Data and Society  
Kevin, David, Rachel, Guannan

1. Introduction

Our lives today have been greatly shaped by technology. Advancements in consumer technology have given us a plethora of modern conveniences, ones that we probably won’t want to let go of. The services that enable these conveniences have been subtly integrating themselves into every aspect of our lives, and now we’ve come to wholly rely on them. We live in a world where our phones can tell our coffee machines to automatically have a pot of coffee ready in the morning, where our thermostats remember when we come home from work and at what temperature we like our homes, where our televisions recommend shows based on what we like to watch, where our maps can tell us when we should leave a party to catch the last train home at night, and the list goes on and on. These services and conveniences are made possible thanks to, yes of course GPS technology and machine learning, but potentially more importantly, the personal data that we surrender every day. The more data we give up, the better the services get. Which is exactly what the corporations that create these services want – for us to turn a blind eye to insecurity in exchange for services – and as corporations continue to feed on data insecurity\(^1\), their control over consumers increases. Like Schneier’s analogy to feudal lords, as our dependence on these services and corporations grow, their ownership of our lives also grows\(^1\). The ultimate consequence of allowing corporations to continue favoring insecurity for profit, is the power and control that they will hold over the users of their services.

2. Supporting Details

As Balzac puts it, “behind every great fortune, there is a crime.” Though most will not dramatically compare data insecurity to crime, the drive behind the data insecurity created by corporations is indeed profits. The business model that corporations build

\(^1\) In this paper, “data insecurity” refers to the act of businesses or corporations collecting more user data than is necessary to provide their service, with the goal of selling, monetizing, or misusing said data.
consists of systems spying on people in exchange for services and relying on user data to make a profit. It is described by Harvard Business School professor Shoshana Zuboff as “surveillance capitalism”. It creates a new marketplace by utilizing personal information as an exchangeable commodity that can be transformed into user predictions.

This is a new form of capitalism nourished by modern technology. Invented by the emergence of the dot-com burst, this business model came to realize that user data was the truly valuable byproduct of online surfing; though initially, it was only interested in whether internet users would generously click on advertisements or not. What came to fruition afterward include the rise of technology corporations, the dominance of Google, and the flourish of Silicon Valley. The surplus of data that can be easily, freely, and conveniently collected by means of "smart" devices and "personalized" services has become the lucrative fodder to feed those ever-growing corporations. And yet, they want more. Claiming faster, safer, and better – in other words, for the good of humanity – corporations even beyond the technology sector have begun to invasively collect our data. A comment from The Washington Post points out this tragedy: “their predictions are about us but not for us[2].”

For these companies, this is still not enough. The smart, data-hungry corporations realized they could do more than automatically collect information generated by us, but they could also automate us. Through economic influence, these corporations can influence our everyday choices. Shockingly, not everyone is aware of this. Think about every time you call “Hey Alexa!” without knowing the little smart device is actually “spying on” you and collecting tons of data beyond what it needs to carry out the task you ask it to[3]. The fact that most people lack the awareness to protect their privacy and are ignorant of how corporations acquire and use personal data actually encourages corporations to keep plundering our security. Indeed, our ignorance is their bliss.

Earlier this year, Apple’s chief executive officer Tim Cook called for a more comprehensive approach to privacy legislation: “Right now, all of these secondary markets for your information exist in a shadow economy that is largely unchecked and out of sight of consumers, regulators and lawmakers[4].” This undoubtedly references
the lack of relevant public policy and the need for new infrastructure to strengthen our understanding of privacy. Obviously, the public and democratic institutions in place today have not devoted enough attention to this mechanism and potential issue.

As corporations continue taking advantage of user data, governments keep busy surveilling the public by similar means, however troublesome it seems[5]. This only exacerbates the situation as victims of data insecurity keep feeding their data to sources that make them insecure, while simultaneously, the justice system is making things worse and even shutting their eyes to the problem. The result is no secret – data insecurity is favored, supported, and sponsored by corporations.

3. Potential Solutions

The problem of large organizations (such as businesses and governments) favoring insecurity is a difficult knot to untangle and defies easy solutions, but ultimately, it is fundamentally a problem of incentives. Organizations favor insecurity because, from their perspective, the benefits they receive from collecting user data far outweigh the costs. Solutions then need to address one or the other of these two major factors. To encourage organizations to favor security over insecurity, we must either reduce the benefits they receive from collecting and storing data or increase the cost.

Not all collection of user data is nefarious, and some collection provides enormous benefits for us as users, citizens, and consumers. For example, it is enormously convenient for our navigation apps to predict where we might want to go next or for hospitals to be able to make predictions about our future health based on our past health data. Whatever solutions we propose ought to encourage organizations to continue to offer and develop these beneficial services. We would like to encourage organizations to derive as much benefit as possible from the most limited information possible. In other words, rather than reducing the benefit that organizations receive from user-data based services, we ought instead to focus on increasing the costs of the most egregious and unnecessary data collection.

In this report we will propose three possible solutions and discuss the benefits and disadvantages of each. More specifically, our proposed solutions are 1) Increasing regulation around the collection and use of user data, 2) Introducing taxation to increase
the cost of acquiring and storing user data, and 3) Shifting public perception of the practice of extreme user data collection by educating the public on the accompanying dangers of these practices.

Perhaps the most straightforward solution to the problem of organizations favoring insecurity by collecting user data is to introduce new regulation in the style of Europe’s General Data Protection Regulation (GDPR) and make it unlawful to collect certain kinds or quantities of user data. This approach has the advantage of having been tested on a large scale in Europe and in other countries where it has, at least initially, seemed very effective in changing the practices of businesses operating in those countries. This approach also has the advantage of being able to be applied equally to the collection of data by governments and by private organizations. One major downside of this approach is the need to create a large bureaucracy to enforce a huge number of sometimes vague regulations.

A second solution, one that has not previously been tried on a large scale, is to introduce a new tax on user data. Under this scheme, organizations (primarily businesses) would be required to pay a certain amount of money for each bit of user data they collect and store. This would encourage companies to produce as much value from as little user data as possible, the exact outcome we hoped to achieve. Additionally, the simplicity of such a tax could make compliance and enforcement of the policy simpler than with a wide-reaching regulation in the style of GDPR. One major disadvantage of this approach is that it applies only to businesses and is not clearly applicable to government collection of data.

A third solution, and perhaps the trickiest, would be to turn public perception against companies that engage in unnecessary data collection through a campaign to educate the public on the many dangers of large scale data collection. This approach has the advantage of not requiring any new legislation of any kind to be introduced or enforced. The major disadvantage is that it is not clear that such an effort could really succeed when the organizations most affected by such a campaign have enormous resources of their own to fight back. Though simple, this approach could be very expensive and ineffective in practice. This solution is closest to what we have actually pursued in the United States up to today.
4. Necessary Infrastructure

The three aforementioned solutions to the problem of corporations collecting unnecessary user data require different infrastructure developments, but share one common and convenient element: no need for additional data storage infrastructure. Each solution calls on companies to collect less user data than they currently do, so no additional data centers or other physical data storage solutions will need to be constructed. Despite this, each solution brings its own unique infrastructure requirements.

The first solution, implementing a GDPR-style legislation, is the most straightforward in the way it solves the problem of corporations collecting unnecessary user data, but is far less straightforward from an infrastructure perspective. In Europe, getting GDPR up and running was a lengthy ordeal taking four years for the legislation to be passed and an additional two years to be fully put into practice[6]. This six year process also involved the creation of multiple advisory bodies, including the European Data Protection Supervisor and the European Data Protection Board[7,8].

The United States currently has no official data protection agency, with the closest regulatory body being the Federal Trade Commission (FTC). The FTC’s main job is to prevent unfair or deceptive trade practices, not promote data security. The FTC has charged some companies for having data breaches on the grounds that their lack of reasonable data security was a deceptive trade practice. The FTC’s ability to do this may not continue, as some companies have begun to protest this and because the United States has no laws or guidelines on what “reasonable data security” measures are[9].

The United States’ complete lack of data protection regulation means that an immense amount of governmental infrastructure would need to be developed to implement this solution. Even then, once the legislation is in place businesses will have to change their employee and protocol infrastructure (introducing roles similar to the “Data Protection Officer” required by GDPR) to maintain compliance with new regulations.

The second solution to corporations favoring data insecurity is to introduce a tax on user data collected by businesses. The infrastructure needs here are similar to—but
more lax than—the first solution. Legislation would still be passed to define and implement the tax, but this legislation should be simpler because it only concerns the amount and type of data being stored, not how or whether it can be stored. Tech giants like Google and Amazon have a history of using loopholes in legislation to dodge taxes\textsuperscript{10}, so implementing this tax in an effective way will be difficult. It may require infrastructure developments such as new administrative bodies in the government to ensure compliance.

The last solution is the trickiest to implement in terms of solving the problem, but the easiest when it comes to infrastructure: changing public opinion. This requires no new legislation, no new advisory bodies, and no new employees at companies (aside from perhaps more public relations personnel). The legwork of this solution will be done through current channels of social and conventional media, requiring no major infrastructure changes.

5. Metrics for Success

No matter how great a proposed solution may be, it is pointless if the solution is unable to actually address the issues at hand. With this in mind, we must consider each of the above solutions within the scope of the problem. All three of the proposed solutions are intended to encourage corporations to favor security by either reducing the benefits they reap or increase the cost of data collection and storage. Each of the metrics that will be explored, will tackle whether each solution does indeed do just that.

One of the most obvious considerations to be made when measuring the success of the solutions would be the amount of data that companies collect. According to a report by DOMO, a mobile, cloud-based operating system service, over 2.5 quintillion bytes of data are created each day, and at least 90% of all the data that we have was created within the last two years\textsuperscript{11}. Google alone collects an exorbitant amount of data. A reporter for the Daily Mail UK estimated that Google was able to collect 569,555 pages worth of data about him in a 12-month period\textsuperscript{12}. If the solutions proposed in section 3 are indeed successful, then surely the sheer amount of data collected everyday would go down. With new legislation surrounding data being collected, and new taxes enacted, the cost of collecting such would begin to outweigh
its benefits – thus encouraging corporations to reconsider collecting such large quantities of data. In addition, if consumers are more educated regarding the collection of their data, they would be more conscious and careful about what permissions they grant to certain applications and what information they make public.

Aside from just the volume of data collected, the type of data collected by corporations can seem very unusual and even unnerving at times. There have been cases where seemingly innocent applications, such as a flashlight app, were requesting access to user location information and even access to their phone microphones\(^{[13]}\) – both, as most people would agree, are things that a flashlight has no business knowing. Furthermore, many services collect data that can seem unnecessary, and it’s difficult to know exactly how they plan on using that data. Therefore, another metric to consider is the level of transparency of what pieces of data are collected and what their intended uses are. If said transparency increases, it is an indication that the proposed solutions, such as legislation, have made some form of impact.

Lastly, assessing public perception of large corporations and data collection can be used as a metric to measure the success of the proposed solutions. According to a 2015 article on the Pew Research Center website, “...Americans also have a pervasive sense that their physical activities may be recorded when they are moving about their daily lives.”\(^{[14]}\) Around 81% of the people that were surveyed agreed that surveillance is hard to avoid. As the proposed solutions encourage corporations to begin favoring security, public perception of these organizations will grow more positive.

6. Next Steps

With three different proposed solutions to solve the data insecurity problem (listed in section three), figuring out exactly what to do next is no small task. To decide which solution to attempt first, we look to how difficult each solution would be to implement. In order of least to most substantial required infrastructure changes (as discussed in section four), we have: 1) changing public opinion, 2) implementing a tax on user data, and 3) implementing a GDPR-style regulation.

The first solution we should attempt is changing public opinion. This can be started almost immediately and at relatively small cost. Using existing low-cost media
channels (social media, internet advertising space, etc.), an ad campaign can be launched to inform the public of the dangers of businesses collecting unnecessary user data. If this campaign shows signs of success (through methods discussed in section five), it can be expanded to more expensive media channels (television, print). If this solution doesn’t show signs of success, it can be scrapped and solutions two (tax) then three (legislation) can be attempted. Again, success should be measured in the early stages of implementation to determine if it should continue.

An additional first step to consider is, in addition to the three proposed solutions, improving general data security. If stopping businesses from collecting user data in the first place will take a long time due to infrastructure needs, an additional measure that could be implemented alongside these long-term solutions is capitalizing on the already growing trend of tech-savvy users preferring strong data security practices. A Pew Research study in 2013 showed that Americans were favoring stronger data security practices more than ever before[15]. Also, security related websites and services are more popular than ever, with more coming into existence every year. These include password managers (KeePass, OnePass, etc.), security checkup websites (Haveibeenpwned), and many major websites implementing two-factor authentication on log-in (Gmail, Blizzard, Steam). Public opinion toward data security is already changing, so a lot of the legwork is done. A good first step is to begin securing the user data that has already been collected while we work toward the greater goal of not collecting that data in the first place.
7. References


think-again/


