Original Research Article

BIG DATA & SOCIETY

Big Data & Society
January–June 2019: 1–16
© The Author(s) 2019
DOI: 10.1177/2053951719848781
journals.sagepub.com/home/bds
SAGE

Clouded data: Privacy and the promise of encryption

Luke Munn, Tsvetelina Hristova and Liam Magee

Abstract

Personal data is highly vulnerable to security exploits, spurring moves to lock it down through encryption, to cryptographically 'cloud' it. But personal data is also highly valuable to corporations and states, triggering moves to unlock its insights by relocating it in the cloud. We characterise this twinned condition as 'clouded data'. Clouded data constructs a political and technological notion of privacy that operates through the intersection of corporate power, computational resources and the ability to obfuscate, gain insights from and valorise a dependency between public and private. First, we survey prominent clouded data approaches (blockchain, multiparty computation, differential privacy, and homomorphic encryption), suggesting their particular affordances produce distinctive versions of privacy. Next, we perform two notional code-based experiments using synthetic datasets. In the field of health, we submit a patient's blood pressure to a notional cloud-based diagnostics service; in education, we construct a student survey that enables aggregate reporting without individual identification. We argue that these technical affordances legitimate new political claims to capture and commodify personal data. The final section broadens the discussion to consider the political force of clouded data and its reconstitution of traditional notions such as the public and the private.

Keywords

Encryption, cloud-computation, privacy, personal data, homomorphic, blockchain

Introduction

The cloud introduces a new scale of observation, computation and control. Its advocates have argued its merits: a flexible utility, delivered on-demand (Buyya et al., 2009: 599); lower barriers to entry, scalable services and support for innovative applications (Avram, 2014: 531); and, at an institutional level, lower capital intensities compared with earlier informatic, media, and communications infrastructure. Driven by the proliferation of data, the intensive processing required by machine learning systems, and the demands of start-ups now dispersed globally from Dhaka to Santiago, the utility-like nature of the cloud in turn conditions and, in turn, is conditioned by today's pervasive and perpetual computing uses. Yet if the cloud can be technically defined as a 'systematized virtualization of data storage and access, the coalescence of processing power' (Coley and Lockwood, 2012: 1), it is a technology that disrupts politically as much as economically. Simultaneously computational architecture and metaphor, the 'cloud' reconfigures the political imagination through the different and often counter-logical realisation of the dialectics between obscurity, on one hand, and the making visible, on the other (Amoore, 2018; Hu, 2015), through data centralisation and accumulation.

This disruptive power has led to an increased sense of urgency in addressing what can be seen as a crisis of privacy. The amassing of data in the cloud has made it a focal point of vulnerability to attacks on personal data. Through the spectacle of media coverage of large-scale compromised consumer databases, threats to cloud-stored data appear to loom large. The rate

Institute for Culture & Society, Western Sydney University, Auckland, New Zealand

Corresponding author:
Luke Munn, Institute for Culture & Society, Western Sydney University, 118 Kahikatea Flat Rd R.D. 4, Albany, Auckland, New Zealand.
Email: 18560060@student.westernsydney.edu.au

of data breaches seems to be accelerating, with the Breach Data Index reporting that over 7 million records are now compromised every day (Gemalto, 2018a). One security commentator noted that 2017 was a 'monumental' year for leaks, observing that 'the number of data records compromised in publicly disclosed data breaches surpassed 2.5 billion, up 88% from 2016' (Gemalto, 2018b). In September of 2017, to take just one example, consumer credit reporting agency Equifax announced one of the largest breaches to date, revealing that 'the names, Social Security numbers, and dates of birth of 143 million US consumers had been exposed' (Gallagher, 2018). Congressional statements made later by Equifax management revealed that much of this information was stored in plaintext, without being obfuscated, encrypted or anonymised (Newman, 2017).

Moreover, data's ability to be combined in new ways complicates attempts to contain and protect personal information. Even when identifiers are hashed out or removed from data, individuals can be re-identified through various techniques (Ohm, 2010). In 2008, streaming giant Netflix made available a massive archive of viewing data in conjunction with a competition that challenged developers to come up with a better recommendation algorithm. While the data was thoroughly anonymised, Narayanan and Shmatikov (2008) demonstrated how it could be cross-referenced against IMDB information in order to identify specific individuals. More recently, De Montjoye et al. (2015) have shown how just the dates and locations from four credit card receipts yielded enough information to identify more than 90% of purchasers.

The attacking of cloud-based vulnerabilities and the adversarial capabilities of techniques like de-anonymisation exert increased pressure on privacy. But as efforts to undermine privacy grow, so does its perceived importance. Microsoft has recently made privacy one of its three 'core pillars' (Nadella, 2018). Facebook plans on hiring 10,000 new employees to address security and privacy in the wake of the Cambridge Analytica scandal (Hautala, 2018). And the European Union's General Data Protection Regulation (GDPR) puts individual privacy at the heart of its legislation (European Union, 2018).

Beyond such individualised concerns, these changes in the feasibility and reach of mass-scale computing put forth new questions about collective data privacy, data security and data sovereignty. The human subject is now interpolated in ways unanticipated in older machines of record. From the hand-written registers of the seventeenth century to the departmental databases of the twentieth, the ways data structured and delineated the person are surprisingly consistent and comparatively thin (Foucault, 2007). Interlinked social media, online health services and student and work histories thicken, intensify and exteriorise the subject's data profiles, criss-crossing private and public institutional interests with individual and group subjects (Amoore, 2014; Mittelstadt, 2017). The technological capabilities offered by data linkage and data analysis allow for a certain subjecthood to be assembled and disassembled in ways that bypass existing legal and ethical frameworks (Amoore, 2017; Cohen, 2019; Mittelstadt, 2017). Technologies that assume the integrity of the individual data subject, such as obfuscation and anonymisation of data, provide only partial protection. Following from what Montgomery and Pool have termed 'experimental publics' (2017), the assembly of these heterogeneous individual traits into ad-hoc clusters might be termed 'combinatorial publics': social ensembles that are made and unmade with the cut of a declarative query or filter operation. Yet the response to the Cambridge Analytica affair and other scandals also illustrates the ways the massification of data produces a political activation of subjects.

We introduce the term 'clouded data' in order to discuss this series of transformations that develop through data accumulation, data privacy and value extraction in the cloud. Used descriptively, the concept refers to the twinned condition of personal information today. Companies and agencies want to unlock the potential value within the data by resituating it within the cloud, a massive process of data centralisation. These cloud-based architectures render data computable and interoperable to a new degree, able to be intensively processed and endlessly recombined with other repositories to generate new insights. At the same time, to protect this highly valuable information, data has also become clouded in the sense of obfuscation, encrypted or distorted to protect it from unwanted surveillance or intrusion. These twinned processes are therefore more than coincident: centralising data makes it more vulnerable, requiring technologies for obfuscation; and the pooling of computational resources in turn makes those technologies computationally tractable and economically feasible.

However 'clouded data' not only describes the technological properties opened up by data in the cloud. It also encompasses political responses to threats and dangers to privacy; the technologies designed to ameliorate such threats and dangers; and, in turn, the ways these technologies themselves open up different political scenarios and different constellations of political actors. This complex movement shows that not only is technology, in a generalised sense, generative of political meaning and implications, but that *different* technological designs produce *different* arrangements of power relations and possibilities for intervention.

Collectively, demand for computational resources, data ownership and access, ease of use, and what can be considered different architectures of privacy, which we explore in detail below, form the parts of technopolitical assemblages that, conversely, can only be understood through analyses of their computational materiality and social practice. Beyond its descriptive value, 'clouded data' denotes then the production of a novel field of differentiated activation, where political concerns can arise, technological responses can be initiated, and political possibilities can in turn be generated around the central issues of data accumulation and privacy.

This field is neither technologically nor politically homogeneous, and in this article we offer an attempt to analyse the multiplicity of arrangements by looking at four of the technological solutions for data privacy. Each expresses what Julie Cohen has referred to as 'privacy by design' (2019), a particular set of 'design, production, and operational practices' that construct a distinct version of privacy and make it available through an infrastructure. In these technical environments, privacy emerges from protocols and feature sets, rather than adhering to an *a priori* normative standard. In other words, we want to explore privacy as affordance rather than abstraction. Understanding the design of these cloud-based technologies implies a certain political economy, a particular arrangement of power, trust and capital, which points in turn to the ways in which they open up new fields of the political, new dependencies between publicness and privacy, and attach new significance to these categories.

The rest of the article explores how clouded data reconfigures privacy. The first section surveys four technologies for data security in the cloud. The second section uses code-based experiments and notional datasets to engage with cloud-based encryption frameworks: in a healthcare context, we posit a scenario around a patient's blood pressure; in a tertiary education context, we work with a student survey. These empirical engagements illustrate how distinctive forms of privacy emerge from particular technical affordances. The final section broadens the discussion to consider the political force of clouded data and its reconstitution of traditional notions such as the public and the private.

Securing the cloud: Four approaches to networked data privacy

We identify and review four cloud-based cryptographic responses to privacy concerns: blockchains, differential privacy, multiparty computation (MPC) and homomorphic encryption. Each technology emphasises a distinctive aspect, staking out a particular territory within the general field of computer security. With its current hold on the public imagination, blockchain represents the first, highly popular approach investigated. Second, as a comparatively unobtrusive means for preserving anonymity in data sets, differential computation foregrounds ease of application as a factor. Next, secure MPC epitomizes what Claude Shannon (1949) defined as 'perfect secrecy': encrypted messages reveal nothing of the key used to encrypt the message. Finally, fully homomorphic encryption (FHE) stresses computability, encrypting data while still allowing it to be operated on.

These four approaches do not indicate mutual exclusivity, nor a definitive articulation of the field itself, but rather sketch a provisional terrain of clouded security today. Our interest is in how each conjures a distinct world of relations between social actors. In the sense Cohen (2019) has suggested, each technology actively coordinates and designs an inflected concept of privacy, and through their respective implementation – in some cases, still highly experimental – intervenes in the unfolding process of data clouding. This, as we discuss later, is as much as a shaping of political imagination as of technical infrastructure.

Blockchain

Blockchain technology encompasses a variety of security models. We begin by describing that used in the most widely known blockchain examples, and then elaborate on more recent versions. Bitcoin and Ethereum are open and public blockchains. Able to be downloaded or inspected by anyone at any time, they operate via what might be termed 'trust-through-transparency.' Distributed among all parties, no one has more information than any other. Any member on the network can send and receive transactions; any member can verify whether blockchain data is consistent and complete. Bitcoin and Ethereum extend this principle of informational symmetry to their security models, which employ public key cryptography to grant all parties theoretically equivalent degrees of anonymity. Each Bitcoin transaction is, for example, sent to an 'address,' a hashed and encoded version of a public key. Though exposed throughout the network, these Bitcoin addresses cannot necessarily identify their owners. However, since total privacy still requires discipline on the part of blockchain members (not exposing their public keys alongside their personal details for instance), this property has been termed pseudonymity rather than strict anonymity. A distributed infrastructure with distinct technical properties – encryption, 'proof of work' checks against fraud, and a data store that can only be appended to, not deleted or edited – seeks to establish an egalitarian or 'trustless' network that democratises, in theory, control over financial transactions, as well as other contractual arrangements (Finextra Research and IBM, 2016).

Yet the distributed character of the blockchain also poses challenges. As a public and inherently complete record, its ledger is permanently open to new exploits that seek to re-identify addresses accompanying transactions. As Primavera De Filippi argues, ‘anyone can retrieve the history of all transactions performed on a blockchain and rely on big data analytics in order to retrieve potentially sensitive information’ (2016: 0). Moreover, blockchain’s distributed ledger records information permanently, and this inability to remove or amend records may violate new privacy regulations such as the GDPR’s Article 17 (European Union, 2018). Different instances of blockchain technologies thus produce different inflections of privacy; users are not exposed to threats in a uniform way. Instead, those with sufficient technical and financial resources can better mask their identities, either by combatting identifying techniques on popular platforms like Bitcoin, or by using more secure but also more complex alternatives like Zcash or Monero.

Partly in response to these criticisms, other blockchain designs explore alternatives with greater security, flexibility and efficiency. These ‘permissioned’ blockchains restrict membership to an invited list of parties, typically at the discretion of a central authority who initiates and governs the blockchain. At least from the blockchain provider’s point of view invitees are no longer even pseudonymous, and may be assigned roles that further constrict their activity. Conversely, the private character of these systems means they can be secured against third-party access: blockchain data is distributed only among authorised members, and is limited to the data specific to the blockchain’s purpose. Permissioned blockchains also need not store all data within the blockchain itself, and can support a hybrid model, where repositories of ‘off-chain’ personal data are pointed to by small ‘on-chain’ references (Zyskind et al., 2015). By only storing references, limiting the number of parties and replacing proof-of-work with simpler consensus procedures, such permissioned blockchains can utilise far fewer network, storage and processing resources. However they also re-establish the central mediating authority – a bank, insurance or healthcare provider – that public blockchains originally sought to bypass. Indeed companies like IBM and Oracle, sensing an opportunity to leverage existing database technologies, have promoted the use of permissioned blockchains for enterprise (Mearian, 2018). In such cases, privacy hinges once again on trusting a central authority and what is often proprietary infrastructure: the servers, databases, encryption standards and security procedures through which such blockchains are administered.

Differential privacy

In differential privacy, privacy is manufactured by making an individual’s contribution to any given data statistic arbitrary or contingent (Dwork, 2006). Differential privacy obscures personal data by introducing noise in statistical datasets in such a way that makes it impossible to deduce whether an individual’s data is part of that dataset or not. Differential privacy advocates explain the concept by positing two worlds: in one world, an individual takes a survey and contributes to a dataset; in the other, she does not. This discrepancy is then formalized systemically, and a corresponding amount of noise added to queries. For differential privacy then, ‘privacy’ is an adjustable value, a parameter on a virtual control knob: dialled up, each record resembles less and less its original form, and the accuracy of statistics declines; dialled down, the ‘true’ shape of a modelled public recrystallizes, and so too do the sharp contours of each individual’s profile (McSherry, 2018). Properly configured, such ambivalence protects individuals at the level of the single record, while still allowing broad trends to emerge when analysed in aggregate. Properties of the public are revealed; properties of the person are not. For pioneers Cynthia Dwork and Aaron Roth, this indeterminacy enables a privacy promise: ‘you will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available’ (2014: 5).

Relative to other approaches, differential privacy has certain affordances: the technology is usable for non-experts, who can run queries without understanding the underlying mechanics; it supports the broad range of queries that analysts are already using; and it integrates with existing data environments, rather than requiring new database architectures (Near, 2018). Near also notes these merits have been adopted in production systems: using differential privacy, Apple has analysed the power consumption of websites and the popularity of emojis without comprising individual privacy, while Google has studied browser malware and traffic analysis in large cities.

Since differential privacy does not address the underlying data itself – depending on configuration, that data is either transformed during initial load, or remains intact but protected while the adjusted data is online – it leaves open the potential for breaches, leaks, or disclosures from adversaries who access and redistribute it. Additionally, and unlike public blockchains, a central authority must control and protect such data; as Dwork and Roth argue, individuals supplying data must ‘assume the existence of a trusted and trustworthy curator who holds the data of individuals in a database’

(2014). While it defuses the ability of analysts to obtain information damaging to any individual, differential privacy presumes both the reputation of the provider and the security of the architecture that delivers responses to queries.

Multiparty computation

Like blockchains and unlike differential privacy, MPC assumes instead the adversarial nature of the network itself. Privacy in this antagonistic environment consists in never trusting any single agent, not even a benevolent curator, with a meaningful dataset. Instead, in the 'Secret Sharing' approach to MPC, information is split into meaningless pieces that are then distributed to a large number of providers for computation and analysis. In this model, privacy is assured not by ciphertexts alone, but also by the fragmented nature of corporate ownership and computer architectures. As Zyskind (2017) explains, 'an attacker would need to compromise *t* servers at any given point in time to get the data back, which is highly unlikely for a large *t*.' Though distributed, the use of the term 'server' here indicates that, unlike blockchains, MPC operates in a more common client-server rather than peer-to-peer network topology; most explanatory diagrams of MPC bear this out, showing a large number of users and a smaller number of servers rather than the coincidence of individual and device.

Though theoretically possible for some time, and debuted in 2008 with Dutch sugar beet prices (Bogetoft et al., 2008), MPC has seen several notable real-world deployments, such as the evaluation of gender pay disparities in Boston (Lapets et al., 2015) and tax fraud in Estonia (Bogdanov et al., 2015). More recently, engineers from Google have discussed how they use MPC to evaluate advertising views or track Android keyboard use while ensuring a degree of privacy (Wood, 2017). At scale, overheads become critical – distributing computation widely across a real-world network like the Internet imposes significant performance costs. To address this constraint, Google's implementation of MPC replaces antagonistic-assuming 'academic' protocols with less severe 'industry' versions that 'require only specific protocols, which can therefore be optimised, and comparatively weak security guarantees' (Wood, 2017).

As these examples suggest, computational cost and complexity ensures the 'parties' in a MPC scheme are most often institutional. These cases of MPC imply a prior arrangement between public and private institutions and the subjects whose data they curate and analyse. Similar to security tools on a personal computer or mobile phone, they seek to be transparent about what they conceal, and in doing so make an implied moral appeal to a security-conscious public. In the contexts accompanying its use to date, MPC seeks to conserve existing institutional–individual relations, and in commercial contexts, offer a feature that differentiates its provider from competitors.

Homomorphic encryption

Homomorphic encryption suggests a solution with a total obfuscation of personal data. In a typical, non-homomorphic context, cloud-based service providers decrypt data in order to run computations and deliver analysis, but this temporary decryption presents an unacceptable vulnerability – privacy is compromised at the moment data is retrieved for computation. The goal of homomorphic encryption is to operate on encrypted data as if it was decrypted, retaining privacy while enabling data analysis. Computation takes place on ciphertexts and generates an encrypted result, which is then returned to the user, who decrypts it. With the advent of asymmetric or public key encryption, FHE imagined a complete set of computing functions based upon support for both additive and multiplicative operations.

Although first suggested in a paper by Rivest, Adleman and Dertouzos in 1978, and further developed by Goldwasser and Silvio Micali in 1982, the possibility of FHE was thought to be practically infeasible. In 2009, Craig Gentry (2009) outlined the first FHE scheme that could handle both addition and multiplication operations, using the mathematical notion of ideal lattices and a technique called 'bootstrapping'. The breakthrough, however, came with significant performance limitations, since the size of the encrypted ciphertext grew enormously with each operation (Schneier, 2009). In 2009 Gentry himself estimated that his scheme would multiply computing time by a factor of a trillion (Greenberg, 2009). Thus homomorphic encryption is highly attractive in a cloud-computing environment, but its performance characteristics have limited its adoption.

Nonetheless much work over the last decade has focused on improving FHE performance through hardware acceleration (Wang et al., 2012), software optimization and prepared datasets – all of which require significant investments. Performance has continued to improve over time (Acar et al., 2017) and most recently a team from Microsoft won the iDash competition with an entry that reduced computation times down to seconds (Çetin et al., 2017). An implementation of the popular HElib library by Shai Halevi, released in March of 2018, claims that optimizations enable speedups of 15× to 75× (Halevi, 2018). Given these order-of-magnitude improvements over a decade, performance is no longer the roadblock for feasible real-world deployments that it once was (Hallman et al., 2018) and this will certainly impact on the dynamics between providers of cloud-as-a-service, data providers and data analysts.

Comparative analysis

In Table 1, we distinguish technical and practical properties of these four security schemes, summarising features and trade-offs. The table shows, for instance, blockchains and MPC are network-dependent, while differential privacy and FHE can operate on single machine; and that MPC and FHE offer current 'state-of-the-art' security, but differ fundamentally in approach.

In the second part of the table, we use these technical properties to derive a series of assessments about how privacy is being configured and designed. These assessments are necessarily fragmentary, since we do not presume theories of privacy can be naively read off the properties of technological systems themselves. The 'reconstructive' case studies below explore more deeply how the individual subject's relation to institutional power may be reworked by two of these systems: MPC and FHE. Yet even the inferences we make here convey some sense of the changing relationship between data, privacy, an individual juridical and political subject, and the queryable 'combinatorial publics' discussed earlier. Blockchains, for instance, hide identifiers in plain sight, while differential privacy requires data be stored by a trusted party.

This last point hints at a further distinction between the four technologies. Blockchain and multiparty computation both emerge from a theoretical security – and in blockchain's case, a further explicitly ideological – desire to decentralise data control (despite the fact that blockchains in practice have tended to become highly centralised). While neither explicitly articulate a centralised computing architecture, both differential privacy and FHE imply powerful centralised computing resources. In the case of differential privacy, determining the degree of noise required to adjust one data record so that it can neither be identified nor perturb aggregate calculations implies control of the complete unencrypted data set by a single provider. In the case of FHE, at least for many sufficiently large, i.e. population-level, data sets, processing power requirements would also imply a dedicated data centre facility. In practice though, both blockchain and MPC have tended also to favour centralised configurations; in the case of famous blockchains like BitCoin, because mining operations have progressed from personal computers to clusters of dedicated mining machines; in the case of MPC, at least in certain cases, because functional systems have tended to be developed under proprietary licenses that favour controlled, i.e. centralised operating environments.

Speculating on encryption: Cases in healthcare and tertiary education

Of these approaches, secure MPC and homomorphic encryption (FHE) have received the least scrutiny in regard to their social implications and by extension, their potential reconfiguration of conceptions of privacy. Applications of blockchain have been a subject of attention in both technology media coverage and IT literature while differential privacy has similarly been comprehensively examined, without the same media acclaim. Due to their comparative novelty and complexity, the distinct form of data privacy constructed by MPC and FHE is less well understood, motivating our selection in the two case studies that follow.

Our method follows those employed in critical code studies, notably in Mackenzie's 'code-based reconstruction' (2018). We develop two such reconstructions, based on Australia's health and education industries, where data privacy is under intense scrutiny and has itself become an explicitly and intensely politicised topic. The institutions of the clinic and the school (Foucault, 2002, 2012) have been paradigmatic sites for the reproduction of sovereignty, power and subjectivity in modern civic society. While claims of the virtualisation of these institutions (through telemedicine and online learning for instance) may be overstated, they are equally essential sites for examining how technologies of 'clouded data' intervene in the establishment of new concepts and practices of data sovereignty. Enacting how new cloud-based technologies might apply to these institutional scenarios – developed and deployed here in much the same ways as they are in the vast literature proselytising cloud-computing – enables, through the differences of these scenarios with the *de facto* conditions of privacy in those industries today, something of the wider reconfiguring of relations between subject, corporation and state to be seen.

While we list several of the affordances and limits of the two approaches, our purpose is not to undertake a technical evaluation of the kind widely used in computer science and information systems disciplines, nor to follow the suggestive possibilities of 'tool criticism' (Van Es et al., 2018) developed in software and media studies. Rather our work constitutes a form of simulation-as-reflexive-practice, designed to anticipate how a future of 'clouded data' might shift political notions of privacy and publics. In the health scenario, we utilise PySEAL, a Python interface to the Simple Encrypted Arithmetic Library (SEAL) FHE implementation developed by Microsoft and open sourced in late

Table 1. Summarizing some key differences in the privacy implemented by cloud-based technologies.

	Blockchain	Differential	MPC	FHE
Technical properties				
Network configuration	Peer-to-peer	Cloud (client / server)	Peer-to-peer; Cloud (multiple servers acting as parties)	Cloud (client/server)
Computational accuracy	N/A	Variable: low–medium (accuracy is inversely proportional to security)	High	Variable: low–medium (accuracy is inversely proportional to security)
Computational cost[a]	Medium	Low	Medium	High
Network cost	Medium	N/A	High	N/A
Security mechanism	Encrypted identifiers; 'mining' to verify new data entries	Statistical noise added to data values	Protocol for computing functions over private data shares	Public key cryptography
Parametric[b]	No	Yes	No	No
Variable strength	No	Yes	Yes	Yes
Computation possible on encrypted values only	No	No	Partial	Yes
Privacy design characteristics				
Data is made	Public and pseudo-anonymous	Fuzzy	Fragmentary but computable	Encrypted but computable
Protection	Total knowledge	Ambiguity	Through networks	Through numbers
Trust placed in . . .	Anonymity of identifiers, and decentralised nature of transaction verification	Data curator, selection of noise parameter to balance accuracy and anonymity; and security of source data	Size and distribution of network	Public key infrastructure, and lack of 'back-door' decryption methods
Key unit	Blocks	Functions	Shares	Ciphertexts
Economic arrangement	Third-party miners (often data centres)	Data centre operator	Variable (incl. data centres)	Data centre operator
Threat to security	Identifiers correlated with individuals; More than 50% of parties are compromised by an adversary	Insufficient noise to de-identify individual records; trusted curator is compromised	More than 50% of parties are compromised by an adversary	Insufficient key length; theoretical compromise of homomorphic encryption

FHE: fully homomorphic encryption; MPC: multiparty computation.

[a] 'Cost' is an approximation of theoretical time or computational complexity. In the case of blockchain, this cost varies enormously depending on implementation (i.e. 'proof of work' vs. 'proof of stake').

[b] 'Parametric' here refers to the dependency of techniques of obfuscation on features of the function or data set being operated on. Differential privacy normally depends, for instance, upon characteristics of the data set or query to determine how much noise to add to individual data values prior to or during computation of results.

2018. In the education scenario, we employ a secure MPC system developed by NTT, San-Shi (2017), which we obtained access to via a partnership with NTT's subsidiary, Dimension Data. In a paper describing San-Shi, Tanaka et al. (2017) claim it is resistant to adversarial attacks from even the majority of parties, offering even stronger protection than conventional multiparty encryption schemes. Each implementation therefore reflects the current state-of-the-art, anticipating cloud security to come rather than mature or widespread systems in use today.

Healthcare

The importance of data security in healthcare has increased with the progressive digitalisation of patient and clinical data. In Australia, the introduction of electronic health records in 2012 has driven efforts to improve data interoperability between private and public health providers and insurance schemes such as Medicare. Such efforts make the case for high levels of healthcare data security all the more pressing. The current version of electronic medical records in Australia, My Health Record, is planned to include every Australian, with an opt-out option, by the end of 2018 (Australian Digital Health Agency, 2018). However, privacy concerns are stalling the planned adoption of cloud storage in the industry (e-Health Strategy, 2018) and threaten to sabotage the plan for universal coverage. In the first couple of weeks after the launch, the rate of opt-outs from the system has raised concerns about the viability of the plan (Stilgherrian, 2018).

In its current state, Australia's e-health system comprises mostly of siloed data systems, which hinders interoperability. Such siloing does limit the damage of data breaches, but does not prevent them altogether. The focus of the Australian Digital Health Agency on health data security has so far been directed predominantly towards legislative measures and controlling the access to the My Health Record database (2018a). One of the measures adopted is the roll out of a new Notifiable data breach scheme, introduced in February 2018, which makes the reporting of personal data breaches mandatory for organisations and entities handling personal data (Office of the Australian Information Commissioner, 2018).

However legislative measures to control access to My Health Record databases have limited efficacy. They address traditional healthcare scenarios – sharing medical data between healthcare providers – but not the challenges brought about by newer technological innovations such as cloud-computing and Big Data analytics. Such measures are expressed in what Cohen (2019) has termed 'liberty-based language of human rights discourse' which 'are both difficult to dispute and operationally meaningless,' especially when such technical operations are opaque and practically inexplicable. More direct critiques have emphasised the inadequate encryption and anonymisation of My Health Record data. In 2016 the bulk of partially encrypted healthcare data shared for research by the Health Ministry was discovered to be vulnerable to re-identification and taken down (Dunlevy, 2016). Analysis revealed that cross-checking with other publicly available databases containing personal and tax information could help re-identify individuals (Culnane et al., 2017). These specific examples point towards a broader inability to completely foreclose de-anonymisation – datasets released in the future could provide the key link to re-identifying individuals or revealing personal data.

In our experimental scenario, a patient takes a blood pressure test at a local clinic. She would like to know whether this reading indicates a risk of hypertension. The clinic has recently learned about a secure cloud-based service able to determine if the patient's blood pressure reading is abnormally high using machine learning techniques. As recent studies suggest, such techniques have been shown 'to provide solid prediction capabilities in various application domains including medicine and healthcare, including in the area of hypertension' (Sakr et al., 2018). Yet the patient is unwilling to share her unencrypted history with online services, concerned that any discovered risk factors might be shared with health insurers or potential employers. Her doctor informs her that her data will be first encrypted, and that the cloud-based service performs its computations solely on that encrypted information. With her consent, the clinic submits the patient's details, including her blood pressure, alongside a database of other, comparable patients – all encrypted. The service determines that the patient's readings are indeed abnormally high, and could be a predictor of hypertension. Such an encrypted and highly focused analysis would allow the patient to make informed choices about lifestyle, diet and potential treatment, while retaining control of her private and highly valuable health data.

We generated and encrypted a small set of blood pressure values, derived from mean and standard deviation values reported for Australia by the World Health Organisation (Kuulasmaa et al., 1999). We developed a simple Python class which would accept (a) the clinic's public key, (b) the encrypted set of previous client records and (c) the patient's blood pressure reading, also encrypted. Without the secret key, objects have no way of deciphering the encrypted data, but can compute meaningful results. Our criteria were intentionally simplistic: assuming blood pressure readings are normally distributed, is the patient's reading in the top 5 percentile (two standard deviations above the mean)? We synthetically set the patient's reading to indicate 'at risk', and then evaluated whether we could determine this once all the records were encrypted and submitted to the cloud service. Figure 1 plots the patient's result (in red) with a set of other randomly generated readings.

We had several challenges implementing even this basic algorithm. Because PySEAL provides no means of performing square root, variable division or number comparison, we were restricted to calculating means and variance. Since variables cannot be compared, the software client needed to test the result of the function with unencrypted data. As noted in comments in the PySEAL example code, calculations are susceptible to the parameters supplied to the homomorphic encryption scheme. Setting these parameter values too low led to calculation errors, while setting them too high introduced a dramatic decline in performance. While our tests were not designed for benchmarking, there was a noticeable performance decrease in calculating the mean even for 25 compared with 10 observations, with parameters set quite low.[1] Beyond 25 observations, both performance and accuracy decreased dramatically.[2] Thus, despite being touted as 'homomorphic encryption in a user-friendly Python package' (Kishore, 2018), we found PySEAL presents challenges even for experienced software developers to use. The project's GitHub page acknowledges it is a 'proof of concept,' and implementing workarounds to calculate functions such as standard deviation is unusually 'low level' (Titus, 2018). The effort to implement such commonplace functions indicates something of the labour required to refactor existing software to incorporate homomorphic encryption. While such functions may eventually be integrated as libraries like SEAL and PySEAL mature, this labour poses a further obstacle to FHE adoption.

Nevertheless, we were able to implement a simplified 'outsourced computation' scenario. As a technical demonstration, the experiment shows a potential

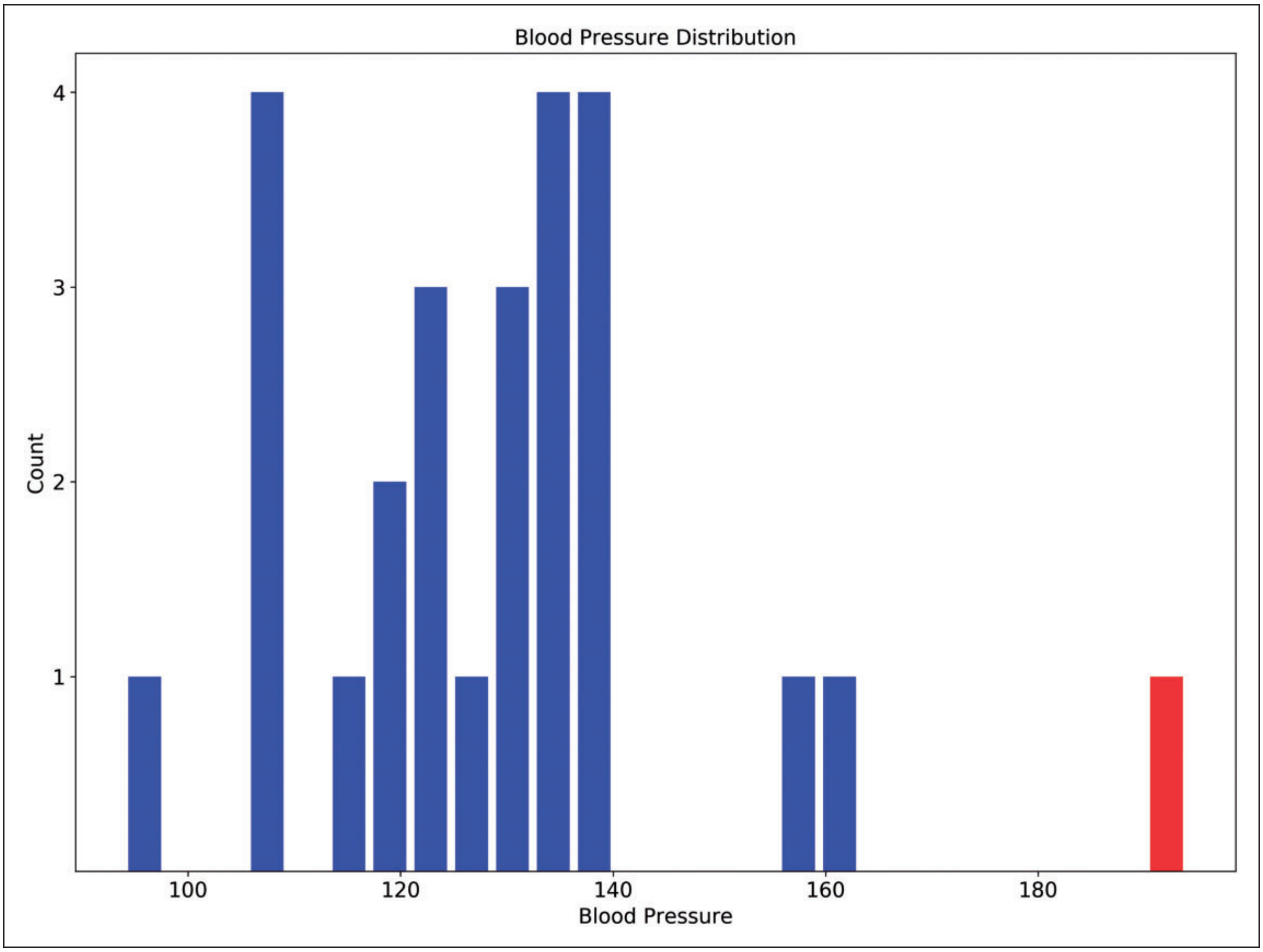

Figure 1. Synthetic blood pressure readings (patient's reading in red).

reconfiguring of assumptions underpinning the privacy disclosures of a national facility like MyHealthRecord. In our naive case, a diagnosis is performed without disclosure of even anonymous plaintext data. The case could be extended to a much wider set of patient data, with personal and identifying characteristics, that could be linked to other data sets in a fully encrypted environment. Under such conditions, disclosures that usually trigger informed consent might instead fall away. In bypassing former privacy limits, encrypted yet computable data sets become a common resource available for analysis by health organisations and indeed any other actors. And yet the cloud that makes such computation possible becomes even more integral, an essential mediator in transactions between client and the health industry. This registers a shift in control from public institutions to private platform corporations.

Tertiary education

The field of higher education is guided by similar concerns about the security of personal data, which are further complicated by the variety of citizenship, migration, financial and social information collected by educational institutions (Australian Government, Department of Education and Training, 2018). Concerns over use of student information have historically focussed on issues of bias and validity (Druckman and Kam, 2011). With the rise of networks and social media, exploitation of student data for research and commercial purposes has begun to receive critical attention (Hewitt and Forte, 2006). Commensurate with the rise of audit culture, universities often survey students to monitor course satisfaction, to boost metrics of engagement, or to gather information for research projects.

Both research and market surveys are often anonymous, but will sometimes include identifying information such as a student ID. In such cases, data privacy policies and university ethics committees will often constrain the ways such identifiers may be used, prohibiting the merging of research data with other databases containing course results or student enrolment records. While such constraints adhere to the university's duty of care toward its students, they limit analysis that could be derived from such merges. In the scenario below, we explore how such analysis might be undertaken with a secure multiparty computational environment.

In this scenario, a research team in a university business school wants to know how well students feel their courses were preparing them for the future job market. They would like to administer a survey to the university's students, with questions like:

> Please state your level of agreement with the following statement:
> 'I feel confident my current course is preparing me for the future job market.'

In addition, the team wants to know how student responses related to their course of study, their place of residence, and economic factors such as student debt and household income. A motivating research question might be: do students from lower socio-economic backgrounds feel more or less positive about how their course is preparing them for future employment?

The team applies to the university's ethics committee for permission to administer their survey. They are informed that their survey can contain basic questions about work preparedness, but not sensitive questions regarding income, background or place of residence, as these would be invasive of privacy. However students' postcodes are captured by the university's enrolment system, and the team does obtain approval from the university's ethics committee to ask for student ID numbers in their survey. The team also explains clearly to all research participants why they are asking for these identifiers, and emphasise they will not be able to use these identifiers to obtain sensitive information from students. After four weeks of running their survey, the team has 1000 survey responses, including attitudes about work preparedness.

They then upload a spreadsheet of these responses to the San-Shi system, where it is encrypted. The same system also has an encrypted copy of student enrolment records, including postcodes. By matching student ID numbers, the team can cross-index their survey with the enrolment records to generate a more comprehensive set of student data. Without being able to look at the original records, the team can generate statistics about responses by postcode. Using measures of socioeconomic disadvantage and cartographic data from the Australian Bureau of Statistics (2018a, 2018b), they then generate a series of maps and tables to explore the data.

Figure 2 shows the distribution of average scores (where 1 = 'Strongly Disagree' and 5 = 'Strongly Agree') across various postcodes in Western Sydney. Barrel distortion is applied to magnify the smaller postal areas surrounding Parramatta. A clustering of low or high response postcodes might indicate that attitudes vary spatially across Western Sydney. The data was generated in *R*, using an inverse logistic function to sample a distribution of responses biased by distance of respondents' place of residence from Parramatta.

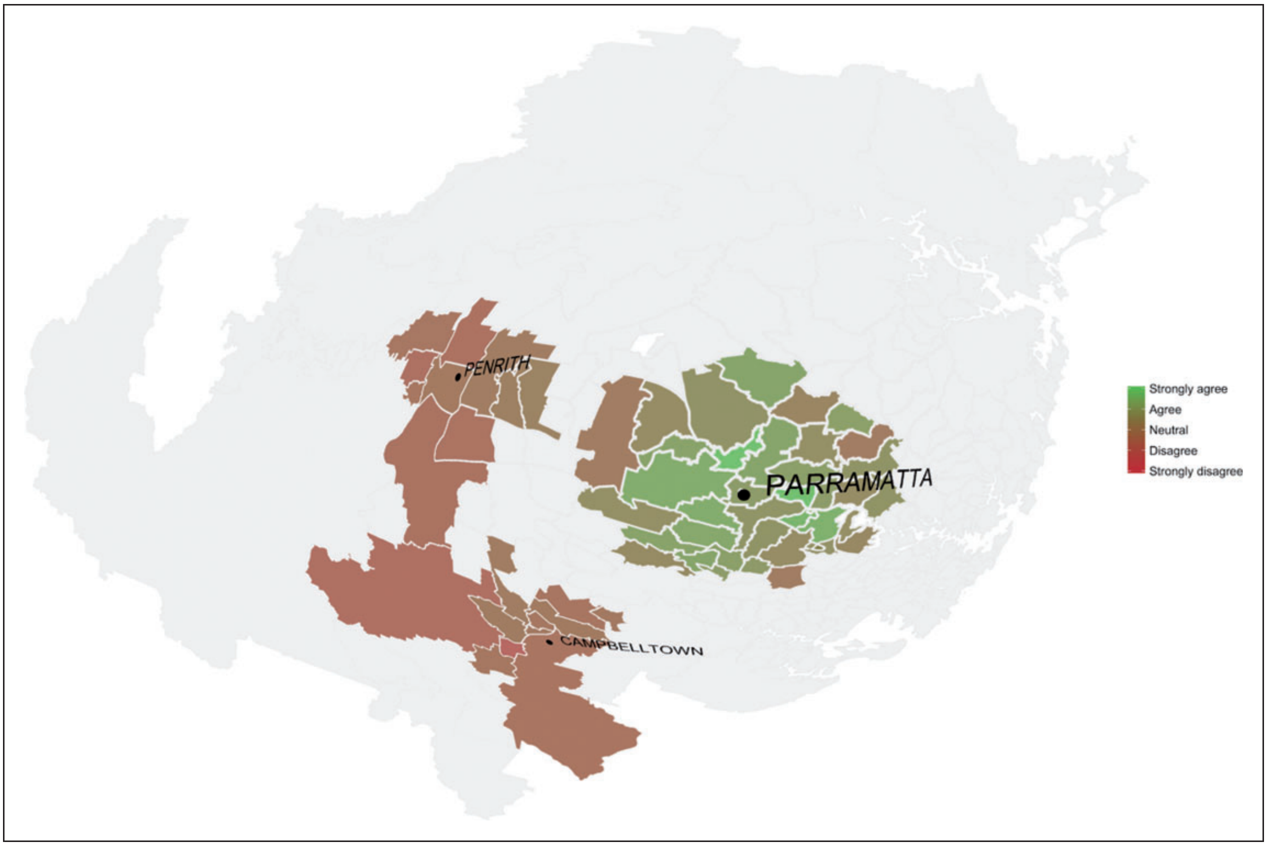

Figure 2. Responses to question on work preparedness, mapped by postcode.

The aggregate responses extracted from the encrypted San-Shi data reflected this biased distribution.

Whereas it was possible with PySEAL to do limited computations with individual encrypted values, here we could not obtain access to any underlying records. In the scenario, this ability allows the School of Business team to comply with ethics while still generating aggregate results. While for testing purposes a single system was operated under the control of a provider, other configurations could include multiple parties who each hold meaningless data shares that are only ever reconstituted in response to queries using a secure protocol. The particular implementation, then, presents a field of possibilities stretching from the singular control exercised by the platform provider to the distributed and decentralised topology of peer-to-peer networks. According to the setup, privacy is configured as either the promise of a trusted corporate guarantor, or a property that rises with the growth of networks. In both cases, control over data by, for example, a state actor is undermined, dispersed toward either a corporate mediator or a multitude of other actors.

Discussion

Our survey of cloud encryption approaches and the experimental scenarios illustrate how technologies construct a particular version of privacy. Each framework has its own implementation of security, its own understanding of trust, its own formalization of roles. Privacy emerges in specific formations based on underlying architectures and embedded assumptions. At a higher level, these technical imaginaries encompass roles and responsibilities, suggesting how cloud-based privacy should work and who should operate it.

Usability provides one way of understanding who a technology is intended for. As both experiments show, current implementations are complex even for experienced technicians to administer, query and programme. In the case of FHE, while performance has improved, integration into real-world projects retains a formidable learning curve. As encryption specialist David Archer (2016) has observed, the requirements ‘to transform programmes into circuits, carefully configure FHE computations, manage encryption and decryption, and other complexities make programming FHE applications the domain of a small number of expert researchers.' For example, to use the SEAL library for basic encryption tasks, 'the first step is to create a new *EncryptionParameters* object, and to set its modulus attributes. The polynomial modulus should be set to a power-of-2 cyclotomic polynomial' (Titus, 2018). Compared with using common statistical functions in data analysis languages like *Python* and *R*, the degree of expertise and requirement for labour involved in homomorphic encryption restrict use to dedicated research and experimental commercialisation environments. The analyst manual for San-Shi is similarly complex. Successful use requires configuration of thresholds for fragmented data, registration of tenants for multiparty sharing, implementation of standard functions such as correlation, and other details not typically part of a data analyst's training.[3] For at least the foreseeable future, such complexity demands the inclusion of the technician in these privacy arrangements; a demand still evident even with more widely available systems like blockchains and differential privacy.

Admittedly MPC and FHE are emerging technologies, at least at the level of implementation. Yet it is precisely at this early juncture that roles are established – 'privacy' becomes a matter for the paid experts of private companies, who offer it back to consumers. In this sense, usability, while anchored in graphical interfaces and help manuals, extends into the broader domain of accessibility. Encryption technologies employ a particular language, assume a certain technical familiarity, and suppose access to necessary computational architectures. In this way, the contextual formation surrounding a technology establishes a gateway, inviting specific publics whilst excluding others. Here, this gateway reinforces the expertise of the cloud provider – expertise offered through the information architectures of data centres and the human resources of security experts. Personal data is entrusted to the professionals.

The dependency between public and private is also evident in questions of data sovereignty and the way it legitimates the move into the corporate cloud. Microsoft, for instance, has taken the lead in developing a homomorphic encryption standard (Microsoft Research, 2017) while also releasing SEAL, a software library that supports it. While SEAL can be embedded into different applications and network configurations, such flexibility belies the practical likelihood that it would operate in data centres with computing power capable of handling homomorphic calculations. In this context, Microsoft's CEO Satya Nadella's public endorsement of FHE in 2018 can be seen as an effort to foreground the importance of privacy precisely in concert with its own highly successful cloud offering. For its part, NTT's San-Shi is envisioned as a privacy toolkit for cloud-providers, encompassing storage, registration of users, delegation of computation to agents and data analysis functions. Both systems argue for a consolidated deployment on data centres, and bind privacy to the platforms that run on them. Encryption, then, is both technical achievement and commercial hinge, underpinning ambitions for market consolidation and reterritorialization, and potentially shifting public trust from the clouds of upstart social media companies to those of incumbent technology firms. Clouded data is not simply the ability to translate mathematical abstractions onto everyday scenarios, but encompasses particular arrangements of research funding, network configurations, protocol standardisation, legal entitlements and delicate enticements to submit institutional data into the safety of the newly secured cloud.

Conclusion

This article has argued that clouded data constructs a political and technological notion of privacy that operates through the intersection of corporate power, computational resources, and the ability to obfuscate, gain insights from, and valorise a dependency between public and private. At an individual level, the obstacles to using and calibrating the parameters of cloud cryptography (and in particular, FHE) point, then, to a limited agency in the control over one's data. Such limits press further on the possibilities to leverage power, and to make claims and demands in a technologically constricted territory of the political. Moreover, by fulfilling (or bypassing) privacy regulations while still serving governmental and market-based interests, cloud cryptography renders collective agency over the control and governance of data more difficult too. Theorists like Gandy (2011) and Morozov (2015) have repeatedly stressed that in determining the commodification and circulation of personal data, this economy shapes politics and exerts significant power.

In responding to the demands of civic and commercial actors, these new cryptographic procedures recondition the sociological imaginary and the political economy of privacy. As they mature, we anticipate they will comprise the heart of efforts to rebuild a shattered public trust of data management by government and corporate institutions. As Facebook hints at a paid tier of its major services (Ellis, 2018), data privacy becomes a discriminating factor, completing a transition in media business models from pay-to-consume to pay-to-stay-private. The processing and network costs of homomorphic encryption and secure MPC will likely be externalised as charges for securitization – a step effectively already taken by the rewards earned by third-party miners in common blockchain implementations. Such developments in turn gesture toward a wildly uneven and unequal future economy of data: the benchmark of privacy set by those who can afford the leasing of computational cycles many orders of magnitude greater than required for unencrypted equivalents, with downward graduations for those with progressively less means, compelled instead to pay for their digital life through third-party monetisation of it.

Finally, the new conditions instantiated by cloud-based encryption seem to shift the conventional understanding of publics. When data can remain clouded and agencies 'never see' the underlying information, then – perversely – privacy becomes a less effective argument for restrictions on data capture or regulations on information use. On an immediate level, this rationalizes more expanded and invasive regimes of data capture. But less obviously, the implied security of this 'always encrypted' data legitimizes its combination and cross-pollination with other datasets. Technically constituted in the moment without the group's knowledge, these 'combinatorial publics' bypass the traditional link between privacy and the individual, forming a kind of ethical loophole. Based on concepts like personal information and a data subject, traditional privacy rights are highly individualized and 'atomistic' (Floridi, 2014). This means that privacy rights and duties do not yet exist for 'algorithmically constructed ad hoc groups' (Mittelstadt, 2017). Indeed a nascent field of 'group privacy' emerging over the last few years has attempted to address the rights of these groups (Taylor et al., 2017). While groups have little control over these profiles, they are both revealing and consequential. Often responding to the concrete demands of a project, the aggregate 'insights' obtainable from such combinatorial publics nevertheless lend them a substantive empirical force, whether leveraged in the commercial arena for business logics or in the civic sphere for legislative policies. Associations of voting preferences with the obscure margins of cultural taste, so well documented in the Cambridge Analytica scandal, highlight the combinatorial affordances of personal data just beginning to be recognized. More work will be necessary to examine exactly how these technologies reconstitute the relation-so central for classical western political thought (Arendt, 1958; Habermas, 1989) – between privacy and the public.

The clouded data condition is part of a rapidly shifting terrain. This still-unfolding technological space, with global economic and political stakes, is not simply a response to social concerns that, once pacified, move to other fields of contestation. Rather it produces a novel imaginary of privacy, with correlate expectations and opportunities for intervention. As more – and more personal – information moves online, and new techniques for exploiting this information emerge, the dichotomy of 'public' and 'private' is (again) challenged. New cloud infrastructures allow data to be shared across sectors and institutions, slipping easily between corporate and state actors. While data takes part in diverse ecologies of power – representation through statistical aggregation, enclosure through encryption, commodification, sharing and hacking – the issue of inviolability of private information and the subject becomes as much a question of political contestation as of technological feasibilities. If the computation of privacy has become newly tractable, the culture of privacy, contested and rapidly shifting, is far from clear.

Acknowledgements

The authors wish to acknowledge the generous support of Dr Laurence Park, of Western Sydney University, and Justin Evans and Tamir Levin of Dimension Data, who provided technical assistance with the San-Shi experiment. In addition, we would like to acknowledge the support of Prof. Ned Rossiter, Prof. Brett Neilson and Tim Horan, who contributed general advice and suggestions for this article.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Our university is a partner with the subsidiary of the manufacturer (NTT) of one of the encryption technologies discussed (San-Shi). We did not receive any funds under this arrangement. However as acknowledged under Funding, the existence of that partnership motivated the university to fund this research.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a small internal grant made available by the University where several of the authors are currently PhD candidates.

ORCID iD

Luke Munn https://orcid.org/0000-0002-1018-7433

Notes

1. The source code and parameter values have been published on GitHub (details supplied on publication).
2. We conducted tests on a server running Ubuntu 16.04, with an Intel Xeon processor E3-12xx at 2.6 MHz and 4GB RAM. We tested compared two functions, average and variance, between unencrypted (supplied by Python *numpy* functions *mean* and *var*) and encrypted (our own 'home-rolled' versions using *PySEAL*'s API). With FHE schemes, there is typically a trade-off between accuracy and performance. This trade-off is managed within

PySEAL through the setting of three parameters: the polynomial modulus, the coefficient modulus and the plaintext modulus (http://130.56.248.129:8889/edit/SEALPythonExamples/examples.py). For calculations with very small numbers, these parameters can be set to small values, which provides a correspondingly low 'noise budget' but better performance. On our hardware configuration, setting polynomial and coefficient modulus to 8192 and the plaintext modulus to 786,433 allowed us to encrypt 25 8-bit numbers – enough to capture blood pressure values – at reasonable speeds. For the calculation of the mean, FHE performance decreased by a factor of 4824 (0.012 vs. 57.577 s, best of three 1000 iterations); for calculation of variance, FHE performance decreased by a factor of 62,834 (0.038 vs. 2,367.476 s, best of three 1000 iterations). These figures are intended to reflect a 'naïve' use of PySEAL's API, and do not constitute a rigorous evaluation. Performance characteristics may vary with other data, parameters, algorithm, hardware and network settings.

3. Based on experimentation with an early (R&D) build of the San-Shi platform.

References

Acar A, Aksu H, Uluagac AS, et al. (2017) A survey on homomorphic encryption schemes: Theory and implementation. *arXiv:1704.03578 [cs]*. Available at: http://arxiv.org/abs/1704.03578 (accessed 19 June 2018)

Amoore L (2014) Security and the Claim to Privacy. International Political Sociology 8(1): 108112. DOI: 10.1111/ips.12044.

Amoore L and Raley R (2017) Securing with algorithms: Knowledge, decision, sovereignty. Security Dialogue 48(1): 310. DOI: 10.1177/0967010616680753.

Amoore L (2018) Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography* 42(1): 4–24.

Archer D (2016) Revolution and evolution: Fully homomorphic encryption. *United States Cybersecurity Magazine*, Summer. Available at: https://www.uscybersecurity.net/csmag/revolution-and-evolution-fully-homomorphic-encryption/ (accessed 19 June 2018)

Arendt H (1958) *The Human Condition*. Chicago, IL: University of Chicago Press.

Australian Bureau of Statistics (2018a) Socio-economic indexes for Australia (SEIFA) 2016. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 (accessed 16 August 2018)

Australian Bureau of Statistics (2018b) 3218.0 – Regional Population Growth, Australia, 2015–16. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/3218.0Main%20Features702015-16?opendocument&tabname=Summary&prodno=3218.0&issue=2015-16&num=&view= (accessed 16 August 2018)

Australian Digital Health Agency (2018) My Health Record. Available at: https://www.myhealthrecord.gov.au/front (accessed 1 July 2018)

Australian Government, Department of Education and Training (2018) *Complete Privacy Policy*. Available at: https://www.education.gov.au/privacy-policy (accessed 30 July 2018)

Avram MG (2014) Advantages and challenges of adopting cloud computing from an enterprise perspective. *Procedia Technology* 12: 529–534.

Bogdanov D, Jõemets M, Siim S, et al. (2015) How the Estonian tax and customs board evaluated a tax fraud detection system based on secure multi-party computation. In: *International conference on financial cryptography and data security*, San Juan, Puerto Rico, 26–30 January 2015, pp. 227–234. Springer. Available at: http://fc15.ifca.ai/preproceedings/paper_47.pdf

Bogetoft P, Christensen DL, Damgard I, et al. (2008) *Multiparty Computation Goes Live*. 068. Available at: http://eprint.iacr.org/2008/068 (accessed 20 June 2018)

Buyya R, Yeo CS, Venugopal S, et al. (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems* 25(6): 599–616.

Çetin GS, Chen H, Laine K, et al. (2017) Private queries on encrypted genomic data. *BMC Medical Genomics* 10(2): 45.

Cohen JE (2019) Turning privacy inside out. *Theoretical Inquiries in Law* 20(1): 1–32.

Coley R and Lockwood D (2012) *Cloud Time*. London: Zero Books.

Culnane C, Rubinstein B and Teague V (2017) *Health Data in an Open World*. 18 December. Melbourne: The University of Melbourne. Available at: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf

De Filippi P (2016) The interplay between decentralization and privacy: The case of blockchain technologies. *Journal of Peer Production* 9. Available at: http://peerproduction.net/issues/issue-9-alternative-internets/peer-reviewed-papers/the-interplay-between-decentralization-and-privacy-the-case-of-blockchain-technologies/ (accessed 30 July 2018)

De Montjoye Y, Radaelli L and Singh VK (2015) Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* 347(6221): 536–539.

Druckman JN and Kam CD (2011) Students as experimental participants. In: *Cambridge Handbook of Experimental Political Science*. Vol. 1. Cambridge: Cambridge University Press, pp. 41–57.

Dunlevy S (2016) Encrypted private medical records released by the Department of Health are vulnerable. Available at: https://www.news.com.au/lifestyle/health/encrypted-private-medical-records-released-by-the-department-of-health-are-vulnerable/news-story/a2f1dc892102cfaa5d1915dd32ad98d8 (accessed 2 July 2018)

Dwork C (2006) Differential privacy. In: Bugliesi M, Preneel B, Sassone V, et al. (eds) *Automata, Languages and Programming*. Berlin/Heidelberg: Springer, pp. 1–12.

Dwork C and Roth A (2014) The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4): 211–407.

e-Health Strategy (2018) *eHealth Strategy for NSW Health 2016–2026*. Sydney: NSW Health. Available at: http://www.health.nsw.gov.au/eHealth/Documents/eHealth-Strategy-for-NSW-Health-2016-2026.pdf

Ellis C (2018) Facebook could introduce a paid subscription service. *Techradar*, 11 April. Available at: https://www.techradar.com/au/news/facebook-could-introduce-a-paid-subscription-service (accessed 25 July 2018)

European Union (2018) Article 17 – Right to erasure ('right to be forgotten'). Available at: http://www.privacy-regulation.eu/en/article-17-right-to-erasure-'right-to-be-forgotten'-GDPR.htm (accessed 20 June 2018)

Finextra Research and IBM (2016) *Banking on Blockchain: Charting the Progress of Distributed Ledger Technology in Financial Services*. January. London: Finextra ResearchAvailable at: https://www.finextra.com/finextra-downloads/surveys/documents/32e19ab4-2d9c-4862-8416-d3be94161c6d/banking%20on%20blockchain.pdf (accessed 2 July 2018)

Floridi L (2014) Open data, data protection, and group privacy. *Philosophy & Technology* 27(1): 1–3.

Foucault M (2002) *The Birth of the Clinic*. Abingdon: Routledge.

Foucault M (2007) *Security, Territory, Population: Lectures at the Collège de France, 1977–78*. Dordrecht: Springer.

Foucault M (2012) *Discipline and Punish: The Birth of the Prison*. London: Vintage.

Gallagher S (2018) Equifax breach exposed millions of driver's licenses, phone numbers, emails. Available at: https://arstechnica.com/information-technology/2018/05/equifax-breach-exposed-millions-of-drivers-licenses-phone-numbers-emails/ (accessed 20 June 2018)

Gandy O (2011) The political economy of personal information. In: Wasko J, Murdock G and Sousa H (eds) *The Handbook of Political Economy of Communications*. Malden, MA: Wiley-Blackwell.

Gemalto (2018a) Data breach statistics by year, industry, more. Available at: https://breachlevelindex.com (accessed 20 June 2018)

Gemalto (2018b) 2017 Data Breach Level Index: Full year results are in. Available at: https://blog.gemalto.com/security/2018/04/13/data-breach-stats-for-2017-full-year-results-are-in/ (accessed 20 June 2018)

Gentry C (2009) *A fully homomorphic encryption scheme*. Dissertation. Stanford University, Stanford, CA, USA. Available at: https://crypto.stanford.edu/craig/craig-thesis.pdf

Greenberg A (2009) IBM's blindfolded calculator. Available at: forbes/2009/0713/breakthroughs-privacy-super-secret-encryption (accessed 20 June 2018).

Habermas J (1989) *The Structural Transformation of the Public Sphere*. Cambridge: Polity.

Halevi S (2018) *HElib: An implementation of homomorphic encryption*. C++. Available at: https://github.com/shaih/HElib (accessed 19 June 2018)

Hallman RA, Diallo MH, August MA, et al. (2018) *Homomorphic Encryption for Secure Computation on Big Data*. Setúbal: SCITEPRESS – Science and Technology Publications, pp. 340–347.

Hautala L (2018) Can Facebook's new hires take on troll farms and data privacy? Available at: https://www.cnet.com/news/can-facebook-mark-zuckerberg-new-hires-take-on-troll-farms-and-data-privacy-after-cambridge-analytica/ (accessed 20 June 2018)

Hewitt A and Forte A (2006) Crossing boundaries: Identity management and student/faculty relationships on the Facebook. Poster presented at CSCW, Banff, Alberta, pp. 1–2.

Hu TH (2015) *A Prehistory of the Cloud*. Cambridge, MA: MIT Press.

Kishore S (2018) PySEAL: Homomorphic encryption in a user-friendly Python package. Available at: https://gab41.lab41.org/pyseal-homomorphic-encryption-in-a-user-friendly-python-package-e27547a0b62f (accessed 8 July 2018)

Kuulasmaa K, Hense HW and Tolonen H (1999) Quality assessment of data on blood pressure in the WHO MONICA Project. WHO MONICA Project e-publications, No. 9.

Lapets A, Dunton E, Holzinger K, et al. (2015) *Web-based Multi-party Computation with Application to Anonymous Aggregate Compensation Analytics*. Boston, MA: Computer Science Department, Boston University. Available at: http://www.cs.bu.edu/techreports/pdf/2015-009-mpc-compensation.pdf

MacKenzie A (2018) Personalization and probabilities: Impersonal propensities in online grocery shopping. *Big Data & Society* 5(1): 1–15.

McSherry F (2018) Lunchtime for data privacy. Available at: https://github.com/frankmcsherry/blog (accessed 27 June 2018)

Mearian L (2018) Will blockchain run afoul of GDPR? (Yes and no). Available at: https://www.computerworld.com/article/3269750/blockchain/will-blockchain-run-afoul-of-gdpr-yes-and-no.html (accessed 20 June 2018)

Microsoft Research (2017) *Homomorphic encryption standardization workshop*. Redmond, WA. Available at: https://www.microsoft.com/en-us/research/event/homomorphic-encryption-standardization-workshop/ (accessed 27 June 2018)

Mittelstadt B (2017) From individual to group privacy in big data analytics. *Philosophy & Technology* 30(4): 475–494.

Montgomery CM and Pool R (2017) From 'trial community' to 'experimental publics': How clinical research shapes public participation. *Critical Public Health* 27(1): 50–62.

Morozov E (2015) Socialize the data centres! *New Left Review* 91(II): 45–66.

Nadella S (2018) *Microsoft CEO: We're Focused on 3 Core Pillars*. Seattle. Available at: http://fortune.com/video/2018/05/07/microsoft-ceo-were-focused-on-3-core-pillars/ (accessed 4 July 2018)

Narayanan A and Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: *IEEE symposium on security and privacy*, Oakland, CA, USA, 18–21 May 2008, pp. 111–125. IEEE.

Near J (2018) *Differential privacy at scale*. Santa Clara, CA, USA. Available at: https://www.youtube.com/watch?v=pk_DCSUayDA (accessed 20 June 2018)

Newman L (2017) 6 fresh horrors from the equifax CEO's congressional hearing. Available at: https://

www.wired.com/story/equifax-ceo-congress-testimony/ (accessed 20 June 2018)

NTT (2017) Performing a statistical analysis of multiple companies' sensitive data. Musashino Research and Development Center, Tokyo. Available at: http://www.ntt.co.jp/RD/active/201702/en/pdf_eng/03/C-28_e.pdf (accessed 4 July 2018)

Office of the Australian Information Commissioner (2018) Mandatory data breach notification comes into force this Thursday. Available at: /media-and-speeches/media-releases/mandatory-data-breach-notification-comes-into-force-this-thursday (accessed 2 July 2018).

Ohm P (2010) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 57: 77.

Sakr S, Elshawi R, Ahmed A, et al. (2018) Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. *PloS One* 13(4): e0195344.

Schneier B (2009) Homomorphic encryption breakthrough. Available at: https://www.schneier.com/blog/archives/2009/07/homomorphic_enc.html (accessed 28 June 2018)

Shannon CE (1949) Communication theory of secrecy systems. *Bell System Technical Journal* 28: 245–249.

Stilgherrian (2018) My Health Record systems collapse under more opt-outs than expected. Available at: https://www.zdnet.com/article/my-health-record-systems-collapse-under-more-opt-outs-than-expected/ (accessed 8 May 2019).

Tanaka S et al. (2017) Secure statistical computation system on encrypted data. In: UNECE Work session on Statistical Data Confidentiality, Skopje, Macedonia, 20 September 2017. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/6_secure_computation_system.pdf

Tanner A (2013) Harvard professor re-identifies anonymous volunteers in DNA study. Available at: https://www.forbes.com/sites/adamtanner/2013/04/25/harvard-professor-re-identifies-anonymous-volunteers-in-dna-study/ (accessed 20 June 2018)

Taylor L, Floridi L and van der Sloot B (2017) Introduction: A new perspective on privacy. In: Taylor L, Floridi L and van der Sloot B (eds) *Group Privacy: New Challenges of Data Technologies*. Philosophical Studies Series. Cham: Springer International Publishing, pp. 1–12.

Titus A (2018) *PySEAL*. C++. Lab41. Available at: https://github.com/Lab41/PySEAL (accessed 27 June 2018)

Van Es K, Wieringa M and Schäfer MT (2018) Tool criticism: From digital methods to digital methodology. Datafied Society Working Paper Series. 28 May. Available at: https://datafiedsociety.nl/tool-criticism (accessed 24 July 2018).

NSW Ministry of Health (2016) *eHealth Strategy for NSW Health 2016–2026*. Sydney. Available at: http://www.health.nsw.gov.au/eHealth/Documents/eHealth-Strategy-for-NSW-Health-2016-2026.pdf

Wang W, Hu Y, Chen L, et al. (2012) Accelerating fully homomorphic encryption using GPU. In: *2012 IEEE conference on high performance extreme computing*, Waltham, MA, USA, 10–12 September 2012, pp. 1–5.

Wood T (2017) Secure MPC at Google. In: Bristol Cryptography Blog. Available at: http://bristolcrypto.blogspot.com/2017/01/rwc-2017-secure-mpc-at-google.html (accessed 20 June 2018)

Zyskind G (2017) Computing over encrypted data. Enigma blog. May 29, 2017. Available at: https://blog.enigma.co/computing-over-encrypted-data-d36621458447 (accessed 20 June 2018).

Zyskind G, Nathan O and Pentland A (2015) Decentralizing privacy: Using blockchain to protect personal data. May, pp. 180–184. IEEE.