thoughts-on-beacon/7584397130/

46. Facebook shuts down Beacon, The Telegraph, September 21, 2009. https://www.telegraph.co.uk/technology/facebook/6214370/Facebook-shuts-down-Beacon.html

47. Gaydar: Facebook friendships expose sexual orientation, Carter Jernigan and Behram F.T. Mistree, First Monday, October 2009. http://firstmonday.org/article/view/2611/2302

48. Project 'Gaydar:' At MIT, an experiment identifies which students are gay, raising new questions about online privacy, Carolyn Y. Johnson, The Boston Globe, September 20, 2009.

49. How Target Gets the Most out of its Guest Data to Improve Marketing ROI, Andrew Pole, Senior Manager, Target, Predictive Analytics World October 2010, http://www.predictiveanalyticsworld.com/dc/2010/agenda.php#day1-8a

50. How Companies Learn Your Secrets, Charles Duhigg, The New York Times, February 16, 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

51. How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did, Kashmir Hill, Forbes, February 16, 2012. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

52. Did Target Really Predict a Teen's Pregnancy? The Inside Story, Gregory Piatetsky, May 7, 2014. http://www.kdnuggets.com/2014/05/target-predict-teen-pregnancy-inside-story.html

53. Apple Q&A on Location Data, April 27, 2011. Apple Computer. http://www.apple.com/pr/library/2011/04/27Apple-Q-A-on-Location-Data.html

54. The one-night stand, quantified and visualized by Uber, Derrick Harris, Gigaom Research, March 26, 2012. https://gigaom.com/2012/03/26/uber-one-night-stands/

55. Uber's Deleted 'Rides of Glory' Blog Post, Who's Driving You?, December 30, 2014. http://www.whosdrivingyou.org/blog/ubers-deleted-rides-of-glory-blog-post

56. Medical Billing Provider and its Former CEO Settle FTC Charges That They Misled Consumers About Collection of Personal Health Data, FTC, December 3, 2014. https://www.ftc.gov/news-events/press-releases/2014/12/medical-billing-provider-its-former-ceo-settle-ftc-charges-they

57. Facebook's 'Year in Review' app swings from merely annoying to tragic, Andrew Peterson, The Washington Post, December 26, 2014. https://www.washingtonpost.com/news/the-switch/wp/2014/12/26/facebooks-year-in-review-app-swings-from-merely-annoying-to-tragic/

58. Inadvertent Algorithmic Cruelty, Eric Meyer, December 24, 2014. http://meyerweb.com/eric/thoughts/2014/12/24/inadvertent-algorithmic-cruelty/

59. My Year Was Tragic. Facebook Ambushed Me With a Painful Reminder, Eric Meyer, Slate. December 29, 2014. http://www.slate.com/blogs/future_tense/2014/12/29/facebook_year_in_review_my_tragic_year_was_the_wrong_fodder_for_facebook.html

60. Facebook Apologizes for its 'Year In Review' Approach, Amit Chowdhry, Forbes, December 29, 2014.

61. Lotus was a major software company that pioneered the development of the spreadsheet and e-mail systems. The company was acquired by IBM for $3.52 billion in 1995.

62. Lotus MarketPlace:Households, Harvard Business School case 9-392-026, June 4, 1996.

63. The Uneasy Ethical and Legal Underpinnings of Large-Scale Genomic Biobanks, Henry T. Greely, Annual Review of Genomics and Human Genetics. 2007. 8:343-64.

64. Weaving Technology and Policy Together to Maintain Confidentiality, Latanya Sweeney, Journal of Law, Medicine and Ethics, Vol. 25 1997, p. 98-110.

65. Ohm, Paul. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization (August 13, 2009). UCLA Law Review, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. Available at SSRN: http://ssrn.com/abstract=1450006.

66. Daniel C. Barth-Jones, The "Re-Identification" of Governor William Weld's Medical Information: A Critical Re-examination of Health Data Identification Risks and Privacy Protections, Then and Now, Pre-Publication Draft-Working Paper, June 18, 2012 Version, http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2076397.

67. Eli Lilly Settles FTC Charges Concerning Security Breach, Federal Trade Commission, January 18, 2002. https://www.ftc.gov/news-events/press-releases/2002/01/eli-lilly-settles-ftc-charges-concerning-security-breach

68. Report to the Public on Events Surrounding jetBlue Data Transfer, Findings and Recommendations, Department of Homeland Security Privacy Office, February 20, 2004. http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_jetblue.pdf

69. Homeland Security—Airline Passenger Risk Assessment, Torch Concepts, February 25, 2003. Archived at https://www.aclu.org/files/FilesPDFs/torch%20concepts%20slideshow%20on%20jetblue%20data.pdf

70. AOL Proudly Releases Massive Amounts of Private Data, Michael Arrington, Tech Crunch, August 6, 2006. http://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/

71. Barbaro M, Zeller Jr. T. A Face Is Exposed for AOL Searcher No. 4417749 New York Times. 9 August, 2006.

72. AOL: This was a screw up, Michael Arrington, Tech Crunch, August 7, 2006. http://techcrunch.com/2006/08/07/aol-this-was-a-screw-up/

73. AOL apologizes for release of user search data, Dawn Kawamoto, CNET, August 9, 2006. http://www.cnet.com/news/aol-apologizes-for-release-of-user-search-data/

74. AOL's $5M Settlement OK'd in Consumer Data Release Suit, David McAfee, Law360, May 28, 2013. https://www.law360.com/articles/445482

75. Checks Mailed for AOL Search Data Class Action Settlement, Dominic Rivera, Top Class Actions, November 20, 2013. https://topclassactions.com/lawsuit-settlements/lawsuit-news/5480-checks-mailed-for-aol-search-data-class-action-settlement/

76. Closing Letter to Reed Freeman, Esq., Counsel for Netflix, Inc., Maneesha Mithal, Associate Director, Division of Privacy & Identity Protection, Federal Trade Commission, March 12, 2010. https://www.ftc.gov/sites/default/files/documents/closing_letters/netflix-inc./

77. Netflix Spilled your Brokeback Mountain Secret, Lawsuit Claims, Ryan Singel, Wired, December 17, 2009. http://www.wired.com/2009/12/netflix-privacy-lawsuit/

78. NetFlix Cancels Recommendation Contest After Privacy Lawsuit, Ryan Singel, Wired, March 12, 2010. https://www.wired.com/2010/03/netflix-cancels-contest/

79. Google's street-level maps raising privacy concerns, Elinor Mills for News.com, USA Today, June 4, 2007. http://usatoday30.usatoday.com/tech/news/internetprivacy/2007-06-01-google-maps-privacy_N.htm

80. Google lets you remove people from Street View, Alex Chitu, Google Operating System, August 26, 2007. http://googlesystem.blogspot.com/2007/08/google-lets-you-remove-people-from.html

81. Google pays another tiny fine in Europe—$1.4M—for street view privacy concerns, Natasha Lomas, April 4, 2014. TechCrunch, http://techcrunch.com/2014/04/04/google-street-view-fine/

82. Jerk, LLC, d/b/a Jerk.com, In the Matter of, Federal Trade Commission, FTC Matter 122 3141, Docket 9361. https://www.ftc.gov/enforcement/cases-proceedings/122-3141/jerk-llc-dba-jerkcom-matter

83. "CVS Caremark Settles FTC Charges: Failed to Protect Medical and Financial Privacy of Customers and Employees; CVS Pharmacy Also Pays $2.25 Million to Settle Allegations of HIPAA Violations," US Federal Trade Commission, February 18, 2009. https://www.ftc.gov/news-events/press-releases/2009/02/cvs-caremark-settles-ftc-chargesfailed-protect-medical-financial

84. "Rite Aid Settles FTC Charges That It Failed to Protect Medical and Financial Privacy of Customers and Employees," US Federal Trade Commission, July 27, 2010. https://www.ftc.gov/news-events/press-releases/2010/07/rite-aid-settles-ftc-charges-it-failed-protect-medical-financial

85. Resolution Agreement: CVS Pays $2.25 Million & Toughens Disposal Practices to Settle HIPAA Privacy Case, US Department of Health & Human Services, January 16, 2009. http://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/CVS/

86. Rite Aid Agrees to Pay $1 Million to Settle HIPAA Privacy Case, US Department of Health & Human Services, July 27, 2010. http://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/rite-aid/index.html

87. "Google, Inc., In the Matter of," FTC Matter 102 3136, https://www.ftc.gov/enforcement/cases-proceedings/102-3136/google-inc-matter. Last updated October 24, 2011.

88. "Snapchat, Inc., In the Matter of," FTC Matter 132 3078, https://www.ftc.gov/enforcement/cases-proceedings/132-3078/snapchat-inc-matter. Last updated December 31, 2014.

89. “God View”: Uber Allegedy Stalked Users for Party-Goers’ Viewing Pleasure,” Kashmir Hill, Forbes. October 3, 2014. http://www.forbes.com/sites/kashmirhill/2014/10/03/god-view-uber-allegedly-stalked-users-for-party-goers-viewing-pleasure/

90. “Is Uber’s rider database a sitting duck for hackers?” Craig Timberg, The Washington Post. December 1,2014. https://www.washingtonpost.com/news/the-switch/wp/2014/12/01/is-ubers-rider-database-a-sitting-duck-for-hackers/

91. “Uber admits employees abused “god view” location tracking,” Jason Mick, Daily Tech, December 18, 2014. http://www.dailytech.com/Uber+Admits+Employees+Abused+God+View+Location+Tracking/article37010.htm

92. “Uber Agrees to Pay $20 Million to Settle FTC Charges That It Recruited Prospective Drivers with Exaggerated Earnings Claims,” Federal Trade Commission, January 19, 2017. https://www.ftc.gov/news-events/press-releases/2017/01/uber-agrees-pay-20-million-settle-ftc-charges-it-recruited

93. “Uber Settles FTC Allegations that It Made Deceptive Privacy and Data Security Claims,” Federal Trade Commission, August 15, 2017. https://www.ftc.gov/news-events/press-releases/2017/08/uber-settles-ftc-allegations-it-made-deceptive-privacy-data

94. Uber Agrees to Expanded Settlement with FTC Related to Privacy, Security Claims, Federal Trade Commission, April 12, 2018. https://www.ftc.gov/news-events/press-releases/2018/04/uber-agrees-expanded-settlement-ftc-related-privacy-security

95. Khanna A. Facebook's Privacy Incident Response: a study of geolocation sharing on Facebook Messenger. Technology Science. 2015081101. August 11, 2015. http://techscience.org/a/2015081101

96. Sweeney L. Only You, Your Doctor, and Many Others May Know. Technology Science. 2015092903. September 29, 2015. http://techscience.org/a/2015092903

97. Craig Brittain, individually, In the Matter of. Agreement containing consent order, United States Federal Trade Commission. https://www.ftc.gov/system/files/documents/cases/150129craigbrittainagree.pdf

98. HealthCare.gov Sends Personal Data to Dozens of Tracking Websites, Cooper Quintin, Electronic Frontier Foundation, January 20, 2015. https://www.eff.org/deeplinks/2015/01/healthcare.gov-sends-personal-data

99. CMS Privacy Notice for HealthCare.gov, May 17, 2018. https://www.healthcare.gov/privacy/

100. Spam King, Simson Garfinkel, Wired Magazine, February 2, 1996. http://www.wired.com/1996/02/spam-king/

101. Messagelabs Intelligence: 2010 Annual Security Report, Symantec, 2010.

102. The Economics of Spam, Justin M. Rao and David H. Reiley, Journal of Economic Perspectives, 26:3, Summer 2012, p. 87-110. http://dx.doi.org/10.1257/jep.26.3.87

103. A 61-million-person experiment in social influence and political mobilization, Robert M. Mond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, Nature 489, 295-298. Sept. 13, 2012.. http://www.nature.com/nature/journal/v489/n7415/full/nature11421.html?WT.ec_id=NATURE-20120913

104. In the Matter of Dialing Services, LLC, Citation and Order: Prerecorded Message Violations, File No. EB-TCD-12-00001812, Before the Federal Communications Commission, Washington, DC, March 15, 2013. https://apps.fcc.gov/edocs_public/attachmatch/DA-13-265A1_Rcd.pdf

105. FCC Enforcement Bureau warns purveyors of Robocalls: “Don’t call us … (or) we’ll call you,” FCC News, March 15, 2013. https://apps.fcc.gov/edocs_public/attachmatch/DOC-319535A1.pdf

106. In the Matter of Dialing Services, LLC: Notice of Apparent Liability and Forfeiture, File No: EB-TCD-12-00001812, Before the Federal Communications Commission, Washington, DC, May 8, 2014. https://apps.fcc.gov/edocs_public/attachmatch/FCC-14-59A1.pdf

107. Robocalls, Federal Trade Commission, Last accessed February 10, 2016. http://www.ftc.gov/robocalls

108. FTC Announces Robocall Challenge Winners, US Federal Trade Commission, April 2, 2013. https://www.ftc.gov/news-events/press-releases/2013/04/ftc-announces-robocall-challenge-winners.

109. How robo-callers outwitted the government and completely wrecked the Do Not Call list, Simon van Zuylen-Wood, The Washington Post Magazine, January 11, 2018. https://www.washingtonpost.com/lifestyle/magazine/how-robo-call-moguls-outwitted-the-government-and-completely-wrecked-the-do-not-call-list/2018/01/09/52c769b6-df7a-11e7-bbd0-9dfb2e37492a_story.html

110. Sprint to pay $7.5 million for unwanted marketing calls and texts in record do-not-call settlement, FCC News, May 19, 2014. https://apps.fcc.gov/edocs_public/attachmatch/DOC-327147A1.pdf

111. "Experimental evidence of massive-scale emotional contagion through social networks," Adam D. I. Kramer, Jamie E. Gullory, and Jeffrey T. Hancok, PNAS, June 17, 2014. 111:24, pp. 8788-8790. doi: 10.1073/pnas.1320040111

112. Controversy Over Facebook Emotional Manipulation Study Grows as Timeline Becomes More Clear, Gregory S. McNeal, Forbes, June 30, 2014. http://www.forbes.com/sites/gregorymcneal/2014/06/30/controversy-over-facebook-emotional-manipulation-study-grows-as-timeline-becomes-more-clear/

113. Editorial Expression of Concern and Correction, Inder M. Verma, PNAS, July 22, 2014. 111:29, p. 10779

114. How the Hell Are These Popular Spying Apps Not Illegal, Kate Knibbs, Gizmodo, January 30, 2015. http://gizmodo.com/how-the-hell-are-these-popular-spying-apps-not-illegal-1682660414

115. Technology tips for domestic violence and stalking victims, Lisa Weintraub Schifferle, Federal Trade Commission. February 5, 2015. https://www.consumer.ftc.gov/blog/technology-tips-domestic-violence-and-stalking-victims

116. Basic HHS Policy for the Protection of Human Research Subjects, Code of Federal Regulations, 45 CFR §46.102 (f) http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html

117. "ULD to website owners: "Deactivate Facebook web analytics", Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, August 19, 2011. https://www.datenschutzzentrum.de/presse/20110819-facebook-en.htm

118. "Private traits and attributes are predictable from digital records of human behavior," Michal Kosinski, David Stillwell, and Thore Graepel, PNAS, 110:15, April 9, 2013. pp. 5802-5805. http://www.pnas.org/content/110/15/5802.full.pdf

---

Authors

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products mentioned are necessarily the best available for the purpose.

This paper is presented with the hope that its contents may be of interest to the general research community. The views in these papers are those of the authors, and do not necessarily represent those of the Census Bureau.

We wish to thank Joan Feigenbaum, Susan Landau, Harry Lewis, Aleecia McDonald, Andrew Odlyzko, Rebecca J. Richards, Claire Stapleton, Latanya Sweeney and Danny Weitzner, all of whom read previous versions of this article and provided useful comments. We also extend our thanks to the anonymous reviewers, whose comments served to tighten and significantly improve this article.

Simson L. Garfinkel is the Senior Computer Scientist for Confidentiality and Data Access at the US Census Bureau and the Chair of the Bureau's Disclosure Review Board. He has published research articles in the areas of computer security, digital forensics and privacy. He is a fellow of the Association for Computing Machinery and holds a PhD in Computer Science from MIT. This article is based on work that Garfinkel performed while he was a Computer Scientist at the National Institute of Standards and Technology (NIST) Information Access Division.

Mary Theofanos is a computer scientist with the U.S. National Institute of Standards and Technology, where she performs research on usability and human factors of systems. Mary is the principal architect of the Usability and Security Program evaluating the human factors and usability of cyber security and biometric systems.

Referring Editor: Latanya Sweeney

Citation

Garfinkel S and Theofanos M. Non-Breach Privacy Events. *Technology Science*. 2018081601. 08 16, 2018. http://techscience.org/a/2018081601