**Alexa Knows More than You Tell Her: How Corporations Control Consumer Data**

STS Research Paper

Presented to the Faculty of the

School of Engineering and Applied Science

University of Virginia

By

­­­Jonathan Lowe

10/16/2017

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Approved: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Caitlin Wylie, Department of Engineering and Society

The term “the internet of things”, commonly referred to as IoT, refers to embedded computers connected to the internet. IoT devices range from thermostats and fitness trackers to cars. IoT devices will be widespread and produce a huge amount of information. If the current trend holds, there will be about 50 billion devices connected to the internet by the year 2020 (Federal Trade Commission, 2015). An average of 150 million points of data is made a day from all of the devices produced by a relatively small IoT company (Federal Trade Commission, 2015). IoT companies will collect huge amounts of possibly extremely personal consumer data, such as sensitive health information, driving habits, and whether or not you are at home (Federal Trade Commission, 2015).

The collection and retention of data collected by IoT systems present privacy risks that did not previously exist.Privacy, in this case, means the right of individuals to have a measure of control over the flow of personal information about themselves. Examples of information that falls under this definition of personal privacy include information about one’s medical records, voting history, academic records, and daily habits. Privacy applies differently to information depending on the sensitivity of that information, the context in which it is shared, and people’s expectations; people expect that their medical records would be kept more private than their recent shopping history.

IoT privacy risks firstly include the fact that data collected by IoT devices though not apparently sensitive can be used to infer private information and thus give corporations significant control over their users’ privacy. Companies have put a lot of effort into predicting what consumers will likely buy based on the information that they have and IoT devices can provide an unprecedented amount of information that they can analyze. Once corporations collect information from IoT devices about their users, they will effectively have exclusive legal control it. There is a whole industry around buying and selling consumer data already and there is little to no regulations governing buying and selling consumer information. Another privacy risk is the fact that many companies that store data on their users have historically misused and insufficiently protected that data; IoT device manufacturers will likely behave the same way. All this begs the question: How should the data produced by IoT systems affect personal privacy? The widespread use of IoT devices will make corporations effectively own a large amount of their users’ personal information, which the companies could easily misuse or put their users’ personal privacy at risk without significant legal ramifications.

To explore this question, I will use contextual integrity to analyze the situation. Contextual Integrity is a theory which states that every social context has norms about how information should be shared and what information is appropriate to share and that there is no context where anything goes (Nissenbaum, 2004). Whenever one of these norms is violated or information is revealed in the wrong context, a person’s privacy has been violated (Nissenbaum, 2004). If you share personal information about yourself, like severe health problems, to a friend, you would not expect them to share that with others. If they did share such personal details, without appropriate reasons, it would violate the social norm of keeping private information about one’s friends secret. Another example is when you go to a medical professional. They are expected to not disclose your medical history; this norm is enforced by law.

Contextual Integrity will act as a framework for analyzing privacy for IoT systems. I will use it to highlight that IoT systems mostly lack social norms governing the flow of consumer data and how IoT data will likely be used in ways other related industries use their consumer’s information. When contextual integrity is violated, I will also show that related industries lack sufficient legal protection for their consumers. Also, I will also cover how related industries can violate contextual integrity in new ways using data mining.

The data collected by IoT devices though maybe not apparently sensitive can be used to infer private information and thus give corporations significant control over their users’ privacy. Data mining is the process of taking large amounts of data and analyzing it to find patterns or relationships that exist between members of certain data points (Keating, 2008). Possible applications of this technique include figuring out if a company is insolvent, if a prospective employee will likely be successful, and what customers will likely purchase which products (Keating, 2008). This technique allows companies to obtain sensitive information about their customers that the customers did not directly provide, thus allowing them to obtain private information outside of an appropriate context. Companies are already using data mining to figure out information about their consumer's given different circumstances.

Retail stores have used data mining to infer private information. The retailer Target has a sophisticated model for predicting how much money a person will likely spend on different items (Corrigan, Craciun, & Powell, 2014). It has proven to be pretty accurate for at least a few types of applications. One way that they used data mining is predicting when a woman is pregnant. Target researchers found that pre-natal mothers usually bought certain items regularly, such as cocoa-butter lotion, a large purse, and magnesium supplements, and came up with a scoring system to predict the likelihood of a woman being pregnant (Corrigan et al., 2014). It is so accurate that Target knew that a teen girl was pregnant before her father (Lubin, 2012). This is at least slightly disturbing and is an example of how data mining can violate contextual integrity. Typically, women keep the fact that they are pregnant in within close family and friends and do not want other entities like retail stores to know. The information that women provide by what they purchase in the original context might be innocuous, but in the surrounding context of other people’s purchasing decisions, that information reveals additional data that is not appropriate in this new context.

IoT devices presents a huge source of new information that can be analyzed using data mining and thus gives companies more control of their customer's privacy. The data from an IoT lighting system could be used to tell when people get home from work, when they go to bed, and when they get up in the morning. Smart fridges that detect what food items you regularly eat could potentially tell how much you likely weigh, whether or not you are lactose intolerant, whether or not you have diabetes, or if you have heart problems by you eat. All of this evidence might suggest that I am arguing that companies should not use data mining. Shouldn’t companies be able to use data mining to improve their services? Yes, they should, but they should be limited in what information that they mine and use and how closely they should link it to individual consumers. Without restrictions on data mining, companies would store and have access to this information and would be able to violate the contextual integrity of the information given to them by consumers.

Once corporations collect information from IoT devices about their users, they will effectively have complete control it. The control that companies have over the data collected about their customers includes the right to sell it, mining it for more information, and releasing it to governments and other third parties. This allows companies to release information in inappropriate contexts and thus violate contextual integrity. As it stands now, companies in different and related industries can effectively do whatever they want with the data they collect about customers.

Search engines are an example of companies that collect information about their users and effectively have control over that information. Not too long ago, Aol released search results to the public that were organized by user ID; people were identifiable from the results (Barbaro & Zeller, 2006). This violated contextual integrity because the information about their user’s search history is expected to be kept only by the search engine. Aol did not face legal consequences for this release and any attempt to peruse such action would probably fail. Once search engines collect information about you they can do practically anything they want without significant legal ramifications (Grimmelmann, 2007). It may not even be a significant legal risk for a search engine to use their user’s data in a way that is against their own privacy policy (Grimmelmann, 2007). This effectively allows search engines to violate the contexts that they themselves have defined in their privacy policy without any legal ramifications.

Other companies like data brokers collect, buy, and sell personal information legally. Data brokers are companies that collect information on consumers, mine it for additional information, and sell it. Data broker companies typically are not very widely known; Acxiom and Datalogix are two of them. Their customer range from antivirus companies to political campaigns (Rostow, 2017). In a 2014 Federal Trade Commission report, the data brokers that were studied gather information from publicly available sources, such as government records and social media, as well as commercial sources; none of them collected information directly from people (Federal Trade Commission, 2014). The information that data brokers collect while validly used in their original contexts, like publicly available records and information collected about costumers by companies for their internal use, is should not necessarily be shared in this new context. Data brokers sell information for different purposes including marketing, risk mitigation, and people search without hardly ever interacting with the people they collect information about (Federal Trade Commission, 2014). Thus, most people don’t see that information about themselves is being bought and sold, which raises questions about whether or not companies should buy and sell this information in this new context.

Buying and selling information about consumers is perfectly legal. With a few exceptions, there are no laws regulating the data broker industry United States on the federal level (Rostow, 2017). Even on the state level, there has been little done to regulate the data broker industry (Rostow, 2017). This highlights the fact that if these companies violate the contextual integrity of consumer’s data, they will likely not face any legal repercussions. Given the quantity and quality of the information produced by IoT systems, it is likely that data brokers will want to get a hold of IoT data either by buying it or acquiring IoT device manufacturers. Also, IoT service providers will be tempted to sell that information because of the amount they can pad their profit margins. This would give data brokers a new source of data that was previously unavailable and new opportunities to violate the contextual integrity of their users’ data.

Since existing companies, which bear a striking resemblance to IoT service providers, have control over consumer data, IoT service providers will likely have the same control over data produced by their customers. IoT service providers will also appropriate the social norms of sharing consumer data from similar industries. All the evidence I provide might make it seem like I am arguing that companies shouldn’t have control over consumer data; shouldn’t companies be able to use information about their customers to improve their services, effectively market to their customers, and come up with new products? Yes, companies should have control over their consumer's data, but they should not sell, release it to third parties, or reveal it in inappropriate social contexts a without their customer’s knowledge or consent.

Companies that store their users’ information have historically misused and insufficiently protected that data; IoT device makers will likely behave in the same way. Such misuse involves releasing personal information about their customers in inappropriate contexts without their costumers’ knowledge or consent and using it in ways that harm consumers in one form or another. While even the most robustly developed software ends up having security vulnerabilities, companies have exposed their user’s data to hackers by not sufficiently building in sufficient security measures into their products, thus allowing unauthorized third parties to access and release information in inappropriate contexts.

Companies have released have released personal data on consumers without their knowledge or consent to different parties that could use that information to directly harm those consumers. PRISM was a secret government program by which the FBI obtained information on people directly from IT companies without a warrant (Gellman & Poitras, 2013). Companies like Yahoo, Google, and AOL have participated in the FBI’s surveillance program (Gellman & Poitras, 2013). This was an edge case of where contextual integrity was violated. Existing informational norms state that law enforcement should have access to private information in criminal investigations even without the person they are investigating’s knowledge or consent, but there should be some restrictions on that power. For example, it is the current norm that police should have a warrant before obtaining information from someone’s private email account. Even though there was a law which made PRISM legal, PRISM violated the expectations of consumers and thus ceased a public outcry. Another example of inappropriate use of consumer data includes the time that InfoUSA sold the information on elderly people which ended up being used by criminals to steal their life’s savings (Duhigg, 2007). Companies continue to offer such services, while collecting millions of dollars, despite being shown evidence that their services are being used to perpetrate fraud (Duhigg, 2007). This shows how companies are willing to give sell information on consumers for profit even though it will be likely be used to directly harm them.

Companies in the past have insufficiently secured their customer’s data, thus inadvertently providing third parties access sensitive information. From May through July, a credit reporting agency called Equifax had a massive security breach in which hackers stole the names, social security numbers, birth dates, addresses of U.S. customers (Equifax, 2017). Initial estimates stated that 143 million Americans were affected by this breach (Equifax, 2017). The scale of this breach is concerning, to say the least. According to their website, the security team was aware of the security vulnerability when it was disclosed by Apache Struts and were taking measures to secure their vulnerable systems (Equifax, 2017). This seems a little bit implausible. Apache announced this vulnerability on March 7, 2017, and released a patch for it the same day (Khudairi, 2017). Equifax had two months to secure their applications, but they only patched their applications after they discovered the method of attack (Equifax, 2017). If Equifax prioritized protecting consumer data, they would have patched their web applications significantly sooner. This incident shows a limitation of contextual integrity. Contextual integrity does not cover the responsibilities of persons to ensure that the information that they have is not taken by other parties as Equifax has here. While that responsibility could be added to social norms concerning information flow, contextual integrity would do well to add some expectations of how securely the parties involved should keep data secret.

In its relatively short life so far, IoT companies have already left customer’s data at risk by insecurely storing and communicating it. Security standards are either fragmented or not present at some IoT manufacturing companies and some IoT devices have no security measures built into them (Chen, 2017). Thus, many IoT devices present little to no obstacles to hacking. A few years ago, TRENDnet, Inc. marketed and sold IoT baby monitors as secure that were not so secure (Riga, 2017). When the baby cameras were hacked in 2012, the footage was posted on the internet and an investigation was launched that found that login information was not encrypted when it was being transmitted and thus capable of being read by attackers (Riga, 2017). This situation is likely to continue. According to a report by the Federal Trade Commission, IoT faces unique security challenges including new IoT companies lacking security experience, the inability to update device software, and the lack of willingness of companies to secure and fix low cost low profit margin devices (Federal Trade Commission, 2015). This situation will allow many more third parties, in this case, hackers, to access and disseminate information outside of their proper social norms than if more effort was put into security.

Since existing companies have historically misused and insufficiently protected their customer’s data, IoT companies will likely do the same. All of this evidence might suggest that companies can’t be trusted with consumer data; aren’t there companies that are trustworthy? Yes, there are many companies that have taken adequate measures to protect their customer’s data and don’t misuse it, but there have been several notable instances where the information released is damaging. So, there should be some legally binding expectations about how companies will use and protect consumer data.

The widespread use of IoT devices will make corporations effectively own a large amount of their users’ personal information, which the corporations could easily misuse or put at risk without significant legal ramifications. Corporations can gain access to more personal information about their users than what their users explicitly gave them by using data mining. If the data that IoT companies collect is misused, the users that produced that information will have little to no legal recourse. In the past, companies abused consumer data and this will likely continue.

Currently, this new technology lacks social norms for the flow of personal information. In many cases it will likely borrow contexts from existing ones: biometric information will likely borrow contexts from existing medical practices. In some cases, this will be adequate, but there will be many in which it is not. Therefore, new informational sharing norms should be formed. Companies should be able to collect and mine data on their customers so that they can improve their services, but they should be limited in what that can do with it. For example, companies should not be able to sell or release it on a whim. Corporations should also have the explicit legal responsibility to protect their consumer data from malicious hackers. These legal responsibilities should be different corresponding to how sensitive that information is.

This paper has some limitations and there are directions future research can take. For one, I did not look directly at tort law or case law for privacy. There could have been additional factors I did not consider. I could have used other STS frameworks such as AND or SCOTT, that could have provided better perspectives on the topic. Future research on the topic could take the direction of looking at existing legal principles on privacy and framing the problem in ANT to get a better perspective on the parties involved.

**References**

Barbaro, M & Zeller, T. (2006, August 9) A face is exposed for AOL searcher no. 4417749. *New*

*York Times.* Retrieved from http://www.nytimes.com/2006/08/09/technology/09aol.html

Chen, L. (2017). *Security management for the internet of things*. (Unpublished master’s thesis). University of Windsor, Canada. Retrieved from https://search.proquest.com/docview/1886841984/abstract/1DF89E7F2A5B4581PQ/2

Corrigan, H. B., Craciun, G., & Powell, A. M. (2014). How does Target know so much about its customers? Utilizing customer analytics to make marketing decisions. *Marketing Education Review*, *24*(2), 159–166.

Duhigg, C. (2007, May 21). Firms sell elderly Americans’ data to telemarketing con artists. *The New York Times*. Retrieved from http://www.nytimes.com/2007/05/21/world/americas/21iht-data.1.5803543.html

Equifax. (2017, September 15). *Equifax releases details on cybersecurity incident, announces personnel changes*. Retrieved from https://www.equifaxsecurity2017.com/2017/09/15/equifax-releases-details-cybersecurity-incident-announces-personnel-changes/

Federal Trade Commission (2014). *Data brokers: A call for transparency and accountability*. Washington, DC: Federal Trade Commission.

Federal Trade Commission. (2015). *Internet of things: Privacy & security in a connected world.* Washington, DC: Federal Trade Commission.

Frizell, S. (2014, November 19). What is Uber really doing with your data? *Time.Com*. Retrieved from http://www.time.com/

Gellman, B. & Poitras, L. (2013, June 7). U.S., British intelligence mining data from nine U.S.

Internet companies in broad secret program. *The Washington Post.* Retrieved from https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497\_story.html

Grimmelmann, J. (2007). The structure of search engine law*.* *Iowa Law Review*, *93*, 1–64.

Keating, B. (2008). Data mining: What is it and how is it used? *Journal of Business*

*Forecasting*, *27*(3), 33–35.

Khudairi, S. (2017, September 14) Media alert: The Apache Software Foundation confirms Equifax data breach due to failure to install patches provided for Apache® StrutsTM exploit. *The Apache Software Foundation.* Retrieved from https://blogs.apache.org/foundation/entry/media-alert-the-apache-software

Lubin, G. (2012, February 16) The incredible story of how Target exposed a teen girl’s pregnancy. *Business Insider.* Retrieved from http://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*, 119.

Riga, S. A. (2017). Two breaches, two enforcement actions, and a ddos attack: Data security and the rise of the internet of things. *Journal of Internet Law*, *20*(9), 3–7.

Rostow, T. (2017). What happens when an acquaintance buys your data?: A new privacy

harm in the age of data brokers. *Yale Journal on Regulation*, *34*(2), 667–707.