https://doi.org/10.48009/2_iis_2019_183-194

Issues in Information Systems
Volume 20, Issue 2, pp. 183-194, 2019

FALSIFYING PERSONAL DATA TO ADDRESS ONLINE PRIVACY ISSUES

Vasilka Chergarova, Nova Southeastern University, vc574@mynsu.nova.edu
Ling Wang, Nova Southeastern University, lingwang@nova.edu

ABSTRACT

Companies today track people's browsing activity and collect their personal information. Amazon transactions cover 50% of the online transactions in the USA, and Google and Facebook cover most of the digital advertising. However, a significant disconnect exists between how the data is being used, sold, or shared, and what the user actually may want. Currently, no law or regulation encompasses how data brokers should handle private information. Several incidents, as Snowden revelations, Brexit, Cambridge Analytica scandal influencing on the United States election vote in 2016, are an example where Big Data collection of personal data is misused. Online users are adopting criminal like behavior to protect their own information. Considering the collection of personal information to be unfair, users are withholding personal data or giving false information. This paper proposes a Design Science Research study methodology to assess if falsifying personal data in online forms prevent the leak of personal information. It is very important to educate students to protect information infrastructure at their future work place and also keep the same practices when it comes to their private information.

Keywords: Information Privacy, Falsifying Personal Information, Risk Mitigation

INTRODUCTION

Low-cost hardware platforms and new data mining frameworks are enabling the rapid growth in private data collection and regulations cannot keep up with the change. Data appetites are getting bigger and bigger every day. Consumer information is being sold to data brokers or marketing firms, where, repacked, it is sold again for profit. One of the biggest data broker company, Acxiom ($1.1 billion for FY2013) advertised transparency about the data they collect. At their portal (aboutthedata.com), users can verify their biography, education, marital status, children, homeownership, mortgage amount, property size, vehicles (make, model, year, color), active investment portfolio over $150k, recent purchases, clothing size, sports, hobbies, pets, text-messaging, cholesterol related products, charity, etc. (Singer, 2013). The portal, currently in 2019, is dysfunctional. When it comes to medical data trading, IMS Health is dominant, achieved $2.6 billion in revenue in 2014 (Turner, 2016). Petabytes of data from pharmacies, insurance companies, medical organizations, federal and state health department are sent automatically to ISM. According to Turner (2016), the company has half a billion dossiers on individual patients, and it is not restricted by the Health Insurance Portability and Accountability Act (HIPAA) of 1996 because the data is anonymized (contain the only year of birth, gender, partial zip code, and doctor's name). In May 2016, IMS Health merged with Quintiles Transnational, the world's largest provider of biopharmaceutical development and commercial outsourcing services. The equity market capitalization is $17.6 billion, and enterprise value is more than $23 billion ("BRIEF-IMS Health and Quintiles to merge," 2016). Other small companies collecting data, and some of the data is available online free or for payment as shown in Table 1. Sweeney (2000) found that combinations of few characteristics (5-digit ZIP, gender, date of birth) are necessary to identify an individual in the USA. The data stream used by the data brokers comes from various online and offline sources, including email, websites, social media, U.S. Census records, retailers, DMVs, real estate records, birth certificates, marriage licenses, divorce records, state professional and recreational license records, voter registration information, bankruptcy, and more. In most cases, the data is often collected unknowingly to the individuals involved (Anthes, 2015). The data is analyzed, synthesized, and in some instances reidentified, then sold to other businesses for a variety of purposes. With the current technology, matching anonymity is just a matter of time and resources. The media and the legal system are questioning the company's practices (Turner, 2016). The USA Federal Trade Commission conducted an in-depth study about the industry and practices of the data brokers (Ramirez, Brill, Ohlhausen, Wright, & McSweeny, 2014). While there exist certain benefits for the consumers who use data brokers' products, the data use can pose risks to consumers, such as the facilitation of harassment, stalking, exposure of domestic violence victims, and other harmful activities.

Table 1 List of data brokers *(Grauer, 2018a)*

Addresses.com	Intelius.com	PhoneDetective.com	thatsthem.com
AddressSearch.com	LookUp.com	Pipl.com	truthfinder.com
AnyWho.com	MyLife.com	PrivateEye.com	usa-peoplesearch.com
Archives.com	Nuwber.com	PublicRecords.com	usidentify.com
BeenVerified.com	PeekYou.com	PublicRecordsNow.com	ussearch.com
Classmates.com	PeopleByName.com	PublicRecords360.com	veromi.net
DOBSearch.com	PeopleFinder.com	Radaris.com	whitepages.com
FamilyTreeNow.com	PeopleLooker.com	ReversePhoneLookup.com	zabasearch.com
InfoTracer.com	PeopleLookUp.com	earchbug.com	zoominfo.com
InstantCheckmate.com	PeopleSearchNow.com	spoke.com	Innovis
InstantPeopleFinder.com	PeopleSmart.com	spokeo.com	

Besides, data brokers store data indefinitely. If such personal data is bought or hacked, it may create a serious security risk that outwaits the benefits. According to a recent IBM study, for 50 million stolen records the cost will be approximately $350 million (*2018 Cost of a Data Breach Study: Global Overview*, 2018). Personal data is processed for political and economic reasons without users' consent. Cambridge Analytica is a data analytics firm, currently under investigation from the UK and the USA, who played a significant role in the Leave campaign for Britain's EU membership referendum (Brexit) and later on was linked to the USA election and Russia (Cadwalladr & Graham-Harrison, 2018). In the Cambridge Analytica scandal, Facebook gave unauthorized access to personally identifiable information (PII) of more than 87 million unsuspecting users. The firm integrated the Facebook data with pre-existing various personal information (browsers history, online purchases, voting results, etc.) to build 5,000+ data points on 230 million people and used to create a micro-targeted adds influencing user's election vote (Isaak & Hanna, 2018). The goal of the suggested study is to investigate how an intentional usage of fake data profiles in online forms affects the data collection by data brokers, as a possible solution to minimize privacy risk presented by the privacy paradox.

LITERATURE REVIEW

Currently, consumer' information is being sold as a commodity for profit. In response, some consumers are adopting criminal-like behavior and providing fake information online. There is limited research on online information disclosure (Smith, Dinev, & Xu, 2011) and why consumers are providing false data in online forms. Users provide inaccurate information in response to increased privacy concerns (Milne & Boza, 1999; Sheehan & Hoy, 2000), as a technique to control the flow of information (Chen & Rea Jr, 2004), to protest corporate overreach (M. Lwin, Wirtz, & Williams, 2007), to mask personal information (Wirtz & Lwin, 2009), express attitude toward a business or a government (Vitell, 2003).

What is Information Privacy

Warren and Brandeis (1890) described general privacy as "the right to be left alone." According to the Nissenbaum (2004) Contextual Integrity theory, the context of the information is crucial, and a single solution that fits all situation concerning privacy is not possible. Several aspects as the nature of information and the situation, the roles of the participants receiving information and their relationship, and terms for sharing and further dissemination of the information, should be considered to determine a violation of privacy (Nissenbaum, 2004). In the quest for providing privacy definition, Smith et al. (2011) determined that information privacy is not anonymity, secrecy, confidentiality, security, and ethics. The context of privacy is related to the type of information (e.g., behavioral, financial, medical, biometric, consumer, biographical), the use of information by sector (e.g., healthcare, marketing, finance), political context (e.g., law enforcement, government, public data, media), and technological applications (Smith et al., 2011). Exploring the construct of information privacy within the information systems domain,

Bélanger and Crossler (2011) defined information privacy as an option for a person to control the disclosure of personal information. Schwartz and Solove (2011, 2012) further segmented the personal information in three information categories where a person can be singled out from others (identified), identification is possible (identifiable), and low risk of identification (non-identifiable information). The lack of having a precise definition of privacy information is creating more challenges for legislative authorities to address issues with privacy violations. What one person can consider private, others may not.

Big Data and Privacy

Today, digital information is stored, shared, combined, and duplicated fast and cheap. The European Data Protection Supervisor (EDPS) defined Big Data as "large amounts of different types of data produced at high speed from multiple sources, whose handling and analysis requires new and more powerful processors and algorithms" (van der Sloot & van Schendel, 2016, p. 113). Exabytes of data are flowing daily from Facebook, Google, ISP, phone providers, credit card transactions companies and more. A study shows that it would take 13,513 Boeing 747 planes to transport one exabyte of data recorded on DVDs (Kuner, Cate, Millard, & Svantesson, 2012). According to Munir, Yasin, Hajar, and Muhammad-Sukki (2015), scientists have to come up with new terms like zettabyte and yottabyte, to accommodate the deluge of data and there are significant benefits from the use of Big Data as increased efficiently, discoveries, variability, and improved performance, better customization, automated online support, and innovated products, and services (Munir et al., 2015). Polonetsky and Tene (2013) argue that Big Data creates tremendous opportunity for the world economy, national security, marketing, credit risk analysis, breakthroughs in medicine, data security, mobility, smart grid, energy use, and urban planning. Big Data represents the "Holy Grail" of online marketing ("The Privacy Legal Implications of Big Data: A Primer," 2013). At the same time, misuse of Big Data can lead to discrimination, overcriminalization, and other restricted freedoms. Facebook has profiles on 1.9 billion users, Google 2 billion, Apple 1 billion, Experian has credit data on 918 million, Equifax 820 million, TransUnion 1 billion, Acxiom 700 million, Oracle 1 billion (Christl, Kopp, & Riechert, 2017). Consumers regularly disregard the online Privacy Policies and Terms & Conditions required by the Federal Trade Commission (FTC) Fair Information Practice Principles, which often do not include how the data is sold further to a data broker ("The Privacy Legal Implications of Big Data: A Primer," 2013). Tracking online activities can create several implications for the user (Bujlow, Carela-Español, Solé-Pareta, & Barlet-Ros, 2015). Credit companies often are using customer's online activities to assess their creditworthiness. Kreditech (Germany) uses up to 8,000 data points to evaluate a loan application, Lenddo (Philippines, Colombia, and Mexico) uses customer' Facebook friends payment history to determine the credibility of the customers, and Kabbage lending (USA) require access to PayPal and eBay borrower's accounts (Lobosco, 2013). Mac users pay on average 30% more than a PC user for a hotel room when using Orbitz website (Simon, 2012) and when using computers at a different location to rent the same car at Hotwire is changing the price with 25% (Elliott, 2009). Polonetsky and Tene (2013) argue that privacy risks should be weighed against Big Data rewards and the society at large should consider who is benefiting from the Big Data analysis, what are the benefits, and the level of certainty of the realization of those benefits. The benefits to the society can outweigh the concerns of risk in individual privacy; however, individuals still don't have access to the aggregated data from various institutions to ripe personal benefits of Big Data (Tene & Polonetsky, 2013).

Regulations Addressing the Collection of Personal Information

The Edward Snowden revelation about the practices of the US National Security Agency (NSA) in 2013 created a clearer picture of the war against information privacy (Ephraim & Discourse, 2017). Bulk collection practices of foreign and domestic information rippled in the European Union's decision to create the General Data Protection Regulation (GDPR) in 2016 which became effective in May 2018. Some of the key points of the GDPR are to synchronize data privacy laws across Europe, protect EU citizens data privacy and reform the way organizations approach data privacy "{Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)}" 2016). The failure to comply can cost up to €20 million or up to 4 percent of annual global revenues of the companies (including USA companies).

At the same time, the USA Federal Communications Commission (FCC) proposed rules that required privacy protection from the Internet Service Providers (ISPs) and to provide opt-out option, customer permission for collecting and sharing data, including user's health and financial details, browsing history, app usage, and geo-location ("FCC Adopts Privacy Rules To Give Broadband Consumers Increased Choice, Transparency And Security For Their Personal Data," 2016). However, as a result of lobbying for ISPs and tech firms, lawmakers were persuaded to dismantle such regulation (Kindy, 2017). Currently, there are no restrictions on what ISPs could do

with the consumer's data. Some states have initiated their own rules for the protection of private information. Some companies offer to opt-out and request for the removal of personal data. However, the consumers are faced with a labyrinth of online pages, various verifications, and in some cases payment. Currently, in the USA, there is no clearly defined legislation addressing the handling of private data from data brokers (Grauer, 2018b). To address privacy concerns, Klosek (2014) argues that a new privacy law should follow the existing regulations applied to use non-public personal information in as the Gramm-Leach-Bliley Act (GLBA) for financial sector, HIPPA for the health sector, Fair Credit Reporting Act for credit reports and state laws as the Massachusetts Data Security regulations. Some states are addressing the issue locally. The California Consumer Privacy Act (CCPA) requires the presence of deletion request, disclose any third parties' companies, and an opt-out option. The Vermont law - H.764, requires data brokers selling data to third parties to implement a written information security program, disclose the type of the collected data, and an opt-out option. Illinois's Biometric Information Privacy Act (BIPA) is addressing companies collecting fingerprints, face scans, or other biometric identifiers from consumers and employees (Lazzarotti & Gavejian, 2019). Despite all the attempt to address privacy concerns, currently, the USA is the wild west for Big Data. Giants like Apple, Amazon, Facebook, Google, and Microsoft spent a combined $64 million to shape U.S. regulations and stave off government scrutiny in 2018 (NG, 2018; Romm, 2018).

Technology Addressing Privacy Invasion

Norberg, Horne, and Horne (2007) present privacy concerns about consumer information been collected by marketers to leverage current technology and data analysis. Technological attributes contributing to the information privacy concerns are direct marketing, internet and e-commerce, data mining and profiling, monitoring and surveillance, communication, ubiquitous computing, and web 2.0 (Smith et al., 2011). Moura and Serrão (2015) argue that Big Data application posits significant challenges to security and propose the use of Social Networks Right Management System with Encryption and Software Defined Networking (SDN), as possible solutions. By using such a system, the user can define personal privacy policy and only connect to social networks that comply with it. Van den Hoven, Blaauw, Pieters, and Warnier (2014) argue that technology should be designed to accommodate privacy requirements in a way to prevent privacy violations. Several antitracking tools and methods already exist. Tracking Protection List (TPL) can block third parties content (Li, Hang, Faloutsos, & Efstathopoulos, 2015), ShareMeNot and Privacy Badger are browser extensions that block Facebook tracking activity (Lécuyer et al., 2014). Additional antitracking tools are AdBlock Plus, Request-Policy, Zend2, Kproxy, SecurityKISS, CyberGost, VPN Services, Tor browser, Privoxy, NoScript, Flashblock, Vanish, Disconnect, and Meddle (Bujlow et al., 2015). Other security habits as opting-out cookies, Do Not Track browser settings, use of privacy-focused search engines as DuckDuckGo, Startpage, Ixquick, private browser mode settings, regular clean-up of browser cache and history, and use of e-mail aliases can also prevent private data collection online (Bujlow et al., 2015). Lécuyer et al. (2014) created the XRay plugin to research sensitive-data targeted advertisement practices used by Gmail, Amazon, and YouTube. In one of the scenarios an individual created multiple fake emails containing words as "cancer," "AIDS," "bankruptcy," and "unemployment" and received numerous advertisements from an insurance company matching the illness-related fake emails. Lécuyer et al. (2014) repeated the experiment with the same result with other sensitive keywords for topics as Alzheimer, cancer, depression, African American, homosexuality, pregnancy, divorce, debt, etc. Enck et al. (2014) created a dynamic taint tracking system TainDroid, to track the flow of privacy-sensitive data from third-party mobile applications and found that 20 applications are misusing user's private information. In an undercover study of 150 smartphone apps and 120 websites in Germany, Herrmann and Lindemann (2016) observed 20% of the owners disclosing personal data to impostors. Meddle is a tool designed by students to filter the traffic from mobile devices, support VPN, blocking advertisement and provide visibility of all established connections on the device (Rao et al., 2013). The application is informing the user on the apps access to personal information (and its destination) and enable the user to decide on how to proceed. Another experimental service called $heriff, allows the online customers to trace dynamic price discrimination based on the location, browser/OS, incoming link, navigation history, and more (Iordanou, Soriente, Sirivianos, & Laoutaris, 2017). Disconnect is a privacy protection tool for browsers and mobile devices which provide real-time visualization of the online trackers and has a VPN capability to block over 5000 trackers (Heitzmann, Jackson, Oppenheim, & Toyens, 2011). Current technologies to prevent web browsers from tracking and recording private information are based on the concept of blocking tracking activities. Despite all available tools addressing such activities the problem persists. The privacy paradox is a perfect example of failing to protect personal information using the available tech tools.

Privacy Paradox

Privacy paradox is the contradiction of a user's information privacy attitude vs. actual behavior (Brown, 2001; Norberg et al., 2007). The privacy paradox is a perfect example of the use of technology failing to protect personal information by trading personal information for a perceived benefit. Norberg et al. (2007) empirically demonstrated that people provide more personal information that they say they will. Several studies estimated the trade-off value of personal information for a perceived benefit, consistent with (Debatin, Lovejoy, Horn, & Hughes, 2009; Lee, Park, & Kim, 2013; Polonetsky & Tene, 2013). In several experiments, Huberman, Adar, and Fine (2005) estimated the price for which people are willing to disclose their age ($57.56) and weight ($74.06). Hann, Hui, Lee, and Png (2007) estimated that improper handling of personal information ($30.49 to $44.62) when the user has to choose convenience vs. privacy protection. The two-phase study by Carrascal, Riederer, Erramilli, Cherubini, and de Oliveira (2013) showed that personal information value is significantly low. The study estimated that the monetary value for which a user agrees to disclose personal information in exchange for benefits, is on average 7€ for a user's browsing history, 25€ for offline personal information (age, address, economic status), 12€ for social networks interaction, 15.5€ for finance websites, 2€ for search activities, and 5€ for online shopping. Contrary to that, Egelman, Felt, and Wagner (2013) study showed that smartphone users are willing to pay more for privacy protection when presented with such an option. In online social networks, the benefits of a social relationship, social validation, self-representation, and diversion and entertainment outweigh the risk of self-disclosing (Debatin et al., 2009; Lee et al., 2013). Another reason for over-disclosure on a social website without concerns for the associated risk is perceived social capital gain (Ellison, Vitak, Steinfield, Gray, & Lampe, 2011; Stutzman, Vitak, Ellison, Gray, & Lampe, 2012). According to Mothersbaugh, Foxx, Beatty, and Wang (2012), privacy paradox occurs from unaccountability for information sensitivity. To provide a better understanding of the privacy paradox, Flender and Müller (2012) apply a quantum theory concept where the decision of human been is determined at the time of the decision but not before. Baek (2014) argues that discrepancy between privacy attitude vs. actual behavior will disappear if a person is offered the option to choose how online companies will use personal information. The privacy paradox makes it harder for government policymakers to address the raised privacy concerns (Kokolakis, 2017). When there is no law to address specifically the collection of private information, people tend to falsify their information online if they consider that the companies are acting unethically (Punj, 2017).

Why People are Falsifying Online

The growing customers' skepticism is fueled by the constant collection of personal information from companies offering a variety of online retail and services. More and more consumers clearly understand the significance of sharing personal data online. In a research survey for consumer perceptions of online registration by Blue Research, 76% admit to having provided incorrect or incomplete information when filling out a form for a new account online (Olson, 2011). Another research commissioned by RSA network security company surveyed 7500 consumers from France, Germany, Italy, UK, and the USA shows 41% to intentionally falsifying personal information online when signing up for product and services and 69% entirely boycotting the companies that are knowingly mishandling personal data (Ismail, 2018). The most frequently falsified types of personal information include a phone number (27%), date of birth (17%), email (16%), address (15%), name and age (14%) (Ismail, 2018). M. O. Lwin and Williams (2003) research on antecedents leading to the fabrication of personal information suggests that attitudes, perceived behavioral control, and perceived moral obligation play significant roles in fabrication behavior, while subjective norms are not. Metzger (2007) used the Communication Privacy Management (CPM) theory (Petronio, 2006) to investigate the relationship between information disclosure and privacy, including withholding and falsifying information, in e-commerce. According to CPM, people believe that they have the right to control their own information using personal privacy rules (Petronio, 2002). The findings suggest that consumers have limits when it comes to disclosing sensitive personal information online. Blank, Bolsover, and Dubois (2014) found that young people tend to do more to protect their data online more than the older ones. To protect private information, young people take varieties of action on social media website such as using pseudonyms and providing false information (Miltgen & Peyrat-Guillard, 2014), applying profile restriction and privacy settings (Hargittai, 2010), and limiting friends, tags, and photos (Young & Quan-Haase, 2013). Boyles, Smith, and Madden (2012) found that 54% of mobile users have decided not to install certain apps (54%) and removing already installed apps (30%) on their phone considering how much personal data the app collects. People falsify their information online for varieties of reasons. Rainie et al. (2013) estimated that 86% of American consumers have falsified or misrepresented their information online. Mehmood, Natgunanathan, Xiang, Hua, and Guo (2016) suggest falsifying data before further distribution to a third party as an access restriction technique against non-trusted online practices. The data can be distorted by using tools as Socketpuppet and MaskMe where individual's real behavior online is obscured by using a

false identity and pretending to be someone else (Xu, Jiang, Wang, Yuan, & Ren, 2014). According to Miltgen and Smith (2018) survey' results, 75% confirm that they have given inaccurate information in response to data requests and that context is playing an important role in the decisions to adopt falsification and withholding behavior. People are more likely to falsify data because the collection of sensitive data creates a sense of discomfort and perception of personal invasion (Malheiros, Preibusch, & Sasse, 2013; Metzger, 2007). Tene and Polonetsky (2013) argue that organizations should be more transparent about their data collection practices and enable users to declare their own policies, preferences, and terms or engagement. When such an option is not offered some users adopt criminal-like behavior and falsify the requested information while filling up online forms gatekeeping the information they are pursuing to prevent the misuse of personal data. A primary motivation for such a behavior is the attempt to escape unsolicited advertising (Malheiros et al., 2013). Poisoning of online data as a risk mitigation strategy for privacy protection is not widely researched and could be further investigated. So far, there is limited research about the reasons that make users disclose, withhold or falsify personal information on web forms (Malheiros et al., 2013).

PROPOSED RESEARCH METHODOLOGY

A study by Choi, Levy, and Hovav (2013) supported the ineffectiveness of self-reporting tools (survey) when it comes to assessing cybersecurity skills. According to Vaishnavi and Kuechler (2004) Design Science Research creates missing knowledge using design, analysis, reflection, and abstraction. The elements of design science research includes identifying the problem, defining possible solution objectives, designing and development, demonstration, evaluation, and communication (Gregor & Hevner, 2013; Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). To assess sensitive-data targeting practices by advertisers, Lécuyer et al. (2014) used the XRay tool. To address the phenomenon of the privacy paradox, the suggested study will create an artifact tool, similar to the Google Chrome browser extension of storing your data for further use ("Fill out forms automatically," 2019) and contains pre-defined fake profiles. The artifact will use fake profiles with less than ten attributes for volunteers to test the concept of falsifying personal data for privacy protection. The quantity and the content of the emails received will be evaluated. The conceptual model, see Figure 1, used to explain the process of collecting the data is based on the Blackbox software testing model.

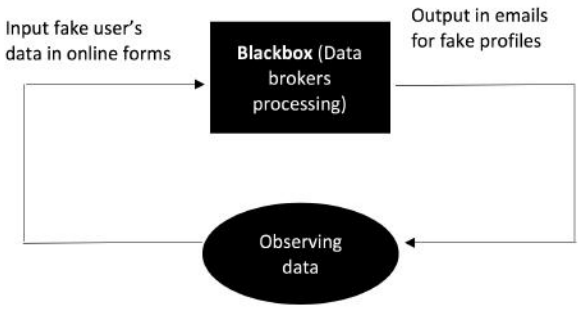

Figure 1 Concept of a Blackbox model

The study will be sequential: create a proof of concept tool, create fake profiles, recruit volunteers, run the experiment for 3-6 months, collect the data from the emails, analyze the data for common themes per profile, compare the results to the existing literature, and assess the participants' opinion of using such strategy for the future via survey. An example of fake profile elements could be a name, address, zip code, age & date of birth, gender, phone number, email address, etc. When a user is presented with a choice of filling out a form online, instead of real information, the user will use the artifact to automatically fill the form with predefined fake information corresponding to the fake identity profile. Every profile has a corresponding newly created real email which will function as a container receiving the targeted email. The data collected from emails will be text mined for common themes to evaluate if there is an effect of the use of fake information. The study will apply quantitative text analysis.

SUMMARY

It is very challenging to define what is considered private information. According to the Nissenbaum (2004) Contextual integrity theory, the context of the information is crucial. No single solution that will fit all exist because what is considered private for one individual may not be the same for others. Current technology to prevent personal data collection from web tracking application is based on blocking the app itself or part of its functionality. The privacy paradox renders such tools ineffective because the user inputs the data voluntarily when not presented with another choice. Currently, in the USA, there is no law addressing the scooping of personal data online and a technical solution to address the privacy paradox. The extent of data collection (private or not) and the use of it in the areas of healthcare, education, housing, and employment is disturbing. The students, unaware of such a practices, may become targets for deceptive cybersecurity attacks and opportunistic and even predatory marketing ploys (Oravec, 2017). It is very important to educate students to protect information infrastructure at their future work place and also keep the same practices when it comes to their private information. Empowering the user to decide what he/she consider private by using falsifying data approach could be one of the solutions.

Limitations and Future Work
The research will only address private online companies and exclude the law enforcement website, IRS, credit website, and similar. If personal data is falsified at such institutions, the act is considered fraud and Identity theft. The research will not present the whole cycle of the fake information due to lack of access to who the fake data is sold. A falsifying personal data approach will create a mixture of real (email) and fake data (profile) to poison the data stream collected by various browsers, applications, and other data collecting apps. The artifact can be expanded to create fake browsing data, a background-run app and do a fake online search based on specified profile parameter.

REFERENCES

2018 Cost of a Data Breach Study: Global Overview. (2018). Retrieved from IBM: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=55017055USEN&

Anthes, G. (2015). Data brokers are watching you. *Communications of the ACM, 58*(1), 28-30.

Baek, Y. M. (2014). Solving the privacy paradox: A counter-argument experimental approach. *Computers in Human Behavior, 38*, 33-42.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: a review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017-1042.

Blank, G., Bolsover, G., & Dubois, E. (2014). *A new privacy paradox: Young people and privacy on social network sites.* Paper presented at the Prepared for the Annual Meeting of the American Sociological Association.

Boyles, J. L., Smith, A., & Madden, M. (2012). Privacy and data management on mobile devices. *Pew Internet & American Life Project, 4*.

BRIEF-IMS Health and Quintiles to merge. (2016). *Market News.* Retrieved from https://www.reuters.com/article/idUSASC08M33

Brown, B. (2001). Studying the Internet experience. *HP Laboratories Technical Report HPL, 49*.

Bujlow, T., Carela-Español, V., Solé-Pareta, J., & Barlet-Ros, P. (2015). Web tracking: Mechanisms, implications, and defenses. *arXiv preprint arXiv:1507.07872*.

Cadwalladr, C., & Graham-Harrison, E. (2018). The Cambridge Analytica Files. *I made Steve Bannon's psychological warfare tool': meet the data war whistleblower*.

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2013). *Your browsing behavior for a big mac: Economics of personal information online.* Paper presented at the Proceedings of the 22nd international conference on World Wide Web.

Chen, K., & Rea Jr, A. I. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *Journal of Computer Information Systems, 44*(4), 85-92.

Choi, M., Levy, Y., & Hovav, A. (2013). *The role of user computer self-efficacy, cybersecurity countermeasures awareness, and cybersecurity skills influence on computer misuse.* Paper presented at the Proceedings of the Pre-International Conference of Information Systems (ICIS) SIGSEC–Workshop on Information Security and Privacy (WISP).

Christl, W., Kopp, K., & Riechert, P. U. J. C. L. (2017). Corporate surveillance in everyday life. 6.

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication, 15*(1), 83-108.

Egelman, S., Felt, A. P., & Wagner, D. (2013). Choice architecture and smartphone privacy: There's a price for that. In *The economics of information security and privacy* (pp. 211-236): Springer.

Elliott, C. (2009). Are online travel agencies quoting higher rates because of your Web cookies? Retrieved from https://www.elliott.org/blog/are-online-travel-agencies-quoting-higher-rates-because-of-your-web-cookies/

Ellison, N. B., Vitak, J., Steinfield, C., Gray, R., & Lampe, C. (2011). Negotiating privacy concerns and social capital needs in a social media environment. In *Privacy online* (pp. 19-32): Springer.

Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B.-G., Cox, L. P., . . . Sheth, A. N. (2014). TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS), 32*(2), 5.

Ephraim, P. E. J. M. M. R. C. A. o. M., Ethics, & Discourse, S. (2017). Whistleblowing and social responsibility in a surveillance system: Appraising the morality of the Snowden disclosures. 50.

FCC Adopts Privacy Rules To Give Broadband Consumers Increased Choice, Transparency And Security For Their Personal Data. (2016). [Press release]. Retrieved from https://www.fcc.gov/document/fcc-adopts-broadband-consumer-privacy-rules

Fill out forms automatically. (2019). Retrieved from https://support.google.com/chrome/answer/142893?co=GENIE.Platform%3DDesktop&hl=en

Flender, C., & Müller, G. (2012). *Type indeterminacy in privacy decisions: the privacy paradox revisited.* Paper presented at the International Symposium on Quantum Interaction.

Grauer, Y. (2018a). Here's a Long List of Data Broker Sites and How to Opt-Out of Them. *Big Data.* Retrieved from https://motherboard.vice.com/en_us/article/ne9b3z/how-to-get-off-data-broker-and-people-search-sites-pipl-spokeo

Grauer, Y. (2018b). What Are 'Data Brokers,' and Why Are They Scooping Up Information About You? Retrieved from https://motherboard.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 337-355.

Hann, I.-H., Hui, K.-L., Lee, S.-Y. T., & Png, I. P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of management information systems, 24*(2), 13-42.

Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday, 15*(8).

Heitzmann, R., Jackson, P., Oppenheim, C., & Toyens, V. (2011). Disconnect. Retrieved from https://disconnect.me/

Herrmann, D., & Lindemann, J. (2016). Obtaining personal data and asking for erasure: Do app vendors and website owners honour your privacy rights? *arXiv preprint arXiv:1602.01804*.

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *IEEE Security & Privacy, 3*(5), 22-25.

Iordanou, C., Soriente, C., Sirivianos, M., & Laoutaris, N. (2017). *Who is fiddling with prices?: Building and deploying a watchdog service for e-commerce.* Paper presented at the Proceedings of the Conference of the ACM Special Interest Group on Data Communication.

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer, 51*(8), 56-59.

Ismail, N. (2018). Fake Data is creeping up on companies worldwide. *Data Analytics & Data Science.* Retrieved from https://www.information-age.com/fake-data-creeping-companies-worldwide-123470687/

Kindy, K. (2017). How Congress dismantled federal Internet privacy rules. *Politics.* Retrieved from https://www.washingtonpost.com/politics/how-congress-dismantled-federal-internet-privacy-rules/2017/05/29/7ad06e14-2f5b-11e7-8674-437ddb6e813e_story.html?utm_term=.fb30a0074651&wpisrc=nl_evening&wpmm=1

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64*, 122-134.

Kuner, C., Cate, F. H., Millard, C., & Svantesson, D. J. B. (2012). The challenge of 'big data' for data protection. *International Data Privacy Law, 2*(2), 47-49. Retrieved from https://dx.doi.org/10.1093/idpl/ips003.

Lazzarotti, J. J., & Gavejian, J. C. (2019). Privacy and Cybersecurity Issues to Watch in 2019. Retrieved from https://www.natlawreview.com/article/privacy-and-cybersecurity-issues-to-watch-2019

Lécuyer, M., Ducoffe, G., Lan, F., Papancea, A., Petsios, T., Spahn, R., . . . Geambasu, R. (2014). *Xray: Enhancing the web's transparency with differential correlation.* Paper presented at the 23rd {USENIX} Security Symposium ({USENIX} Security 14).

Lee, H., Park, H., & Kim, J. (2013). Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. *International Journal of Human-Computer Studies, 71*(9), 862-877.

Li, T.-C., Hang, H., Faloutsos, M., & Efstathopoulos, P. (2015). *Trackadvisor: Taking back browsing privacy from third-party trackers.* Paper presented at the International Conference on Passive and Active Network Measurement.

Lobosco, K. (2013). Facebook friends could change your credit score. Retrieved from https://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html?hpt=hp_t2

Lwin, M., Wirtz, J., & Williams, J. D. (2007). Consumer online privacy concerns and responses: a power–responsibility equilibrium perspective. *Journal of the Academy of Marketing Science, 35*(4), 572-585.

Lwin, M. O., & Williams, J. D. J. M. L. (2003). A Model Integrating the Multidimensional Developmental Theory of Privacy and Theory of Planned Behavior to Examine Fabrication of Information Online. *Marketing Letters, 14*(4), 257-272. Retrieved from https://doi.org/10.1023/B:MARK.0000012471.31858.e5. doi:10.1023/B:MARK.0000012471.31858.e5

Malheiros, M., Preibusch, S., & Sasse, M. A. (2013). *"Fairly truthful": The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure.* Paper presented at the International Conference on Trust and Trustworthy Computing.

Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., & Guo, S. (2016). Protection of big data privacy. *IEEE access, 4*, 1821-1834.

Metzger, M. J. (2007). Communication privacy management in electronic commerce. *Journal of computer-mediated communication, 12*(2), 335-361.

Milne, G. R., & Boza, M.-E. (1999). Trust and concern in consumers' perceptions of marketing information management practices. *Journal of interactive Marketing, 13*(1), 5-24.

Miltgen, C. L., & Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. *European Journal of Information Systems, 23*(2), 103-125.

Miltgen, C. L., & Smith, H. J. (2018). Falsifying and withholding: exploring individuals' contextual privacy-related decision-making. *Information & management*.

Mothersbaugh, D. L., Foxx, W. K., Beatty, S. E., & Wang, S. (2012). Disclosure antecedents in an online service context: The role of sensitivity of information. *Journal of service research, 15*(1), 76-98.

Moura, J., & Serrão, C. (2015). Security and privacy issues of big data. In *Handbook of research on trends and future directions in big data and web intelligence* (pp. 20-52): IGI Global.

Munir, A. B., Yasin, M., Hajar, S., & Muhammad-Sukki, F. (2015). Big data: big challenges to privacy and data protection. *International Scholarly and Scientific Research & Innovation, 9*(1).

NG, A. (2018). Google, Facebook and Amazon spending more than ever lobbying Congress. *News:Politics.* Retrieved from https://www.cnet.com/news/google-facebook-amazon-spending-more-than-ever-lobbying-congress/

Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev., 79*, 119.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs, 41*(1), 100-126.

Olson, M. (2011). Research Study: Consumer Perceptions of Online Registration and Social Sign-in. Retrieved from https://www.janrain.com/blog/research-study-consumer-perceptions-online-registration-and-social-sign

Oravec, J. A. (2017). *Emerging "cyber hygiene" practices for the Internet of Things (IoT): Professional issues in consulting clients and educating users on IoT privacy and security.* Paper presented at the Professional Communication Conference (ProComm), 2017 IEEE International.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems, 24*(3), 45-77.

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*. Albany, NY, US: State University of New York Press.

Petronio, S. (2006). Communication Boundary Management: A Theoretical Model of Managing Disclosure of Private Information between Marital Couples. *Communication Theory, 1*(4), 311-335. Retrieved from https://dx.doi.org/10.1111/j.1468-2885.1991.tb00023.x. doi:10.1111/j.1468-2885.1991.tb00023.x

Polonetsky, J., & Tene, O. (2013). Privacy and big data: making ends meet. *Stan. L. Rev. Online, 66*, 25.

The Privacy Legal Implications of Big Data: A Primer. (2013). Retrieved from https://www.infolawgroup.com/blog/2013/02/articles/big-data/the-privacy-legal-implications-of-big-data-a-primer

Punj, G. (2017). Consumer intentions to falsify personal information online: unethical or justifiable? *Journal of Marketing Management, 33*(15-16), 1402-1412.

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. (2013). Anonymity, privacy, and security online. *Pew Research Center, 5*.

Ramirez, E., Brill, J., Ohlhausen, M. K., Wright, J. D., & McSweeny, T. (2014). *Data brokers: A call for transparency and accountability*. Retrieved from Washington, DC: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Rao, A., Kakhki, A. M., Razaghpanah, A., Tang, A., Wang, S., Sherry, J., . . . Mislove, A. (2013). Using the middle to meddle with mobile. *CCIS, Northeastern University, Tech. Rep., December*.

{Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)}. (2016). *Official Journal of the European Union, L119*, 1-88. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC.

Romm, T. (2018). Google led a multimillion-dollar tech industry lobbying blitz in 2018, records show. *The Switch.* Retrieved from https://www.washingtonpost.com/technology/2019/01/23/google-led-multimillion-dollar-tech-industry-lobbying-blitz-records-show/?noredirect=on&utm_term=.89a6be349518

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *NYUL rev., 86*, 1814.

Schwartz, P. M., & Solove, D. J. (2012). Pii 2.0: Privacy and a new approach to personal information. *Privacy and Security Law Report*.

Sheehan, K. B., & Hoy, M. G. (2000). Dimensions of privacy concern among online consumers. *Journal of public policy & marketing, 19*(1), 62-73.

Simon, M. (2012). Is Orbitz steering Mac users toward pricier hotels? Retrieved from https://www.cnn.com/2012/06/26/tech/web/orbitz-mac-users/index.html

Singer, N. (2013). A Data Broker Offers a Peek Behind the Curtain. Retrieved from https://www.nytimes.com/2013/09/01/business/a-data-broker-offers-a-peek-behind-the-curtain.html?_r=0

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS quarterly, 35*(4), 989-1016.

Stutzman, F., Vitak, J., Ellison, N. B., Gray, R., & Lampe, C. (2012). *Privacy in interaction: Exploring disclosure and social capital in Facebook.* Paper presented at the Sixth International AAAI Conference on Weblogs and Social Media.

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco), 671*, 1-34.

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property, 11*(5), xxvii.

Turner, A. (2016). How Data Brokers Make Money Off Your Medical Records. *Scientific American, 314*(2), 26-27. Retrieved from https://www.scientificamerican.com/article/how-data-brokers-make-money-off-your-medical-records/. doi:doi:10.1038/scientificamerican0216-26

Vaishnavi, V., & Kuechler, W. (2004). Design sciences research in information systems. Retrieved from http://www.desrist.org/design-research-in-information-systems/

Van den Hoven, J., Blaauw, M., Pieters, W., & Warnier, M. (2014). Privacy and information technology.

van der Sloot, B., & van Schendel, S. (2016). Ten questions for future regulation of big data: A comparative and empirical legal study. *Journal of Intellectual Property Information Technology and Electronic Commerce Law, 7*, 110.

Vitell, S. J. (2003). Consumer ethics research: Review, synthesis and suggestions for the future. *Journal of business ethics, 43*(1-2), 33-47.

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 193-220.

Wirtz, J., & Lwin, M. O. (2009). Regulatory focus theory, trust, and privacy concern. *Journal of Service Research, 12*(2), 190-207.

Xu, L., Jiang, C., Wang, J., Yuan, J., & Ren, Y. (2014). Information security in big data: privacy and data mining. *IEEE access, 2*, 1149-1176.

Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication & Society, 16*(4), 479-500.