Chapter

Risks of Privacy-Enhancing Technologies: Complexity and Implications of Differential Privacy in the Context of Cybercrime

William Stadler

Abstract

In recent years, the swift expansion of technology-enabled data harvesting has infiltrated modern life and led to the collection of massive amounts of private data. As a result, the preservation of individual privacy has become a salient concern for the general public. Combined with an increase in the frequency and prevalence of cybercrime, more of the public now face the very real risk of privacy loss associated with illegitimate use of private data. Differential Privacy has emerged as a relatively new privacy-preserving method with the potential to significantly reduce the likelihood of harmful data disclosures stemming from malicious use. However, research has not explicitly investigated Differential Privacy from the perspective of criminal justice or examined the utility of Differential Privacy as a possible situational crime prevention measure to cybercrime. Therefore, this chapter explores the proliferation of cybercrime through advances in technology and briefly examines other privacy-preserving methods before discussing the possible use of Differential Privacy as a viable countermeasure to cybercrime. The chapter concludes with a discussion of several practical considerations related to the use of Differential Privacy as a tool in the fight against cybercrime and offers recommendations for future research.

Keywords: cybercrime, Differential Privacy, privacy-enhancing technologies, technology-enabled crime, situational crime prevention

1. Introduction

The production and consumption of data has been increasing with the ubiquity of the internet [1], and with this, the benefits that accompany innovations and advances in computing technology, such as those stemming from artificial intelligence and machine learning, are increasingly relevant to a growing number of industries and applications [2]. However, our reliance on technology and consumer connectedness, coupled with rapid growth in the aggregation and liquidity of personalized data, has made us more vulnerable to cybercrime victimization and the malicious use of private data [3, 4]. The challenge of securing confidential information is becoming one of the key issues in our digital world [5].

Recently developed privacy-enhancing technologies and methods are being touted as possible solutions to mitigate privacy risks associated with inadvertent disclosure and guard against sinister data incursions resulting from cybercrime. One such possibility is Differential Privacy [6], which represents a new security paradigm designed to meet the growing number of privacy risks which accompany data stewardship, particularly for those entrusted with safeguarding data. Differential Privacy was conceived to simultaneously harness the power of information contained in "big data" while substantially reducing the likelihood of harmful data disclosures resulting in possible malicious use [7].

The commercial benefits and costs of privacy enhancing technologies have been widely studied, particularly as consumer data sharing and consumption has grown through distributed systems and Internet of Things (IOT) devices and applications such as smartphones, televisions, medical equipment, appliances, and wearables. However, because of its emergence as a promising new approach to computational analysis, far less has been written about the implications of Differential Privacy, including the merits and limitations of the sophisticated techniques created in the context of this definition. Similarly, research aimed at the advantages, pitfalls, and practical challenges of adopting differentially private approaches has been limited. Literature on Differential Privacy has yet to explore the applied use of this privacy-preserving approach in the context of contemporary crime and justice threats, including cybercrime. Scholarship has generally tended to avoid important, and arguably necessary, cross-disciplinary collaborations between technical science disciplines such as computer science and social science disciplines like criminal justice.

Therefore, through the lens of the criminal justice discipline, this chapter will explore the use of Differential Privacy as a possible cybercrime prevention technique in the context of the massive digital ecosystem that has emerged over the last two decades. We begin with a discussion of the recent proliferation of cybercrime that has arisen through advances in technology, followed by a brief examination of evolving privacy protections which led to the rise of differential privacy, as both a general tenet and assortment of techniques for advancing data security. We then speculate on the use of Differential Privacy as a situational crime prevention countermeasure to cybercrime, and review potential challenges to its use. The chapter concludes with an attempt to stimulate future research and interest in cross-disciplinary exploration of this relatively new privacy-enhancing approach, particularly with respect to its potential to reduce risk, combat crime, and preserve the confidentiality of data for consumers and those most vulnerable to cybercrime victimization.

2. The proliferation of cybercrime through technology

As the general public engages more with online environments and participation in connected routines that produce personal data becomes more common to everyday life, new criminal opportunities emerge in the form of cybercrime [8]. Though the concept of cybercrime is open to interpretation and has resulted in several competing definitions, broadly defined, cybercrime involves technology-related offending that takes place in the online environment [9] and is "committed using a computer, network, or hardware device" [10]. More importantly, cybercrime represents a serious economic and national security threat to the United States and to other countries around the world [11, 12]. Research has revealed that theft of private data through cybercrime is continuing to grow [1], resulting in a substantial need for promising new definitions and approaches, as well as new laws [13], aimed at the protection of personal data and individual privacy. Differential Privacy is one

of many approaches with the potential to prevent or significantly blunt the harmful consequences associated with cyber criminality by influencing the means through which organizations and agencies protect sensitive information from exploitation and malicious use. Yet, cybercrime itself has largely remained on the periphery of the criminology discipline as a marginalized topic [3], and research on information security in the context of cybercrime has remained limited as a result, perhaps because of the complexities associated with the crimes and spatial and temporal distance between offenders and victims [13]. Further, research in crime and justice literature on both the theoretical and practical use of technical privacy methods, such as Differential Privacy, is virtually non-existent.

Meanwhile, the spread of data-driven technologies are generating a multitude of ways for public and private-sector entities to induce the creation and dissemination of personal data which also inadvertently enables cybercriminals access to information that people would rather keep to themselves. Ironically, the recent trend toward distributed computing and the decentralization of control and access to smaller computer systems and network resources has also increased the likelihood of cybercrime [14]. Once data have been generated and exist somewhere, the malicious use of that data becomes more likely, creating greater potential for victimization and harm to individuals and to organizations alike. Thus, two related issues become tantamount when considering the practical utility of Differential Privacy as one of many possible countermeasures to cybercrime. First, it is important to understand how cybercrime threats are evolving and expanding to ensure that subsequent prevention and interdiction measures are designed with specific cybercrime threats in mind. Second, it is also necessary to consider how detection and attribution capabilities have evolved in relation to the changing threat landscape so that cybercrime enforcement and investigation methods also meet changing demands.

2.1 Threat expansion and evolution

The world community has been increasingly expressing concern about the use of advanced computing and AI for criminal purposes [15]. And in recent years, advances in technology have undoubtedly increased the frequency and prevalence of cybercrime activity, resulting in an expansion of possible threats to systems and data worldwide [13]. Given the breadth of information captured and widely available today about each individual on earth, people might assume that the magnitude of the internet and related "systems" as well as volume of data being transmitted provides adequate protection against disclosures of personal data. Individuals sharing this view may also conclude that the odds of becoming a victim are low and that more robust technical countermeasures to cybercrime are unnecessary. However, this perception is a fallacy; vulnerability to victimization is not uniformly distributed, nor are contemporary acts of cybercrime targeted only at single persons or entities. The size and scale of cybercrime capabilities and efforts has increased commensurate with advances to computing power and precision, perhaps resulting in modern cyber-predators posing greater risk to larger groups of individuals than ever before [15].

This fact is becoming more evident as the United States and other countries around the world grapple with increasingly serious cases of cybercrime which strain the integrity of data protection measures in both public and private sectors. Dozens of high-profile and illegal data breaches have occurred in the U.S. over the last handful of years that resulted in the compromise or theft of massive amounts of private information, including with eBay [16], JP Morgan Chase [17], Sony [18], Adobe, Equifax, and LinkedIn [19], as well as with U.S. political organizations [20] and voter registration records [21]. Highly sophisticated gangs, organized crime

groups, and terrorist organizations are also using computer and communication technologies to steal, smuggle, blackmail, sell drugs, and conduct a variety of other criminal activities on a much larger scale to finance their operations [22]. To be sure, cybercriminals are becoming more knowledgeable and skilled, and they appear to be systematically attacking larger and more sensitive databases with increasing brazenness and alarming frequency.

Recent advances in privacy technology have to some degree equipped data guardians with more tools to systematically prevent inadvertent data disclosures resulting from legitimate use. With respect to cybercrime, the contribution of new innovations has also enabled private corporations and government agencies, including those serving prevention, enforcement, or regulatory functions, to better deter, investigate, and detect instances of nefarious activity and cybercrime attacks resulting in privacy fissures. Yet, on the whole, governments and private entities frequently appear to be playing catch-up. Growth of distributed systems, AI, and novel privacy enhancing technologies which strengthen the capabilities of data producers and distributors have also produced unintended consequences, including conditions favorable to hostile actors gaining the motivation, means, and cover to access private information and conceal malicious activity [23]. Moreover, typical privacy protections have achieved limited success because they are inattentive to the opportunistic aspects of cybercrime [14]. Commonly deployed data protection tactics may generate a false sense of security while inadvertently softening crime targets by making them more attractive, accessible, and unguarded to allow cyber-criminals opportunities to conceivably initiate attacks on private information more easily. The resulting "target softening" stems directly from the shift toward complex software, interconnected data networks, and distributed systems in the modern IoT infrastructure which remain inadequately guarded and vulnerable to penetration via more sophisticated techniques [5]. While innovations and capabilities advancements undoubtedly enable more sophisticated applications, they also enable adversaries to collect information and deliver exploits specifically tailored to target systems [24].

The frequency of hostile attacks will also likely increase as artificial intelligence capabilities become more powerful and widespread, evolving and expanding the very nature of existing cybercrime threats while simultaneously spawning new threats. Indeed, there is reason to expect that intrusions enabled by the growing use of AI among cybercriminals will be finely targeted at the complex vulnerabilities created by AI systems and become more effective at exploiting the weaknesses left in their wake [15]. The emergence of machine learning algorithms, in particular, has effectively boosted adversary capabilities to run complex and repeatable problem-solving operations against unfortified positions without human intervention, providing cybercriminals with technical scalability and automation which has historically been beyond their reach. The ability of cybercriminals to more intelligently and systematically assault numerous targets at once will likely exacerbate an already challenging problem facing cyber security practitioners in which criminals must only find one flaw in a vast system, whereas database and systems administrators must account for all possible weaknesses to protect system integrity [25]. Even the most inept cyber-criminal need only exploit a single path of vulnerability among the complex and increasing number of data ingestion points, whereas data guardians face the increasingly difficult task of protecting against all conceivable threats to privacy [26].

2.2 Threat detection and attribution

While cybercrime offenses against privacy may in some ways be synonymous with traditional non-violent "street" crimes, such as those against property, because

they involve the theft, corruption, or destruction of assets held and valued by a property owner, there may be a tendency to address them like ordinary crimes. However, the nature of technology-based privacy crimes varies in several important ways. Chief among these is the fact that cybercrimes often carry an inherently lower risk of detection, due to significant spatial separation and temporal distance between offenders and victims. Additionally, privacy-related offenses may also be obscured due to their velocity, automation, and complexity [27]. Thus, the adoption of new computing innovations and methods, such as machine learning, by cybercriminals will likely continue to challenge existing cybercrime detection and attribution methods. In particular, cyber-assaults against distributed systems may be of such increasing scale and complexity that forensic detection and attribution efforts will suffer markedly. Research has already shown cybercriminals to be savvy, having migrated away from easily detectable attacks that were recently commonplace toward more stealthy aggressions that are often indistinguishable [24].

For similar reasons, cybercrime threats will presumably expand and diversify as a natural byproduct of the automation computing innovations have permitted. In this regard, human capital costs of cybercriminals attempting intrusions into databases containing personal information are likely to decline as they leverage the scalable use of AI systems to complete tasks that would ordinarily require extensive human labor, intelligence and expertise. Those cost savings might naturally translate into expanding the pool of actors with which to initiate attacks, increasing the rate at which attacks are carried out, and growing the set of prospective targets. Thus, the acquisition of AI capabilities among cybercriminals will expand their operations to spawn new attacks that would be otherwise impractical for humans. Malicious actors will purposely target and exploit the growing multitude of vulnerabilities of AI systems deployed by those entrusted with stewardship and fortification of data, thereby deepening the threat to the privacy of individuals represented in such data.

3. Evolving privacy methods

While the influence and intrusion of technology into the public sphere has unintentionally created new opportunities for cyber victimization, various approaches to counter emerging threats have developed and evolved out of privacy requirements engineering. These methods have enabled the design, analysis, and integration of security and privacy requirements during systems implementation for traditional and cloud architectures to better support and protect data [28]. Further, novel privacy definitions have been created, resulting in several systematic approaches to minimize the likelihood of unintended data disclosures. Differential Privacy represents one of the newest, and perhaps most promising, privacy definitions aimed at preserving the privacy of individuals and groups whose data is published and/or accessible for public- and private-sector research and data analysis, as well as product and service development and enhancement. Yet a variety of other techniques continue to persist.

3.1 Prior anonymization techniques

As the scale of consumable data generated by society has grown, so too have the mechanisms for shielding the information and individuals represented in such data. Historically, curators of large databases attempted to protect individual privacy through the de-identification of datasets using a variety of algorithmic data anonymization techniques. These have included stripping or suppressing identifying

information such as names, dates of birth, and other personal information out of data that is released for consumption, or through replacement of some data values with generalized quasi-identifiers. In effect, the data elements generated from these processes have represented approximations of data or a broad category of values to achieve the property of "k-anonymity"—anonymization resulting from data that is indistinguishable from that produced by another individual in the same dataset [29]. Through these practices, curators reasonably believed anonymity could be assured—that personal identifiable information (PII) contained within the data could not be distinguished or used to discover the identity of individuals or groups of individuals represented in the data [30]. However, we now know that these earlier methods for protecting individual privacy have been afflicted with vulnerabilities, resulting in "de-identified" datasets being prone to exploitation or attack, particularly where the value of sensitive attributes is not diverse enough or when sufficient background knowledge is known by would-be attackers [31]. In such circumstances, individuals might face unintentional risk of cybercrime victimization and identification resulting from inference attacks and algorithms deployed against databases to reconstruct case-specific identities through whatever limited, sensitive data is contained in a given database, or through the fusion of extracted data with external sources [32].

Numerous examples have been cited where de-identified data published for legitimate use was nevertheless systematically exploited to uncover individual identities (see [33–35]). Though some privacy breaches may not involve nefarious intent and therefore result in relatively benign consequences, the growing number of intentionally harmful and illegal privacy intrusions should elicit concern among privacy advocates and information security practitioners. Further, subsequent research has also revealed that not all k-anonymity algorithms provide uniform, privacy-preserving protections [36] and that some can inadvertently distort data to a point where both its integrity and utility are appreciably diminished [37]. Thus, it is clear that prior efforts to counter privacy risks have not gone far enough. While more recent techniques such as l-diversity and t-closeness have incrementally advanced the security of personal-level data, they may also be vulnerable to exploitation as the liquidity of data and proliferation of artificial intelligence in today's contemporary world continue to advance [38, 39]. Yet, despite these notable concerns, many of the deficient database de-identification techniques referenced above, which fail to truly anonymize participants and protect their confidentiality, continue to persist as commonplace practices in commercial industries and the larger research community [34].

3.2 Emergence of Differential Privacy

Recognizing the need for a more robust privacy approach, Differential Privacy was developed in the early 2000s. While it was not explicitly intended to guard against cybercrime, Differential Privacy represents a deliberate attempt to overcome many of the foreseeable privacy challenges identified above by seeking true anonymity in datasets. With this definition and the use of differentially private processes, personal information can, in theory, be more adequately protected from cybercrime activity by avoiding the availability or release of raw data and instead enabling a replica database upon which queries are run containing modified (but statistically similar) versions of person-level data. Thus, Differential Privacy represents an enhanced level of privacy protection in the evolving data security model, resulting in virtually no disclosure risk. It achieves this by obscuring individual identities with the addition of mathematical "noise" to particular data elements, consequently concealing a small sample of each individual's data [40]. According to

its proponents, Differential Privacy virtually guarantees that the removal or addition of a single database item does not appreciably affect the outcome or validity of any analysis. Stated another way, this data perturbation technique ensures that the probability of a statistical query producing a given result is virtually the same whether it is conducted on an unadulterated dataset or one containing modified or synthetic data [40]. Thus, the true benefit to Differential Privacy is that there is quantifiably lower risk associated with its use over alternative methods aimed at systematically safeguarding personal data. In turn, individuals' data should be more rigorously defended from theft or illegitimate use when differentially private methods are used.

Because Differential Privacy was conceived as a more rigorous definition of anonymizing data and protecting confidentiality than prior methods, its popularity has grown in recent years, with several commercial entities enabling Differential Privacy algorithms for use on a massive scale for data generated in the private sector. For example, Apple has intentionally deployed Differential Privacy techniques to discover and analyze usage patterns of large numbers of iPhone users without compromising the privacy of individuals [40]. In this instance, Differential Privacy algorithms executed by Apple analyze iOS user data with the published goal of improving and enhancing end-user experiences with various iOS applications such as iMessage (text messaging), through which functions such as auto-correct, suggested words and phrases, and emojis can become more intuitive [41]. In a similar example of commercial use, Google has employed Differential Privacy algorithms in its analyses of Chrome web-browser usage to discover the prevalence of malicious software hijacking computer and application settings without user knowledge [42].

There has even been expanded use of Differential Privacy in the public sphere, with the U.S. Census Bureau recently announcing its plan to more rigorously protect the confidentiality of individual-level data than in years past. Prior to the most recent census, this federal agency attempted to obscure person-level information by substituting raw data beneath the census block level with comparable data from another block to ensure the validity of population-level statistics. However, beginning with the 2020 Census, "noise" will be purposely injected into all data emanating below the state geographic level [43] to achieve "advanced disclosure protections" [44]. This instance of Differential Privacy use represents one of the first by a federal agency broadly responsible for the collection and provision of data for public use, and is likely to serve as a possible model for other federal, state, and local data stewards.

Given its intent, generally positive reviews, and notable use in a handful of public and private sector instances, it is somewhat remarkable that Differential Privacy has failed to gain widespread adoption as a data protection measure since its introduction in 2006. Though Differential Privacy has indeed become an information security standard with database computation and analysis in computer science research, resulting in numerous algorithms aimed at strengthening privacy, practitioner adoption of Differential Privacy in applied settings has been slow to gain traction [45]. Similarly, while Differential Privacy has indeed spawned important new lines of data privacy research, much of that work has been theoretical or simulated and proven to be less suitable for application to real-world situations [4]. To date there have been few empirical examinations of the practical application of Differential Privacy, despite the existence of important concerns surrounding its viability, including possible tradeoffs that arise between achieving heightened privacy protections and preserving the utility of data produced through differentially private queries [46]. Despite the obvious and substantial lag between the emergence of Differential Privacy as a definition worthy of research and its acceptance as a pragmatic and commonly employed approach in real-world scenarios, it is

important to consider the relevance and utility of Differential Privacy as a possible cybercrime countermeasure in anticipation that its use will someday become pervasive.

4. Cybercrime prevention through Differential Privacy

Though Differential Privacy is applicable to a number of industries and scenarios, its potential as a cybercrime prevention and risk mitigation measure is intriguing and warrants deeper exploration. From a criminological standpoint, differentially private approaches might best be deployed as technical situational crime prevention (SCP) measures to deter prospective attacks against sensitive data, or at the very least, minimize their harms. Generally speaking, situational crime prevention represents a data-driven approach to reduce the physical opportunities for crime by concentrating on the specific conditions, settings, and circumstances which produce the conditions favorable to criminality [47]. Further, the approach explicitly suggests that crime prevention can only be accomplished by systematically analyzing the details of a given crime problem and then introducing strategies for blocking, reducing or removing the opportunities that enable a particular crime to take place [14]. Thus, the most viable strategy to combat crime is through the informed management, design, and manipulation of a particular environment that would ordinarily be conducive to crime [48]. While SCP has mostly been utilized to examine and respond to traditional forms of criminality, such as burglary, robbery, and theft, it has direct applicability to cybercrime, given that acts of cybercrime share many similarities with property crimes. By examining important contextual attributes associated with specific cybercrime events, such as the technical means and steps through which an attack on data may be committed and how a database containing private information may be made less attractive or be better protected, cybersecurity practitioners can develop competent proactive strategies to reduce the presence and attractiveness of criminal possibilities for would-be offenders [14].

Situational Crime Prevention efforts are generally intended to achieve three goals: increase the overall risk to criminals, increase the effort they would be required to expend to engage in a crime, and decrease the reward associated with an act of crime [49]. In practice, exploration of a given network or computerized system through the perspective of situational crime prevention might first enable the identification of various targets that represent higher-value for cybercriminals. In turn, those high-value targets would be the first and most likely to receive heightened privacy protections. For example, databases that contain sensitive information about individuals or groups which, if disclosed, might hold potential monetary value and likely result in physical or financial harm, would be ideal candidates for Differential Privacy protections. Once identified, possible cybercrime targets might be "hardened" and made less attractive through the intentional adulteration of data in an effort to obscure personal information. The intent of this tactic would be to reduce the likelihood of an attack, because the risk and effort for a cybercriminal initiating an assault on that target would be considerably greater than in situations where differentially private techniques are not used.

The act of safeguarding data clearly carries direct costs for data stewards and information security practitioners, but attacks against data also carry similar costs for the attacker, both in terms of the resources required to mount an attack and potential costs if an attack is detected and subsequently punished. Unless the expected return from an attack is greater than the risk-adjusted costs of the attack, the attack will be uneconomical and become a less attractive target for an offender. Thus, the injection of noise into an otherwise high-value, sensitive dataset through

Differential Privacy algorithms might ensure that attackers would have to work harder and still be unable to derive much, if any, value from stolen data, despite the data still remaining useful for legitimate purposes. As a data perturbation method, Differential Privacy stands to more securely protect the privacy of individuals and appreciably diminish the utility of the entire corpus of stolen data, thus negating an attacker's reward motive. With advance knowledge of the use of differentially private techniques on high-value databases, cybercriminals might altogether be deterred from exerting the effort to wage an attack, given the minimal value of the data relative to the cost of waging such an attack.

5. Practical challenges

Despite confidence in Differential Privacy as a promising new tool in the war against cybercrime, it is not a panacea. A number of practical concerns remain that may slow the adoption of this approach in the near-term and challenge its use as a viable cybercrime countermeasure. Each of the following challenges should be examined more thoroughly to guide future decision-making for the use of Differential Privacy in real-world settings. Chief among these concerns are the trade-offs that accompany the use of Differential Privacy, specifically, where the costs associated with using differentially private methods are balanced against the benefits of doing so. Second, while the likelihood of privacy intrusions originating external to a given system might fall with the use of Differential Privacy, there is a possible shift in risk from external to internal threats that is likely to accompany the use of Differential Privacy in a variety of applied settings. Similarly, as use of Differential Privacy grows, adversaries will also be increasingly more likely to take advantage of advances in computing power, launching a virtual "arms race" between cybercriminals and those responsible for protecting sensitive data. Lastly, but perhaps most limiting for the use of Differential Privacy, particularly in crime and justice settings, there remains a very real concern about the practical challenge of resourcing the skilled human capital needed to develop, enable, and continually support Differential Privacy techniques.

5.1 Tradeoffs

An important implication of Differential Privacy is that its use results in two significant tradeoffs that should be factored into decisions regarding whether, when, and how to use the method. In the first tradeoff, the validity or accuracy of a given set of data may be reduced with a corresponding attempt to increase privacy. For example, the near-guarantee of total anonymity in a dataset can only be attained at some proportional reduction to the utility of that dataset. This challenge is commonly referred to as the "privacy budget" [50]. In this regard, tipping the scales in favor of greater privacy protections by injecting noise into data will provide a clear privacy benefit to the individuals whose personal information is contained in a given database. However, the adulteration of data resulting from a differentially private technique may also unintentionally produce imprecise statistical measures of a given phenomenon and lead to invalid conclusions derived from analysis of the data. The risk associated with this situation is that conclusions drawn from adulterated data under legitimate use scenarios, either by researchers or practitioners, might be faulty, because they are based on inaccurate data.

One cautionary example of this challenge is a pharmacogenetic study conducted by Fredrikson et al. [50]. The research evaluated the clinical effectiveness of a commonly prescribed blood-thinner using machine-learning models, while

differentially private algorithms were enabled to significantly reduce privacy risk for study participants. While the study yielded success in appreciably reducing privacy risk for study participants, according to the data, that success came at an increased risk of patient adverse health events and mortality. Though the study itself was simulated to examine the impact of Differential Privacy on a real-world clinical situation, the possible implications are clear; using differentially private algorithms to produce synthetic data may lessen privacy risks, but consequently result in a variety of unintended consequences to the conclusions of research, or in a worst-case scenario, to the same people Differential Privacy is meant to protect.

In addition to the tradeoff concern relating to the privacy budget, Differential Privacy also requires a tradeoff between the costs of deploying the privacy protections and the relative value of the data assets being protected. The values of data assets differ widely. Some targets might contain high-value, sensitive information, such as personal identifiers, credit card information, passwords, social security numbers, and insurance information that can be used maliciously to steal an identity or file false Medicare claims. Cybercriminals would likely view these targets as attractive and initiate attacks against the databases to steal such information. Therefore, databases containing highly sensitive data need extremely high-assurance protections. Other targets may contain personal data but of a less sensitive variety, including Netflix subscriptions, personal shopping preferences, search term use, or website visits. The value of these data may have lower transactional value for cybercriminals looking to exploit personal information. Thus, datasets containing these sources of information would presumably require weaker assurance protections.

A scenario where both high-and low-value assets are guarded requires that hazard-based decisions be made about the effort devoted to protecting each set of assets from cybercriminals. For example, security practitioners should explore what must be done to sufficiently protect high-assurance assets from possible intrusion, and what minimum level of effort would be required to protect low-assurance assets. Treating low-assurance assets the same as high-value would lead to the irrational use of resources. Therefore, practitioners should carefully consider tradeoffs to the privacy budget and efforts required to protect assets when choosing to implement differentially private approaches.

5.2 Shifting risk and the impending arms race

While the adoption of Differential Privacy techniques may provably strengthen defenses against traditional cybercrime threats directed at the theft of personal information from a database, their use may also coincide with a sizeable shift in where risks originate and how they evolve. For instance, there is already mounting concern among researchers and practitioners that new innovations and technology advances will transform the very nature of systems integrity and vulnerability, particularly with the growth of artificial intelligence, which will result in a "double-barreled threat" to high-value data repositories [14, 51]. In the traditional cybercrime model, criminal threats are generally thought to arise from an external source, spatially distant from the data being protected. However, internal threats to systems and data are now garnering additional attention, as cybercrime attacks are being more frequently initiated by organizational insiders [52]. The growing likelihood and simultaneous nature of these dual threats significantly increases the effort necessary to keep an infrastructure and its data secure, which will represent a significant ongoing challenge for many industries and organizations already struggling to provide robust information security [51].

Further, as Differential Privacy continues its incremental expansion beyond the realm of research toward use in applied settings, the resources and costs required for enabling Differential Privacy and other sophisticated privacy protections will also evolve. So too will the costs for cybercriminals intent on defeating the stronger protections afforded by differentially private systems. Cybercriminals are already taking advantage of more powerful computational resources and sophisticated approaches, requiring the investment of data guardians to continue increasing proportionally to keep pace. As a result of the commodification of computing technology, there is a brewing cybercrime "arms race" where information security practitioners will be constantly expected to respond in tit-for-tat fashion to complex and powerful threats from hostile actors [53]. As a result, to avoid having information security devolve into a never-ending game of "whack-a-mole" to combat emerging threats, individuals responsible for data security policy and practice must develop comprehensive strategies for data management and the use of privacy-preserving tools like Differential Privacy. However, the creation of such policies requires careful consideration of the origin and nature of threats to the data for which organizations have responsibility.

5.3 Resource constraints

Finally, and despite its potential as an automated method of systematically safeguarding data, Differential Privacy, much like artificial intelligence (see [54]), will only be as useful as the skilled humans that enable and support it. Unfortunately, some of the most pressing information security concerns facing a majority of organizations today include the limited number of skilled security personnel employed and the number who are readily available for employment [54]. While Differential Privacy strategies offer the realistic promise of protecting data for organizations that cater to consumers, significant barriers to the implementation and use of advanced privacy-enhancing technologies remain for organizations and agencies in the public sector that curate data for the most vulnerable populations, such as patients, prisoners, the disabled, and juveniles. Differential Privacy use to date has taken place primarily in the private sector, within organizations that have the financial and intellectual resources to pursue novel and costly privacy protections. However, research suggests that federal agencies do not have the relevant expertise or resourcing to implement differential privacy for the data they curate [55]. This is evident in the fact that to this point the U.S. Census Bureau is the only federal agency known to have initiated a systematic effort to employ Differential Privacy with the data it curates. The increasing sophistication of prospective cybercriminals and growing complexity of privacy enhancing technologies, including Differential Privacy algorithms needed to protect sensitive data, requires a level of data security expertise and sophistication that is simply not readily available throughout the federal public sector. In turn, this limitation is likely to be amplified at the state and local agency level, where funding for and expertise in skilled information security personnel are even more severely restricted than with the federal government.

Though expertise and a skilled labor force will become more common with the pervasiveness of Differential Privacy and other privacy-preserving technologies, it is sure to take time. And even then, organizations in the public sector may continue to face the difficulty of competing against private sector organizations to hire and retain personal capable of developing and enabling the use of robust privacy measures on vulnerable data.

6. Conclusion

This unprecedented era of technology "connectedness" and "big data" has virtually assured that we will never be left entirely alone, and that our views of privacy will be forever changed by the digital means through which we now interact with the world around us [56]. Similarly, this new way of life raises the ever-present specter of devastating privacy risks resulting from cybercrime that compromises or steals the personal data we all generate and share. Given its potential as a pragmatic tool for organizations and data stewards in the war against growing cybercrime threats, Differential Privacy fits well within a situational crime prevention framework and possibly represents a model to guide future privacy requirements engineering and protections directed exclusively at the issue of cybercrime. Therefore, policymakers and practitioners would be well-served to engage in empirical exploration of the implications that Differential Privacy and situational crime prevention collectively have on existing and new forms of cybercrime that are likely to emerge in the future.

Notwithstanding the practical challenges identified above, there is a continuing need for exploration and development of data privacy and disclosure methods that match our shifting data culture and also maintain the public's trust in institutions and industries [4]. The pursuit of these aims can be achieved through greater consideration of securing software systems early in development and implementation lifecycles and through a more dedicated focus on expanding privacy requirements engineering research [28, 57]. Additional attention should be directed more specifically at Differential Privacy as a unique design and implementation method. Research on the practical application and limitations of privacy enhancing technologies like Differential Privacy within the context of cybercrime remains necessary. In this regard, future studies might wish to employ a "no-free-lunch theorem" and investigate some of the popular misconceptions about Differential Privacy and its vulnerabilities, such as making no assumptions about how data are generated, that it protects personal information despite an attacker having knowledge of other individuals represented in the data, and that it is defensible to arbitrary background knowledge [58]. Doing so would ensure that subsequent use of Differential Privacy does not inadvertently contribute to future privacy-related challenges.

Generally speaking, most research exploring risks to individual privacy have been aimed squarely at consumer protection in the private sector. And while the average consumer should be cautious about the risks associated with sharing data for commercial use, there are other groups for which data privacy becomes a more considerable challenge. Vulnerable populations such as patients, children, the indigent, the elderly, inmates, undocumented immigrants, the civilly committed, and the mentally ill, are some of the most frequently studied populations, but are among the least likely to have the sufficient protections from data privacy intrusions. Efforts should be made to correct this imbalance by finding opportunities to make costly privacy-enhancing technologies available to public sector agencies.

There is also a significant need and opportunity for cross-disciplinary collaboration with respect to cybercrime and privacy-related research. Scholars from technical and social science disciplines are encouraged to join forces to expand the scope and breadth of research on the many threats to privacy which stem from cybercrime. They should also work together to investigate the variety of promising opportunities for preventing and responding to cybercrime threats, including Differential Privacy. Doing so would undoubtedly contribute to the development and spread of more appropriate and accessible approaches to the preservation of privacy.

DOI: http://dx.doi.org/10.5772/intechopen.92752

Ultimately, before moving forward with any Differential Privacy or any other privacy-enhancing technologies, data scientists, researchers, and practitioners should collaborate and carefully explore the consequences of this evolution in data protection. Additional resources and effort should be dedicated to the careful appraisal of privacy protections for person-level data in a variety of public and private scenarios. Failure to do so will likely result in more frequent and severe cybercrime breaches of critical infrastructure and significant privacy implications for individuals and groups whose data is widely available and easily accessible.

Author details

William Stadler
Saint Martin's University, Lacey, WA, USA

*Address all correspondence to: wstadler@stmartin.edu

References

[1] Mivule K. Utilizing noise addition for data privacy, an overview. In: Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012). 2012. pp. 65-71

[2] Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Design Direction [Internet]. 2018:1-101. Available from: https://arxiv.org/pdf/1802.07228.pdf [Accessed: 24 March 2020]

[3] Diamond B, Bachmann M. Out of the beta phase: Obstacles, challenges, and promising paths in the study of cyber criminology. International Journal of Cyber Criminology [Internet]. 2015;**9**(1):24-34. Available from: http://www.cybercrimejournal.com [Accessed: 03 April 2020]

[4] Snoke J, Bowen CM. Differential privacy: What is it? AMSTAT News [Internet]. 2019. Available from: https://magazine.amstat.org/blog/2019/03/01/differentialprivacy/ [Accessed: 03 April 2020]

[5] Wilner AS. Cybersecurity and its discontents: Artificial intelligence, the Internet of Things, and digital misinformation. International Journal: Canada's Journal of Global Policy Analysis [Internet]. 2018;**73**(2):308-316. DOI: 10.1177/0020702018782496. Available from: http://journals.sagepub.com [Accessed: 17 April 2020]

[6] Dwork C. Differential privacy. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Berlin, Heidelberg: Springer; 2006. pp. 1-12

[7] Dwork C. Differential privacy: A survey of results. In: Agrawal M, Du D, Duan Z, Li A, editors. Theory and Applications of Models of Computation. Berlin, Heidelberg: Springer; 2008. pp. 1-19

[8] Clarke RV. Technology, criminology and crime science. Crime and Deviance in Cyberspace. 2004;**10**(1):441-450

[9] Holt TJ, Bossler AM. Cybercrime in Progress: Theory and Prevention of Technology-Enabled Offenses. New York: Routledge; 2015

[10] Gordon S, Ford R. On the definition and classification of cybercrime. Journal in Computer Virology. 2006;**2**(1):13-20

[11] Anderson R, Barton C, Böhme R, Clayton R, van Eeten MJG, Levi M, et al. Measuring the cost of cybercrime. In: Bohme R, editor. The Economics of Information Security and Privacy. Berlin Heidelberg: Springer; 2013. pp. 265-300

[12] Tabansky L. Cybercrime: A national security issue? Military and Strategic Affairs. 2012;**4**(3):117-136

[13] Subramanian R, Sedita S. Are cybercrime laws keeping up with the triple convergence of information, innovation and technology? Communication IIMA [Internet]. 2006;**6**(1):39-50. Available from: https://scholarworks.lib.csusb.edu/ciima; https://scholarworks.lib.csusb.edu/ciima/vol6/iss1/4 [Accessed: 15 April 2020]

[14] Hinduja S, Kooi B. Curtailing cyber and information security vulnerabilities through situational crime prevention. Security Journal [Internet]. 2013;**26**(4):383-402. Available from: www.palgrave-journals.com/sj/ [Accessed: 16 April 2020]

[15] Khisamova ZI, Begishev IR, Sidorenko EL. Artificial intelligence and problems of ensuring cyber

security. International Journal of Cyber Criminology. 2019;**13**(2):564-577

[16] Finkle J, Chatterjee S, Maan L. EBay asks 145 million users to change passwords after cyber attack. Reuters [Internet]. 2014. Available from: https://www.reuters.com/article/us-ebay-password/ebay-asks-145-million-users-to-change-passwords-after-cyber-attack-idUSBREA4K0B420140521 [Accessed: 15 April 2020]

[17] Silver-Greenberg J, Goldstein M, Perlroth N. JPMorgan Chase hacking affects 76 million households. The New York Times [Internet]. 2014. Available from: https://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/ [Accessed: 15 April 2020]

[18] Cieply M, Barnes B. Sony cyberattack, first a nuisance, swiftly grew into a firestorm. The New York Times [Internet]. 2014. Available from: https://www.nytimes.com/2014/12/31/business/media/sony-attack-first-a-nuisance-swiftly-grew-into-a-firestorm-.html [Accessed: 15 April 2020]

[19] Swinhoe D. The 14 biggest data breaches of the 21st century. CSO Online [Internet]. 2020. Available from: https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html [Accessed: 15 April 2020]

[20] Perlroth N. D.N.C. says it was targeted again by Russian hackers after '18 election. The New York Times [Internet]. 2019. Available from: https://www.nytimes.com/2019/01/18/technology/dnc-russian-hacking.html [Accessed: 16 April 2020]

[21] Mazzei P. F.B.I. to Florida lawmakers: You were hacked by Russians, but don't tell voters. The New York Times [Internet]. 2019. Available from: https://www.nytimes.com/2019/05/16/us/florida-election-hacking-russians-fbi.html [Accessed: 15 April 2020]

[22] Broadhurst R, Grabosky P, Alazab M, Bouhours B, Chon SK. Crime in cyberspace: Offenders and the role of organized crime groups. SSRN Electronic Journal. 2013:1-42. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2211842 [Accessed: 24 March 2020]

[23] Cline S, Aronoff J. With great power comes great responsibility: Utilizing privacy technology for the greater bad. arXiv:200100226 [Internet]. 2019;**1**. Available from: http://arxiv.org/abs/2001.00226 [Accessed: 26 March 2020]

[24] Provos N, Rajab MA, Mavrommatis P. Cybercrime 2.0: When the cloud turns dark. Communications of the ACM. 2009;**52**(4):42-47

[25] Dickson B. The darker side of machine learning. TechCrunch [Internet]. 2016. Available from: https://techcrunch.com/2016/10/26/the-darker-side-of-machine-learning/ [Accessed: 06 April 2020]

[26] Herley C. Security, cybercrime, and scale. Communications of the ACM. 2014;**57**(9):64-71

[27] Lallement P. The cybercrime process: An overview of scientific challenges and methods. International Journal of Advanced Computer Science and Applications. 2013;**4**(12):72-78

[28] Pattakou A, Kalloniatis C, Gritzalis S. Security and privacy requirements engineering methods for traditional and cloud-based systems: A review. In: Dini P, editor. Proceedings of the CLOUD COMPUTING 2017 8th International Conference on Cloud Computing, GRIDs, and Virtualization. Athens, Greece; 2017

[29] Sweeney L. k-anonymity: A model for k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems. 2002;**10**(5):557-570

[30] Sweeney L. Weaving technology and policy together to maintain confidentiality. Journal of Law, Medicine & Ethics [Internet]. 1997;**25**(2-3):98-110. DOI: 10.1111/j.1748-720X.1997.tb01885.x. Available from: http://journals.sagepub.com [Accessed: 26 March 2020]

[31] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. ℓ-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data [Internet]. 2007;**1**(1):3. Available from: http://portal.acm.org/citation.cfm?doid=1217299.1217302 [Accessed: 24 March 2020]

[32] Ciriani V, Capitani di Vimercati S, Foresti S, Samarati P. In: Yu T, Jajodia S, editors. Secure Data Management in Decentralized Systems. New York, NY: Springer; 2007. pp. 291-321

[33] Barbaro M, Zeller T. A face is exposed for AOL searcher no. 4417749. The New York Times [Internet]. 2006. Available from: http://www.nytimes.com/2006/08/09/technology/09aol.html?ei=5090&e [Accessed: 26 March 2020]

[34] Heffetz O, Ligett K. Privacy and data-based research. The Journal of Economic Perspectives. 2014;**28**(2):75-98

[35] Narayanan A, Shmatikov V. Robust de-anonymization of large datasets (how to break anonymity of the Netflix Prize dataset). arXiv:cs/0610105 [Internet]. 2006:1-24. Available from: http://arxiv.org/abs/cs/0610105 [Accessed: 26 March 2020]

[36] Ayala-Rivera V, McDonagh P, Cerqueus T, Murphy L. A systematic comparison and evaluation of k-anonymization algorithms for practitioners. Transactions on Data Privacy. 2014;**7**:337-370

[37] El Emam K, Dankar FK. Protecting privacy using k-anonymity. Journal of the American Medical Informatics Association. 2008;**15**(5):627-637

[38] Li N, Li T, Venkatasubramanian S. T-closeness: Privacy beyond k-anonymity and-diversity. In: IEEE International Conference on Data Engineering. 2007

[39] Wong RCW, Fu AWC, Wang K, Yu PS, Pei J. Can the utility of anonymized data be used for privacy breaches? ACM Transactions on Knowledge Discovery from Data [Internet]. 2011;5(3):1-24. doi 10.1145/1993077.1993080. Available from: https://dl.acm.org [Accessed: 15 April 2020]

[40] Green M. What is Differential Privacy?—A Few Thoughts on Cryptographic Engineering [Internet]. 2016. Available from: https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/ [Accessed: 24 March 2020]

[41] Apple Inc. Differential Privacy Team. Learning with Privacy at Scale [Internet]. 2017. Available from: https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html [Accessed: 27 March 2020]

[42] Erlingsson Ú. Learning statistics with privacy, aided by the flip of a coin. Google Online Security Blog [Internet]. 2014. Available from: https://security.googleblog.com/2014/10/learning-statistics-with-privacy-aided.html [Accessed: 25 March 2020]

[43] National Conference of State Legislatures. Differential Privacy for Census Data Explained [Internet]. 2020 Available from: https://www.ncsl.org/research/redistricting/differential-privacy-for-census-data-explained.aspx [Accessed: 26 March 2020]

[44] United States Census Bureau. Disclosure Avoidance and the 2020 Census [Internet]. 2020. Available from: https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html [Accessed: 30 March 2020]

[45] Machanavajjhala A, He X, Hay M. Differential privacy in the wild: A tutorial on current practices & open challenges. Proceedings of the VLDB Endowment. 2016;**9**(13):1611-1614

[46] Hsu J, Gaboardi M, Haeberlen A, Khanna S, Narayan A, Pierce BC, et al. Differential privacy: An economic method for choosing epsilon. In: Proceedings of the Computer Security Foundations Symposium [Internet]. IEEE Computer Society; 2014. pp. 398-410. Available from: http://arxiv.org/abs/1402.3329 [Accessed: 25 March 2020]

[47] RVG C. Situational crime prevention: Theory and practice. British Journal of Criminology. 1980;**20**(2):136-147. Available from: https://heinonline.org/HOL/Page?handle=hein.journals/bjcrim20&id=148&div=19&collection=journals [Accessed: 16 April 2020]

[48] Clarke RV. Situational Crime Prevention: Successful Case Studies. 2nd ed. Guilderland, New York: Harrow and Heston; 1997

[49] Clarke RV, Homel R. A revised classification of situational crime prevention techniques. In: Lab SP, editor. Crime Prevention at a Crossroads. Cincinnati, OH: Anderson; 1997

[50] Fredrikson M, Lantz E, Jha S, Lin S, Page D, Ristenpart T. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In: Proceedings of the 23rd USENIX Security Symposium [Internet]. San Diego, CA; 2014. p. 17. Available from: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/fredrikson_matthew [Accessed: 27 March 2020]

[51] Higgins GE. Cybercrime: An Introduction to an Emerging Phenomenon. New York: McGraw-Hill; 2009

[52] Greitzer FL, Moore AP, Cappelli DM, Andrews DH, Carroll LA, Hull TD. Combating the insider cyber threat. IEEE Security and Privacy. 2008;**6**(1):61-64

[53] Mansfield-Devine S. The malware arms race. Computer Fraud & Security. 2018;**2**:15-20

[54] Maher D. Can artificial intelligence help in the war on cybercrime? Computer Fraud & Security. 2017;**2017**(8):7-9

[55] Reiter JP. Differential privacy and federal data releases. Annual Review of Statistics and Its Application [Internet]. 2019;**6**:85-102. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-030718-105142 [Accessed: 24 March 2020]

[56] Jerome J. Big data: Catalyst for a privacy conversation. Indiana Law Review [Internet]. 2014;**48**(1):213-242. Available from: https://heinonline.org/HOL/Page?handle=hein.journals/indilr48&id=229&div=12&collection=journals [Acessed: 16 April 2020]

[57] Mouratidis H, Argyropoulos N, Shei S. Security requirements

engineering for cloud computing: The secure tropos approach. In: Domain-Specific Conceptual Modeling: Concepts, Methods and Tools. Switzerland: Springer International Publishing; 2016. pp. 357-380

[58] Nguyen HH, Kim J, Kim Y. Differential privacy in practice. Journal of Computing Science and Engineering. 2013;7(3):177-186