# Big Data and Individual Privacy in the Age of the Internet of Things

Mackenzie Adams

> *"Who could deny that privacy is a jewel? It has always been the mark of privilege, the distinguishing feature of a truly urbane culture. Out of the cave, the tribal teepee, the pueblo, the community fortress, man emerged to build himself a house of his own with a shelter in it for himself and his diversions. Every age has seen it so. The poor might have to huddle together in cities for need's sake, and the frontiersman cling to his neighbors for the sake of protection. But in each civilization, as it advanced, those who could afford it chose the luxury of a withdrawing-place."*
>
> Phyllis McGinley (1905–1978)
> Pulitzer Prize-winning author and poet

The availability of "big data" and "smart" products are credited with advancing solutions to complex problems in medicine, transportation, and education, among others. However, with big data comes big responsibility. The collection, storage, sharing, and analysis of data are far outpacing individual privacy protections, whether technological or legislative. The Internet of Things (IoT), with its promise to create networks of networks, will magnify individual data privacy threats. Recent data breaches, exposing the personal information of millions of users, provide insight into the vulnerability of personal data. Although seemingly expansive, there are core individual privacy issues that are central to current big data breaches and anticipated IoT threats. This article examines both big data and the IoT using examples of data privacy breaches to illustrate the impact of individual data loss. Furthermore, the article examines the complexity of tackling technological and legislative challenges in protecting individual privacy. It concludes by summarizing these issues in terms of the future implications of the IoT and the loss of privacy.

## Introduction

Across most domains, societal functioning has become increasingly dependent on information and communication technology, as well as the management of massive data streaming through physical and virtual environments. The generation of this extensive data, formal or informal in structure, has led to its referral as "big data", a nomenclature pointing to not only sheer size, but also to the speed with which it is generated and the complexity in organizing and analyzing it (Berman, 2013; Chen et al., 2014). Big data has emerged as an area of significant interest in research and applications for organizations dealing with or anticipating an overwhelming flow of data. Individual privacy regarding big data has especially taken hold as a central issue affecting different technology areas as connectivity and information sharing have far outpaced data protection efforts (Perera et al., 2015).

Widely publicized breaches of large databases exposed significant and escalating threats to individual privacy and control over personal data. In 2005, a security breach of an American health insurance company, Anthem, led to the theft of personal information of more than 78 million customers (Mathews, 2015). The information included names, dates of birth, social security numbers, and income data, all of which were likely sold in underground markets. The total number of affected individuals and the sensitive nature of large data

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

breaches are alarming; they also point to an urgent need to the convergence of technology, legislation, user policies, and awareness in protecting privacy.

Big data and individual privacy protection are further complicated by the evolution of networks of networks, also referred to as the Internet of Things (IoT). This new paradigm promises to enable existing and future devices to be connected to local and virtual networks and, eventually, communicate autonomously with these networks and other devices for functions such as gathering and analyzing data (Borgohain et al., 2015). For instance, new applications are enabling users to check the status of their home appliances from their smartphones, monitor private property, and synchronize their devices while increasing the likelihood of exposing the large amount of data collected and stored in these devices and networks to other individuals and entities.

According to Russo and colleagues (2015), by 2020, there will be over 200 billion sensor devices that are interconnected. These sensors will be found in home electronic systems, health monitoring equipment, cars, and smartphones. Their economic impact will also be tremendous, according to the authors who estimate that, by 2025, their market will be worth approximately $3 trillion per year. As the surface area of data expands exponentially through the IoT, the implications of individual privacy threats of this pervasive interconnectivity are immense. Current breaches of large databases and their impact provide insights into how the future of big data and the IoT is shaped. It becomes of significant importance to explore how the collection, storage, sharing, and analysis of big data can be complex and multifaceted and how it can bridge the worlds of technology and application development, privacy legislation, and consumer/user privacy protection processes.

This article examines the implications of compromised individual privacy in the age of the IOT as it relates to big data. First, it provides definitions and descriptions of the widely used terms "big data" and the "IoT". It clarifies the parameters used by researchers in studying and writing about both phenomena, as well as touches upon vulnerability that expose the privacy of individuals' data to unauthorized access, loss, or theft. Next, it examines the extent to which recent big data breaches have exposed the vulnerability of personal data. The examples illustrate the different pathways and impact of individual data loss. Then, the article places issues and challenges of data privacy loss into the context of the age of the IoT, and it emphasizes the fundamental complexity of the IoT and the how it is likely to present further technological, legislative, and user experience challenges to protecting individual privacy. Finally, the article integrates and summarizes the previous sections by examining opportunities in security and individual privacy protection in the age of the IoT.

The underlying assumption of the article is that the collection of data from IoT devices and customization based on the collected data create vulnerabilities in individual data privacy. As a framework to guide the discussion, Figure 1 provides an overview of individual privacy when big data is examined in the age of IoT. Individual privacy is threatened when data is collected and a data breach can expose an individual's private data; it is also threatened when companies and individuals, under the pretext of assumed consent to provide a custom experience, use the collected data. The roles of technological and legislative solutions in protecting individual data privacy continue to change and evolve.

## Big Data, IoT, and Data Privacy

Big data, as a concept, has been around for two decades since being used by Cox and Ellsworth (1997). While initially referring to extensive volumes of scientific data, big data has since been defined in a number of ways. Boyd and Crawford (2012) argue that it "is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets", whereas Hashem and colleagues (2015) propose that big data has three characteristics: i) numerous, ii) cannot be cat-

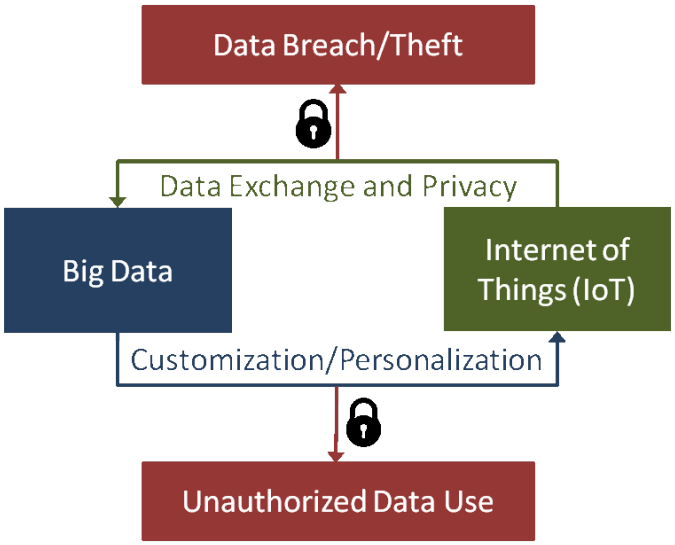


**Figure 1.** A framework for big data and individual privacy

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

egorized into regular relational databases, and iii) generated, captured, and processed rapidly. Chen and Lin (2014), on the other hand, define big data as "the exponential growth and wide availability of digital data that are difficult or even impossible to be managed and analyzed using conventional software tools and technologies".

The most commonly known definition was suggested by IBM (Malik, 2013; Schroeck et al., 2012), which proposes that big data is characterized by any or all of the following three attributes: volume, variety, and velocity. *Volume* reflects the tremendous amounts of data created from a number of sources and across different platforms such as mobile devices and applications and smart grids, as well as social media such as Facebook. The sheer volume of big data is likely to increase substantially as IoT-enabled technology will continue to be designed to generate data from multiple devices and sources. *Variety* refers to the nature of data generated. For instance, structured data from geographic information systems as well as unstructured data from websites are found in numerous formats. *Velocity* reflects the speed with which data is not only generated from a myriad of sources, but the frequency of data capture, analysis, and the application of information in decision making. Hashem and colleagues (2015) have added a fourth "v" to the IBM definition, "value", noting that it is the "most important aspect of big data; it refers to the process of discovering huge hidden values from large datasets with various types and rapid generation". Thus, *value* refers to the actual use of the data collected. Physical devices or sensors may not, by themselves, provide data that can be used for predictive modelling in medicine or retail, for instance. However, multiple devices and sensors can provide data that, when aggregated, provides valuable information upon analysis.

Big data, therefore, is likely about the above four attributes and their scaling to ever greater numbers of devices, infrastructures, and networks. At its core, big data describes the wide availability of data in digital form, with a concomitant presence of data mining and knowledge-generation capability across numerous networks.

*Mining big data*

The collection and storage of large volumes of data has held the promise of data-driven discovery in diverse fields including scientific research, healthcare, industry, manufacturing and education (Chen et al., 2015; Malik, 2014). Massive volumes coupled with wider availability aimed to fulfill this promise through the development of data exploration and mining technologies. The purpose of data mining, therefore, is to uncover useful and novel information from data stored in large databases, thereby being predictive or descriptive. This is an especially important development in fields reliant upon large data for making those predictions to be generalized across populations such as medicine and commerce. The data mining process, in general, involves several major steps whereby data is cleaned, transformed, and mined for information.

Big data and the use of machine learning algorithms have become inextricably linked with data mining recently. A main reason is that datasets have grown larger and more complex, and traditional learning methods of managing such volumes while extracting useful data have fallen short. Furthermore, while the volume of data has increased, its quality has remained inconsistent; data mining efforts face low quality, multi-form data across numerous applications and systems, and are further complicated by the lack of effective security solutions to share such data. As noted by Shukla (2015):

*"I use the term big data a bit too generically to include machine learning and data mining even when the data is not necessarily 'big'. Especially when the Internet of Things becomes a reality in improving the lives of people, improving quality of automation systems, and improving transportation system performance, machine learning and data mining will be ready to deliver technologies, algorithms, and possibly products that can be directly used to make those systems perform in the most optimal fashion, adapting to changing situations, and securing the system against hackers who would certainly want to disrupt such systems or try to breach privacy of people who will be connected to such networks."*

Data mining for effective decision making may seem innocuous from the perspective of private data exposure. Aggregate forms of data, such as those collected by search engine programs or presented in census information, are expected to remove key pieces of identifying information while retaining others for the purpose of analysis (Boyd & Crawford, 2012; Liu, 2014). For instance, census data collected may aggregate ages to arrive at descriptive statistics for age groups, but will be expected to not provide access individual identifying information such as names and addresses. However, these expectations are outside the control of individuals whose data may be stored, transferred, shared, and analyzed by different individuals and organizations. As both data volume and data mining interest increase in the IoT paradigm, the issue of privacy becomes more urgent.

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

*Internet of Things (IoT): Current paradigm and anticipated reality*

From a review of recent literature, it is apparent that the IoT encompasses an understanding of how networks of networks will connect devices, infrastructure, and systems, among others, through a new Internet. The review shows that the IoT is referred to by researchers and practitioners as "a vision", "a new paradigm", "an area of research", "an emerging global Internet based information architecture", "next step evolution of our today Internet", "a growing technology", and "a new form of computation". Perera and colleagues. (2015) define the IoT as a "network of networks, in which, typically, a massive number of objects, things, sensors, devices are connected through the information and communications infrastructure to provide value-added services".

A comprehensive definition of the IoT is also presented by Russo and colleagues (2015), who state that:

*"The Internet of Things (IoT) is an integrated part of the Future Internet and can be defined as a dynamic global network infrastructure with self-configuring capabilities based on standard and inter-operable communication protocols, where physical and virtual things have identities, physical attributes and virtual personalities; they use intelligent interfaces and are seamlessly integrated into the information network."*

The IoT promises unprecedented advancements across knowledge-based industries and fields. According to a review of literature on the IoT (Russo et al., 2015) from its earliest conceptions in the late 1980s to 2015, numerous future characteristics behind these advancements are proposed by researchers as follows:

- Evolution in communication, not only human–human and human–things, but things–things as well, reflecting an increasing role of autonomous communication among devices and artificial intelligence research and application.
- Optimization of energy consumption through network infrastructures and remotely monitored systems designed to reduce consumption. Smart homes are an example whereby devices can be programmed to autonomously communicate and can affect such things as temperature settings and electricity consumption.
- Wider opportunities to develop technologies and tools through the creation of Internet-connected devices.
- Greater role in development of technologies in medicine, critical infrastructures, and smart cities. Recent advances in continuous patient monitoring, including in-hospital and out-of-hospital applications are strong examples of such technologies.

On a wider, societal scale, IoT applications are numerous and wide-ranging given that they are used in commercial, environmental, and critical infrastructure settings (Chen et al., 2014). It is expected that, with an increased capability in analyzing large data, high-quality information will guide such functions as monitoring air quality and pollution indices, as well as monitoring food as it is transported across the globe. The agricultural industry can exploit in-ground sensors and irrigation-control software to automate its soil management, while reducing costs associated with inclement conditions (Russo et al., 2015). Commercial applications have noted the ever-increasing role of supply chain management and logistics, both of which are made more efficient and cost-effective when connected devices are programmed to provide basic decision-making capability.

In summary, the IoT will allow billions of objects, such as mobile devices, and virtual environments to exchange data. With machine learning, devices and environments may exchange such data autonomously while extracting meaningful data. However, the IoT – by definition – is complex and covers extensive data landscapes, structures, and contexts. This complexity has serious implications in securing information flowing from individuals' devices to the networks of the IoT. To further complicate the exposure of private data, cloud computing environments essentially upload the 'minute details of one's life to virtual environments that are targets for privacy breaches (Maras, 2015; Matzner, 2014; Perera et al., 2015). The IoT is a developing target for interconnectivity of devices and environments in a network of networks. The potential entry points and vulnerabilities to data privacy breaches are also developing, and a key question is whether security measures can be concomitantly interoperable and scalable. However, breaches of large datasets are a reality, and recent years have shown how vulnerable individual data is to loss of control, theft, and exploitation.

## Privacy Loss and Big Data Breaches

Privacy of individual data is expectedly complex and multi-faceted, extending across technological, legal, commercial, and financial domains (Punagin & Arya, 2015). The loss of personal information to unauthorized

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

and illegal means did not start with the Internet; individuals were likely to lose their financial information such as credit card statements or social insurance numbers from thieves rummaging through personal effects or property. The widespread digitization of everyday living, from financial transactions to personal communication, to business dealings, however, has exposed individual information to unauthorized access to entities from across the globe (Bekara, 2014). In the process, it has prompted a revisiting of privacy threats and an examination of individual privacy and control of data generated by our activities as a right deserving of user and legal protections. It remains that the right to the massive data collected currently through databases – which are expected to be interconnected, sometimes autonomously, through the IoT – has legal frameworks and privacy-enhancing technologies but they are lagging to provide adequate protections (Han et al., 2014; Maras, 2015).

Some examples of big data collection may seem mundane. Currently, most smartphones are enabled with location sensors, providing real-time data to be collected on an individual's whereabouts and activities (Rghioui, et al., 2015). As more devices are enabled to provide similar information, we observe that cars also provide data on location, while household efficiency and security protection are connected to handheld devices. Taken together, the information from disparate devices provide extensive information on individual and behavioural patterns, which is a privacy concern (Schroeck et al., 2012; van de Pas & van Bussel, 2015). This situation is similar to the collection of browsing history and purchasing behaviour used to tailor online activities to an individual. However, they are also similar in exposing individuals to the loss of their information.

Big data collection and mining are also promising to transform the quality of individuals' lives in innumerable ways. In healthcare, for instance, health information collection is now enabled in many everyday devices such as iPhones or FitBits, providing continuous data collection of key health behaviour, a function reserved in the past through medical intervention to a limited number of people (Suciu et al. 2015; Tsai et al., 2014). Abinaya, Kumar, and Swathika (2015) examined the application of the IoT in devising an information system based on the ontology method. The researchers explored a system that aimed to connect emergency medical services with hospital-based services.

The implications of this data collection and storage, and the ability to provide real-time analysis and provision to healthcare providers, represent a revolutionary advancement in health monitoring and preventive care (Abinaya et al., 2015). With an increase in big data analytics and technology, the large, raw health data collected from these and other devices can provide valuable information about the individual's health, as well as population-level information that previously would have only been available through formal, large studies. Once again, however, privacy risks are inherent in the collection, storage, and exchange of this data. Individuals may lose control of who views their information, which has the potential to result in exposure of health conditions and practices, but may also have ramifications for employment and health insurance (Borgohain et al., 2015; Krotoszynski, 2015).

High profile data breaches, especially of businesses, often dominate media coverage of data security compromises because they often involve the information of numerous clients and customers. A data breach is said to have occurred when individuals' data has been subjected to unauthorized access, resulting in the exposure of confidential, protected, or sensitive information. The personal, financial, and legal impact of data breaches can be tremendous (Sen & Borle, 2015). Individuals whose information is stolen or accessed can suffer identity and financial losses, and have sensitive information such as health conditions or personal behaviour scrutinized and exposed. Organizations that are breached are also likely to suffer financial and proprietary information losses, as well as reputation compromises. Organizations that collect extensive personal data from their customers, such as healthcare institutions and banks, are particularly vulnerable to such losses. According to the Ponemon Institute Report (2014), the impact of data breaches' on individuals, mostly linked to identity thefts, are implicated in a loss of $16 billion approximately from nearly 13 million individuals. The average cost per incident was estimated to be nearly $6 million for organizations in the United States. The report also cites that identity theft is the dominant consumer fraud complaint to the United States Federal Trade Commission (FTC).

*Financial and private institutions*

A number of illustrative cases of big data breaches in recent years have shed light on the nature and impact of individual data security compromises. In the previously mentioned Anthem health insurance company breach, approximately 78 million people had their company private records illegally accessed (Mathews, 2015). Breached data included their identifying numbers along with names, dates of birth, and social security

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

numbers, as well as income data. The nature of the data stolen reflects the high risks of identity-theft schemes. The hacking of a large database of JPMorgan Chase bank affected a similar number of individuals, approximately 76 million (Silver-Greenberg et al., 2014). The database hackers gained access to applications that were run on the bank's computers where they were able to exploit a known vulnerability. The hackers were able to access personal details such as names, addresses, and phone numbers, although the company had released a statement that other personal data such as dates of birth were not included in the hacked databases.

Given the wide use of social media and networking sites, it was inevitable that a large data breach would occur. The Canadian owned social dating site, Ashley Madison, was hacked in 2015, exposing the company's internal servers, company bank account data, and staff salary information (Solomon, 2015).

*Public institution breaches*

Although they are attractive targets for big data breaches, financial institutions are not the only organizations that are targeted for malicious access. Approximately 191 million American voters' personal information was exposed on the open Internet due to an incorrectly configured database (Finkle & Volz, 2015). While not considered a malicious act, it is, nonetheless, a data breach that exposed the personal details such as name and address, as well as party affiliations of voters in all 50 States and Washington, DC. Another governmental body exposed the individual private data of millions of American military veterans when a breach occurred at the National Archives and Records Administration (Singel, 2009). The breach was traced back to a defective hard drive that the organization had sent to the external vendor for repair. However, it was later discovered that the data recorded in the drive was not destroyed before being sent to the vendor.

Those seeking illegal access to data are, at times, motivated by nation-state purposes. An example of a public institution breach through such a purpose is the hacking of the Office of Personnel Management (OPM) in the United States by the Chinese state (Nakashima, 2015). The organization informed approximately 4 million current and former federal employees that their personal data had been accessed illegally. Representing the biggest data breach of federal employees in recent history, the OPM breach exposed personal identifying information such as social security numbers, human resources' related information, and job assignments.

Critical infrastructures are also a target for big data breaches. The San Francisco Public Utilities Commission warned approximately 180,000 thousand of its customers that a data breach had exposed their personal information to illegal access (Mills, 2011). Specifically, customers' account numbers and personal identifying information such as names and addresses were breached. According to the organization, the breach occurred when an unsecured server was infected with viruses through an open port.

*Individual data loss impact and protection*

The above-illustrated cases of big data breaches provide insight into both the vulnerability of personal data and the impact of its loss. Organizations must work to secure personal data by ensuring that only information that is required is collected from users and customers. Ensuring that only required information is collected will force both individuals and organizations to realize that data has to be protected, and the less personal/sensitive data collected, the less likely that it is breached (Maras, 2015). Furthermore, to safeguard personal information, it is crucial that storage and transportation processes are embedded with security measures. The above examples of inadvertent data breaches show that carelessness, poor follow-through, and lack of accountability can be just as harmful as intentional hacking or malicious behaviour.

Organizations should also consider effective and periodic ways to discard personal information collected from individuals, especially when that information is no longer required in its raw forms. To reduce the risk of unused servers becoming the target of data loss, users and organizations should be diligent in pursuing strict and accountable processes for discarding data. As explored in this article, there are important implications of inconsistent data management and handling processes that will surely be magnified in IoT environments (Maras, 2015; Samani et al., 2015). When the absolute volume of data exchanged increases exponentially in such environments, even the most diligent of systems can "lose track" of personal information, especially as data is streamed from new devices and objects.

It is also important to consider that the public–private sphere of policies and protections are at times blurred in the context of data exchanges (Schroeck et al., 2012; van de Pas & van Bussel, 2015). For example, policies to limit data collection in public institutions may not exist in private organizations. Governments are likely limited in how they can impose data protection measures in

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

the "private sphere". Standardization of measures, even within public institutions, is a challenge due to a potential impact of data exchange limits. Potential employees or health insurance seekers are expected to provide their personal information. If they do not, they may not be insured or considered for employment. The same can be said for everyday aspects of life including securing loans, buying or renting places of residence, and even enrolling in colleges and universities. Thus, individuals in society cannot opt out of disclosing their personal information to private and public entities, but such disclosure comes with the risk that their private information may be exposed in a data breach.

## Individual Privacy Issues and Challenges in the IoT

As more IoT-enabled devices and systems are created, more individual data privacy issues and challenges emerge, especially as big data analytics and technology are positioned to search for value in this data. It is generally insightful to examine "on-the-ground" applications of the IoT against emerging privacy concerns. For instance, Rghioui and colleagues (2015) examined the lack of consideration of data security and privacy in the IoT-based wireless body area network (WBAN). Specifically, the researchers reviewed various devices that are now attached to patients physically to monitor health outputs such as cardiac function. These devices have allowed patients to become more mobile while continuously monitored by their healthcare providers and transmitting data through the WBAN. Rghioui and colleagues (2015), however, found that, despite the tremendous advancement in health monitoring offered by these devices, the WBAN networks were largely open to outside access with external IP hosts, which could compromise data integrity, disrupt communication between the mobile devices and the networks, and expose personal health information to unauthorized individuals.

Rghioui and colleagues (2015) proposed a number of solutions for the management of security keys through encryption, which would consider patient mobility and a device's resource constraints. The solutions they proposed address a number of important factors in addressing IoT data-privacy issues, namely, data integrity, scalability, mobility, and key connectivity. Data integrity is an especially important factor whereby encryption keys ensure that no unauthorized access occurs in the transfer of information among devices and the networks. Scalability is also important given that a key challenge in security measures in the IoT is whether a network can remain stable as more devices are added to it. Although their proposed solutions in managing privacy concerns in a healthcare setting are technologically focused, their paper sheds light on overarching issues in securing the integrity and access to large volumes of data in an IoT environment, while continuing to scale up the technology to serve more patients in greater health monitoring functions.

In addition to healthcare, smart grids are an area of exploring big data privacy issues and challenges in the IoT. Bekara (2014) examined security as a determining factor in the expanded application of the IoT and smart grids. iIn a number of IoT-based smart infrastructure contexts, such as homes, cars, and appliances, inherent data privacy and security issues include: impersonation/identity spoofing; eavesdropping; data tampering; authorization and control access issues; privacy issues; compromising and malicious code; and availability and denial-of-service issues and cyber-attacks. Thus, individual data privacy in an IoT-based smart grid is largely compromised through exposure of personal information to unauthorized access, especially in the context of device-device and device-network communication. Similarly, Bekara (2014) has highlighted privacy and security challenges related to scalability; mobility; deployment over large areas; legacy systems; constrained resources; heterogeneity in implemented protocols and communication stacks; interoperability; bootstrapping; trust management; and latency or time constraints.

Other researchers examining IoT-enabled technologies, especially those affecting individuals and households, note similar data privacy and security challenges. Punagin and Arya (2015) also explore the various opportunities presented through IoT-based technologies such as healthcare, mobility, smart grids, law enforcement, and e-commerce. The researchers note a number of similar privacy and security challenges as well, such as identity/sensitive attribute disclosure. As noted earlier, it is expected that big data is an aggregate of individual data, and that various methods of de-anonymizing individual-level data will be available. However, published data may be susceptible to external linkage attacks where hackers and other attackers can link the publicly available data to the de-anonymized one. Narayanan and Shmatikov (2008) were able to de-anonymize a Netflix data set, linking it to individual user profile data from an entertainment repository website, while Sweeney (2002) de-anonymized a hospital's anonymized health records by linking the data set with publicly available information.

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

Punagin and Arya (2015) also include automated recommendations. Big data may expose a person's behavioural patterns (websites visited, pages clicked, etc.) on their social networking site. The automated recommendation may be a data breach if the person has not provided consent. Finally, the researchers list predictive analysis as a security challenge. According to the authors, retailers could use big data analytics to conduct regression analysis on individual purchase habits and patterns, and use them to make predictions about future behaviour. Although this approach may be used widely on a population level, at an individual level, it is a data privacy compromise, especially when consent is absent.

Hashem and colleagues (2015) examined data privacy and security challenges in cloud computing applications. Cloud computing refers to distributed data-processing platforms, and it is one of the building blocks of IoT-based technologies. The authors note, "big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast-growing cloud-based applications developed using virtualized technologies" (Hashem et al., 2015). Given this, privacy and security challenges include big data mining and analytics, which access personal data to create information such as location-based services and recommendations. The authors argue that this use of individual data exposes individual privacy to profiling, loss of control, and theft. The authors further note that control over individual data falls under rules of transparency and accountability that exist between users and organizations, and these rules must be clarified in cloud computing given the high chance of individual privacy compromise.

In their analysis of individual data privacy in the era of the IoT, Perera and colleagues (2015) focus on the inherent assumptions and understandings of which users must be aware when connecting to the Internet with their devices. For instance, the authors note that, when individuals use free online services, such as Facebook and email, they must be aware that they are signing on to become sources of business data. This data is likely used by the service owners to improve services; however, it may also be used to conduct predictive analyses or may be given to affiliate businesses and organizations. Consent may or may not be sought for these actions. Perera 'and colleagues (2015) predict that consumers may find themselves weighing the "free" aspect of online services against their privacy protection in connecting to IoT-enabled technologies. This is especially the case with these technologies continuing to gather more intimate personal information such as health metrics and daily living behaviours. If not paying outright for privacy protections, individuals may opt to limit how their data is used in exchange for continuing to use free services.

## Individual Data Privacy Protection in the IoT

Earlier, we noted that big data and IoT-enabled technologies have outpaced the development of legal and user-privacy protection frameworks'. Weber (2015) argues that today's IoT devices are designed to minimize the likelihood that data transmitted across devices and networks will be at risk for tampering and interception. However, he notes that existing protocols and compression technologies for the movement of large volumes of data are limited. Furthermore, the technological limitation is coupled with legislative ones that have not caught up with fast advancements in the field. There is little argument that privacy is considered a right, and individual user protections are necessary to safeguard this right (Maras, 2015). Legal data-protection laws and privacy laws are limited, however, by the type of data created, collected, transmitted, and exchanged. For instance, the European Data Protection Directive (DPD) legislates data if it is deemed private (Weber, 2015).

*Privacy definition and legislation*

The definition of privacy is understandably diverse and broad. In 1968, Westin defined "information privacy" as "the right to select what personal information about me is known to what people". The definition is dated but has a core value of "right" in controlling individual information disclosed to others – a value that is significant even in the era of the IoT. Ziegoldorf, Morchon, and Wehrle (2014) proposed an IoT-relevant definition of privacy that is reflective of current technological innovation and data exchange. The authors' definition of privacy in the IoT is the "guarantee to the subject for 1) awareness of privacy risks imposed by smart things and services surrounding the data subject; 2) individual control over the collection and processing of personal information by the surrounding smart things; and 3) awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere" (Ziegoldorf et al., 2014).

Privacy legislation aims to provide a balancing force against business and commercial enterprises' ever-increasing chase of data that services market and advertising needs. With appropriate legislation, individual

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

privacy protections place the values of personal information control and use as prime values in this balancing act. The 1948 Universal Declaration of Human Rights recognized privacy as a fundamental human right, while most countries' constitutional rights include privacy (Ziegoldorf et al., 2015). The United States passed the first known legislation on information privacy more than 40 years ago through the 1974 US Privacy Act, whereby fair information practices (FIPs) were established. The FIPs were developed to hold a number of core values regarding individual information including "the principles of notice, consent, individual access and control, data minimization, purposeful use, adequate security, and accountability" (Ziegoldorf et al., 2015).

Regardless of the core values and principles underlying current existing privacy legislations, there are fundamental challenges in the era of the IoT, as pointed out by several researchers (Krotoszynski, 2015; Maras, 2015; van de Pas & van Bussel, 2015; Ziegoldorf et al., 2015). One important challenge is the definition of "personal" in a number of concepts such as "personally identifiable information". Attributes such as date of birth and financial information as identifying attributes in definitions may vary by legislation or jurisdiction. This variability makes it a challenge to have a single privacy definition that could apply across different technologies and applications in the IoT that are developed and managed by different entities.

A second challenge identified by researchers is how legal frameworks and legislations lag behind applications going live and being used by millions worldwide. Ziegoldorf and colleagues (2015) note that the European Commission passed a law against the tracking of web users in 2011; this legislation comes nearly 20 years after users starting browsing the web. In a similar vein, IoT-enabled technology is developing at a much faster rate than legislation could and should. It remains that many jurisdictions have not legislated the sale of user data on websites that offer their services free, such as email. Thus, users are likely to receive promotional and other marketing information once they have registered to use a free site. With IoT technologies, Ziegoldorf and colleagues (2015) argue that it is unclear whether "personal" information in the future will include readouts from health monitoring devices or home smart meter readings.

A third challenge for privacy legislation in the IoT is unique to the paradigm: the speed with which data is exchanged and the volume of data involved both make it unlikely that data privacy breaches will even be known to individuals. Unlike previous data breaches that could be linked often to financial fraud or identity theft directly, and thus individuals were made aware of them through their credit reports and financial statements, the loss of personal information from multiple devices is more insidious. One can lose data privacy aimed to individualize advertising without a physical loss of assets or exposure of private data in a public platform. For example, output from a medical device could be used by others to tune their advertising, but it still reflects loss of personal information.

*Privacy protecting solutions in the IoT*

Protecting individual data privacy in the IoT will bridge legislative and technological solutions, in addition to addressing social, cultural, and political factors. The purpose behind any data privacy protection solution will be compliance; however, there are a number of challenges that impede such compliance. If system development does not integrate sufficient privacy-protecting capabilities, expanding them upon and beyond deployment is often costly, unwieldy, or not possible (van de Pas & van Bussel, 2015). Similarly, when protection solutions include policy and user documentation that are vague in language, inadequate in scope, and non-enforceable across applications and systems in an IoT environment, compliance is also affected. Spiekermann and Cranor (2009) provide a framework whereby privacy can be protected through two major routes: privacy-by-architecture and privacy-by-policy.

Privacy-by-architecture aims to incorporate privacy-preserving functionalities into the earliest stages of system development. For instance, while gathering system requirements, the engineers and developers will aim to build capabilities that minimize the collection of personal data or provide anonymization functionality during the information lifecycle. Privacy-enhancing technologies use this process in their development. Privacy-by-policy, on the other hand, holds "notice and choice" as a central value in developing privacy-protecting policies. Spiekermann and Cranor (2009) note that, despite the expedience of this approach, it has multiple issues that fall short of providing effective protection of individual data. Specifically, organizations such as companies, service providers, and data-collecting governmental bodies can readily draft privacy policies that maximize their access to individuals' personal information while writing the protection components in vague language that is difficult to understand (Maras, 2015). When these entities incorporate language that is designed to provide defense against future lawsuits, privacy-protecting policies become even more

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

incomprehensible to the average user (Krotoszynski, 2015; Samani et al., 2015). For privacy-by-policy to provide effective protection solutions to individual data in the era of IoT, it will address these challenges through user-controlled language and parameters, as we will note shortly.

*Privacy-by-architecture*

Currently, there are a number of privacy-enabling technologies that are deployed to provide some protections. Suciu and colleagues (2015) looked at how to secure e-health architecture through a search-based application, CloudView. Specifically, they noted how cloud middleware received data from heterogeneous devices and integrated data from healthcare platforms, which at times compromised the security of user information. Their proposed search-based application protects this user information by ensuring that data is stored and processed as close as possible, in both space and time, to its location of creation and consumption. The researchers also supported non-functional requirements in the solution such as reliability and security through well-designed integration of physical resources and remote devices, thus "things" and gateways. Finally, the application ensured the distribution of on-the-spot inferred content, instead of raw data. This quality of the solution reflected resource efficiency and scalability of the system so that more IoT-enabled devices and objects can be added.

In their overview of security protecting solutions in cloud applications, Hashem and colleagues (2015) note a number of approaches, including the development of a reconstruction algorithm for privacy-preserving data mining. They also note that a privacy-preserving layer can be applied over a MapReduce framework to reduce risks to privacy caused by data indexing. The privacy-preserving layer makes certain that data privacy is preserved before it is further processed, while ensuring that other data processing applications can be integrated. Given that many privacy-protection solutions are resource intensive and, thus, cannot be scaled in IoT environments, the researchers propose a solution that is an "upper bound privacy leakage constraint-based" approach. To make encryption of data feasible in cloud computing, the solution helps identify which intermediate datasets should be encrypted rather than encrypting all. The benefit is that protection of data can be effective without incurring the cost and time of encrypting all datasets in various states of cleaning, transformation and analysis.

Henze and colleagues (2016) also provide privacy-protection solutions for cloud-based IoT technologies and applications. The authors do so by allowing users to enforce their privacy requirements before their sensitive data is uploaded to the cloud. The solution also enables developers of cloud services' to integrate this privacy functionality into existing IoT-enabled devices. The core requirements of a system that integrates the IoT and cloud computing in privacy-critical application areas are as follows (Henze et al., 2016):

1. *Data security* ensures that data access is controllable by the owner of the data. Security design and mechanisms have to be robust and flexible enough to allow owners to change their mind about access in the future.

2. *Transparency by design,* on the other hand, and as recommended by van de Pas and van Bussel (2015), ensures that data-usage documentation is incorporated into the design and implementation of a cloud service so that users have transparency regarding how their information will be accessed and by whom.

3. Similarly, *privacy-aware development* ensures efficiency in enhancing privacy-protection capabilities by supporting these functionalities early in the development process.

4. *User-controlled* data use and handling shift the control of data access and use to the individual end user rather than the developer or service provider.

5. *Adaptable user control* allows for differential expertise in these end users to tailor data access and use control to their needs in the future.

*Privacy-by-policy*

Protecting individual data privacy through policy is common practice that is, similar to legislation, likely to fall short in IoT contexts. Both the lack of clarity in language and poor classification of "private" or "personal" information across applications and systems are important factors. However, there are steps towards privacy protection solutions through policy that address these factors. Lu and colleagues (2015) propose an attribute-based privacy information security classification (PISC) model that classifies information into categories based on the degree of security and privacy. Each classification is designed to have a security goal that determines the nature of encryption, access control, and a time limit for access.

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

Punagin and Arya (2015) argue that privacy protection can often come at the expense of utility and access to online services, with a resultant restriction in the information provided as more security measures are implemented. They propose that users should be in control of how much of their private information they are willing to share with others, at the risk of exposure, to achieve better utility. Thus, "data collection and usage mining becomes transparent and users understand what data is being mined and how it is used, they may be willing to share their personal information with increased confidence" (Punagin & Arya, 2015).

Other data privacy researchers propose that policy-carrying data is an effective solution for incorporating user control into the development of data-protection policies. Padgeta and Vasconcelosb (2015) acknowledge that the "who", "when", and "how" concerns of data access must be captured in policies to protect data privacy. They propose that a way to capture the wording and manner of access controls over data, and the ability to link that with clarity with the data through what they term "policy-carrying data" (PCD). According to the authors, the PCD sets parameters for the transmission, storage, use, and disposal permissions. The formalized process would provide very specific instructions to how pieces of data can be used and by whom. The following is an example of a PCD proposed by the authors:

*"Lab managers can access 500 records of my data. If an interested party requested 1,000 records, the server would (i) check the credentials of the requester (who needs to be registered); (ii) grant access to 500 records (a message would provide reasons for not providing the 1,000 records); (iii) update the record of that requester with respect to that PCD. Further requests from the same party would be rejected with a suitable justification."* (Padgeta & Vasconcelosb, 2015)

There are several qualities of this PCD that address privacy policy challenges presented above. One quality is the specificity of the data use and control. There is an upper limit, with a provision on how to handle more requests for data. The PCD also has clear language that is controlled by the data owner. Rather than vague, often standard, language about the use of data, it provides clear parameters and consequences for requests beyond those parameters. More importantly, it places transparency as a core factor in communicating data access and control wishes.

Saroiu, Wolman, and Agarwal (2015) also propose the use of PCD to provide individual data-privacy protections in cloud-based applications. The authors argue that, instead of expensive and difficult-to-implement technological solutions, individuals should use a simpler approach before uploading to a cloud environment any data they deem private. Their form of PCD, as a terms-of-service document similar to the one used by sites and service providers already, will allow data owners to be the ones to dictate how their data will be used. The main purpose of the PCD proposed is to bind the user's data to the policy parameters and conditions of use. Therefore, an individual can be explicit in opting out of (or into) some data uses or in setting time/volume limits as proposed by others.

It is interesting to note that the proposal by Saroiu and colleagues (2015) uses encryption in a novel way. It compels the cloud services' providers to be compliant with the PCD that the data owner attaches to the data. It does so by using ciphertext-based attribute-based encryption (CPABE). Following their reading of the PCD, the service providers must build a number of attributes that are compliant with the policy parameters and conditions. If the attributes are not compliant, the decryption fails and the data is not available in the environment. Similar to the Padgeta and Vasconcelosb (2015) approach, this PCD places data owner control as a core value in creating a policy-driven data protection solution.

## Conclusion

Protecting personal data in the era of big data and the IoT requires a multi-faceted approach that places data owner control as a core value of its solutions. Individuals must be not only aware of the data they generate and share across devices and platforms, but they must also understand the security risks and implications of a breach. Whether technology or policy, or a combination, is used to protect individual data, it must be done with users controlling who accesses their information and in what manner. And, importantly, data owners should not be penalized for accessing the advantages of an increasingly connected, data-rich world of information and communication technology with an increased risk of privacy loss and exploitation.

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

## About the Author

**Mackenzie Adams** is Co-Founder and Creative Director at SOMANDA Inc., and she is a recent graduate of the Technology Innovation Management (TIM) program at Carleton University in Ottawa, Canada. As an avid learner and serial entrepreneur, Mackenzie is always seeking new challenges to continue evolving and expanding her interests, knowledge base, and skills. Her interests span the fields of artificial intelligence, quantum computing, EdTech, and FinTech. Her passion is to find and cultivate the next generation of innovators in underserved communities.

## References

Abinaya, V., Kumar, V., & Swathika, K. 2015. Ontology Based Public Healthcare System in Internet of Things (IoT). *Procedia Computer Science,* 50: 99–102. https://doi.org/10.1016/j.procs.2015.04.067

Bekara, S. 2014. Security Issues and Challenges for the IoT-Based Smart Grid. *Procedia Computer Science,* 34: 532–537. https://doi.org/10.1016/j.procs.2014.07.064

Berman, J. J. 2013. Introduction. In *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information:* xix–xxvi. Boston: Morgan Kaufmann.

Borgohain, T., Kuman, U., & Sanyal, S. 2015. Survey of Security and Privacy Issues of Internet of Things. *International Journal of Advanced Networking and Applications,* 9(11): 20–26.

Boyd, D., & Crawford, K. 2012. Critical Questions for Big Data. *Information, Communication and Society,* 15(5): 662–679. http://dx.doi.org/10.1080/1369118X.2012.678878

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V., & Rong, X. 2015. Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks,* 11(8). https://doi.org/10.1155/2015/431047

Chen, M., Mao, S., Zhang, Y., & Leung, V. 2014. *Big Data: Related Technologies, Challenges, and Future Prospects.* Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-06245-7

Chen, X-W., & Lin, X. 2014. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access,* 2: 514–525. http://dx.doi.org/10.1109/ACCESS.2014.2325029

Cox, M., & Ellsworth, D. 1997. *Managing Big Data for Scientific Visualization.* ACM SIGGRAPH '97, August 1997.

Finkle, J., & Volz, D. 2015. Database of 191 Million U.S. Voters Exposed on Internet: Researcher. *Reuters,* December 29, 2015. Accessed April 10, 2017: http://uk.reuters.com/article/us-usa-voters-breach-idUKKBN0UB1E020151229

Gol, H. S. 2016. Integration of Wireless Sensor Network (WSN) and Internet of Things (IOT): Investigation of Its Security Challenges and Risks. *International Journal of Advanced Research in Computer Science and Software Engineering,* 6(1): 37–40.

Guo, B., Zhang, D., Wang, Z., Yu, Z., & Zhou, X. 2012. Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things. *Journal of Network and Computer Applications,* 36(6): 1531–1539. https://doi.org/10.1016/j.jnca.2012.12.028

Han, G., Chan, S., Shu, L., & Hu, J. 2014. Security and Privacy in Internet of Things: Methods, Architectures, and Solutions. *Security and Communication Networks,* 7(11): 2181–2181. http://dx.doi.org/10.1002/sec.1065

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. 2015. The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues. *Information Systems,* 47: 98–115. https://doi.org/10.1016/j.is.2014.07.006

Henze, M., Hermerschmidt, L., Kerpen, D., Häußling, R., Rumpe, B., & Wehrle, K. 2016. A Comprehensive Approach to Privacy in the Cloud-Based Internet of Things. *Future Generation Computer Systems,* 56: 701–718. https://doi.org/10.1016/j.future.2015.09.016

Kelly, G. 2014. eBay Suffers Massive Security Breach, All Users Must Change their Passwords. *Forbes,* May 21, 2014. Accessed April 10, 2017: http://www.forbes.com/sites/gordonkelly/2014/05/21/ebay-suffers-massive-security-breach-all-users-must-their-change-passwords/

Krotoszynski, R. J. Jr. 2015. Reconciling Privacy and Speech in the Era of Big Data: A Comparative Legal Analysis. *William and Mary Law Review,* 56(4): 1309.

Liu, C. 2014. External Integrity Verification for Outsourced Big Data in Cloud and IoT: A Big Picture. *Future Generation Computer Systems,* 49: 58–67. http://dx.doi.org/10.1016/j.future.2014.08.007

Lu, X., Qu, Z., Li, Q., & Hui, P. 2015. Privacy Information Security Classification for Internet of Things Based on Internet Data. *International Journal of Distributed Sensor Networks,* 11(8). http://dx.doi.org/10.1155/2015/932941

Malik, P. 2013. Governing Big Data: Principles and Practices. *IBM Journal of Research and Development,* 57(3/4): 1–13. https://doi.org/10.1147/JRD.2013.2241359

Maras, M.-H. 2015. Internet of Things: Security and Privacy Implications. *International Data Privacy Law,* 5(2): 99–104. https://doi.org/10.1093/idpl/ipv004

Mathews, A.W. 2015. Anthem: Hacked Database Included 78.8 Million People. *The Wall Street Journal,* February 24, 2015. Accessed April 10, 2017: http://www.wsj.com/articles/anthem-hacked-database-included-78-8-million-people-1424807364

Matzner, T. 2014. Why Privacy Is Not Enough Privacy in the Context of "Ubiquitous Computing" and "Big Data". *Journal of Information, Communication & Ethics in Society,* 12(2): 93–106. http://dx.doi.org/10.1108/JICES-08-2013-0030

Mills, E. 2011. SF Utilities Agency Warns of Potential Breach. *CNET,* June 2, 2011. Accessed April 10, 2017: http://www.cnet.com/news/sf-utilities-agency-warns-of-potential-breach/

# Big Data and Individual Privacy in the Age of the Internet of Things

*Mackenzie Adams*

Nakashima, E. 2015. Chinese Breach Data of 4 Million Federal Workers. *The Washington Post,* June 4, 2015. Accessed April 10, 2017: https://www.washingtonpost.com/world/national-security/chinese-hackers-breach-federal-governments-personnel-office/2015/06/04/889c0e52-0af7-11e5-95fd-d580f1c5d44e_story.html

Narayanan, A., & Shmatikov, V. 2008. Robust De-Anonymization of Large Sparse Datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy,* SP '08: 111–125.

O'Leary, D. 2013. Artificial Intelligence and Big Data. *IEEE Intelligent Systems,* 28(2): 96-99. https://doi.org/10.1109/MIS.2013.39

Padgeta, J., & Vasconcelosb, W. W. 2015. Policy-Carrying Data: A Step Towards Transparent Data Sharing. *Procedia Computer Science,* 52: 59–66. https://doi.org/10.1016/j.procs.2015.05.020

Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. 2015. Big Data Privacy in the Internet of Things Era. *IT Professional,* 17(3): 32–39. https://doi.org/10.1109/MITP.2015.34

Ponemon Institute. 2014. *The Cost of Data Breach Study.* Traverse City, MI: Ponemon Institute.

Punagin, S., & Arya, A. 2015. Privacy in the Age of Pervasive Internet and Big Data Analytics: Challenges and Opportunities. *International Journal of Modern Education and Computer Science,* 7(7): 36–47. http://dx.doi.org/10.5815/ijmecs.2015.07.05

Rghioui, A., Aziza, L., Elouaai, F., & Bouhorma, M. 2015. Protecting E-Healthcare Data Privacy for Internet of Things Based Wireless Body Area Network. *Research Journal of Applied Sciences, Engineering and Technology,* 9(10): 876–885.

Russo, G., Marsigalia, B., Evangelista, F., Palmaccio, M., & Maggioni, M. 2015. Exploring Regulations and Scope of the Internet of Things in Contemporary Companies: A First Literature Analysis. *Journal of Innovation and Entrepreneurship,* 4(11). http://dx.doi.org/10.1186/s13731-015-0025-5

Samani, A., Ghenniwa, H. H., & Wahaishi, A. 2015. Privacy in Internet of Things: A Model and Protection Framework. *Procedia Computer Science,* 52: 606–613. https://doi.org/10.1016/j.procs.2015.05.046

Saroiu, S., Wolman, A., & Agarwal, S. 2015. Policy-Carrying Data: A Privacy Abstraction for Attaching Terms of Service to Mobile Data. In *HotMobile '15: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications:* 129–134. New York: Association for Computing Machinery. https://doi.org/10.1145/2699343.2699357

Schroeck, M., Shockley, R., Smart, D., Romero-Morales, J., & Tufano, P. 2012. *Analytics: The Real-World Use of Big Data.* IBM Global Business Services.

Shukla, S. 2015. Editorial: Big Data, Internet of Things, Cybersecurity: A New Trinity of Embedded Systems Research. *ACM Transactions on Embedded Computing Systems,* 14(4): 1–2. https://doi.org/10.1145/2820608

Sen, R., & Borle, S. 2015. Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems,* 32(2): 314–341. http://dx.doi.org/10.1080/07421222.2015.1063315

Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porsini, A. 2015. Security, Privacy and Trust in Internet of Things: The Road Ahead. *Computer Networks,* 76: 146–164. https://doi.org/10.1016/j.comnet.2014.11.008

Silver-Greenberg, J., Goldstein, M., & Perlroth, N. 2014. JPMorgan Chase Hacking Affects 76 Million Households. *The New York Times,* October 2, 2014. Accessed April 10, 2017: http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/

Singel, R. 2009. Probe Targets Archives' Handling of Data on 70 Million Vets. *Wired,* October 1, 2009. Accessed April 10, 2017: http://www.wired.com/2009/10/probe-targets-archives-handling-of-data-on-70-million-vets/

Smith, J., & Lee, M. 2015. Massive Hack of 70 Million Prisoner Phone Calls Indicates Violations of Attorney-Client Privilege. *The Intercept,* November 11, 2015. Accessed April 10, 2017: https://theintercept.com/2015/11/11/securus-hack-prison-phone-company-exposes-thousands-of-calls-lawyers-and-clients/

Solomon, H. 2015. Popular Canadian-Owned Dating Sites Including Ashley Madison Hacked. *IT World Canada,* July 20, 2015. Accessed April 10, 2017: http://www.itworldcanada.com/post/popular-canadian-dating-sites-including-ashley-maddison-hacked

Spiekermann, S., & Cranor, L. F. 2009. Engineering Privacy. *IEEE Transactions on Software Engineering,* 35(1): 67–82. https://doi.org/10.1109/TSE.2008.88

Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., Halunga, S., & Fratu, O. 2015. Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure e-Health Applications. *Journal of Medical Systems,* 39(11): 1–8. https://doi.org/10.1007/s10916-015-0327-y

Sweeney, L. 2002. k-anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 10(5): 557–570. https://doi.org/10.1142/S0218488502001648

Tsai, C. W., Lai, C. F., & Vasilakos, A. V. 2014. Future Internet of Things: Open Issues and Challenges. *Wireless Networks,* 20(8): 2201–2217. https://doi.org/10.1007/s11276-014-0731-0

van de Pas J., & van Bussel G. 2015. 'Privacy Lost - and Found?' The Information Value Chain as a Model to Meet Citizens' Concerns. *The Electronic Journal Information Systems Evaluation,* 18(2): 185–195.

Weber, R.H. 2015. Internet of Things: Privacy Issues Revisited. *Computer Law & Security Review,* 31(5): 618–627. https://doi.org/10.1016/j.clsr.2015.07.002

Westin, A. F. 1968. Privacy and Freedom. *Washington and Lee Law Review,* 25(1): 166–170.

Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. 2014. Privacy in the Internet of Things: Threats and Challenges. *Security and Communication Networks,* 7(12): 2728–2742. http://dx.doi.org/10.1002/sec.795

**Citation:** Adams, M. 2017. Big Data and Individual Privacy in the Age of the Internet of Things. *Technology Innovation Management Review,* 7(4): 12–24. http://timreview.ca/article/1067



**Keywords:** cybersecurity, Internet of Things, IoT, big data, privacy, smart devices, data breaches