 doi:10.20944/preprints201908.0239.v1

Ethical Access Control in the Era of Data Breaches

A.S.M. Kayes, M.J.M. Chowdhury, Fatma Mohammed, Alex Ng, Paul A. Watters, and Patrick Scolyer-Gray
La Trobe University, Melbourne, Australia
{a.kayes, m.chowdhury, f.mohammed, alex.ng, p.watters, p.scolyer-gray}@latrobe.edu.au

Abstract

The worldwide interconnected objects, called Internet of Things (IoTs), have been increasingly growing in the last several years. Different social media platforms and devices are continuously generating data about individuals and facilitate the technological and the social convergence of their Internet-based data and services with globalized users. These social and device-related IoTs create rooms for data breaches as such platforms provide ability to collect private and sensitive data. We assert that data breaches are fundamentally failures of access control - most users are too busy or technically ill-equipped to understand access control policy expressions and decisions. We argue that this is symptomatic of globalised societies structured by the conditions of algorithmic modernity; an era in which our data is increasingly interdependent on, and enmeshed with, ever more complex systems and processes that are vulnerable to attack. Ethically managing data breaches is now too complex for current access control systems, such as Role-Based Access Control (RBAC) and Context-Aware Access Control (CAAC). These systems do not provide an explicit mechanism to engage in decision making processes, about who should have access to what data and when, that are involved in data breaches. We argue that a policy ontology will contribute towards the development of Ethical CAAC better suited to attributing accountability for data breaches in the context of algorithmic modernity. We interrogate our proposed Ethical CAAC as a theoretical construct with implications for future policy ontology models and data breach countermeasures. An experimental study on the performance of the proposed framework is carried out with respect to a more generic CAAC framework.

Keywords: *Ethical access control, Context-aware access control, Data breaches, Responsibility model, Policy model, Cost model*

1. Introduction

The dynamism of the Internet that is intrinsic to the Digital age has involved the production and consumption of tremendous quantities of user-generated data, especially via social media platforms. Along with the development of this corpus of big data, online social media platforms (such as Facebook, Instagram, Twitter, LinkedIn etc.) facilitate the technological and the social convergence of their Internet-based services with globalized users. In recent years, open data sharing platforms have connected users and communities, through services, such as publishing and sharing user-generated data and information resources, including photos, stories and videos. Often, this sharing has occurred without much thought given to inappropriate data disclosure or unauthorized data access. The unprecedented success of social media represents one of the many 'runaway engines' of late modernity, and have already demonstrated that they are routinely involved in major data breaches. For example, Facebook was forced to acknowledge that millions of American users had been farmed for their private information, which was then sold and used by a third-party political firm to influence a series of US elections, thus demonstrating the manifold forms of data breaches afforded by the conditions of contemporary society [1] [2].

The traditional data warehouse [3] and most recent data lake [4] technologies have been used to implement the big data servers, such as public, private and hybrid cloud centres [5]. These big data servers have been posing serious challenges with regards to data and information sharing, access control, privacy violation and relevant service agreements. One of the important issues revolved around here is *how to educate end users about the importance of privacy for the IoTs* [6], as well as how *to increase privacy-awareness in today's dynamic era of social IoTs* [7]. Indeed, these existing service/business models need to be revised to deal with the dynamic aspects of data, particularly to maintain the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

trust of users and improve the quality of decision making capability when confronted by the need to integrate users' data from multiple sources [8].

Data breaches have thus become an unwelcome but very familiar feature of late-modern societies [9]. Newspaper reports are filed on a near daily basis in which the exploits of malicious actors (state-sponsored, cybercriminals, terrorists) are described and explained to an immense (and globalised) online public. The epidemiology of knowledge is an issue in and of itself, publicity compromising purportedly 'secure', private, and even state networks and data-theft *en masse* are now commonly perceived by the public as if such behaviours are 'the natural order of things'.

In general, one of the big challenges is *to identify who is responsible for what actions and subsequently who is accountable for the consequences of certain actions*. The ethical responsibility of relevant parties involved in the social scenarios are also need to be identified [10]. How *to estimate different tangible and intangible costs of data breaches* is also an associated research challenge, considering the consequences of a data loss or a data breach (e.g., loss of reputation) [11].

In the literature, a significant number of access and privacy control solutions have been proposed to control users' access to data and information resources, for example, context-aware access control mechanisms specific to manage data (e.g., [12]), privacy-aware role-based mechanisms specific to manage sensitive data (e.g., [13]). The security and privacy policy models are the basic contributions of these mechanisms, which are the main cornerstones of today's data-driven organizations, ranging from database management systems to operating systems.

In this context of impotent data security and ambiguous privacy policies, we see a problem in and of itself, but also an accompanying issue of remediation, the same complexities of algorithmic modernity [14] that have made individual interrogations of our seemingly impenetrable and interdependent nebulas of technologies so difficult to navigate have also obfuscated the attribution of accountability for instances of data breaches. In short, another subsequent challenge is *who is responsible for what data, when, for how long, for what purposes and how can that responsibility be codified in a suitably transparent and ethical manner?*

1.1. The Contributions

The contributions of this work are mainly two-folds. A new generation of *Ethical Access Control Approach* is proposed, combining the benefits of existing access control policies with elements of a traditional model of responsibility attribution to estimate the cost against data breaches. We then propose a policy ontology and a reasoning model, extending a relevant earlier ontology-based CAAC policy framework [15] [16], incorporating the responsibility model into the data access policies. In addition, we demonstrate a prototype architecture along with an empirical evaluation of our proposed ethical access control solution.

1.2. Background

For those who do not wish to exploit their employees through forcing them to engage in criminal activities such as corporate espionage need not fear falling behind in the polls. A wealth of valuable data can simply be purchased (or stolen) from social media platforms - analytics companies such as Cambridge Analytica [2] that have similarly taken advantage of algorithmic modernity at tremendous expense to the global public's perception of the efficacy of data-security measures that have (or have not) been deployed, as well as the credibility and ethical integrity of politicians and major corporate entities alike. Although expensive to the end-users, the current order of things ensure extraordinary returns for investors and an elite group of stakeholders who have achieved a kind of wealth that is of a scope well beyond that which can be quantified by pecuniary measures.

Giddens originally described late-modernity as a "Juggernaut"; something that pushes everything out of its path "a runaway engine of enormous power which, we can drive to some extent, but which also threatens to rush out of our control" [17]. Giddens' conceptualisation of modernity is based firmly in a critique of various political-economic systems and associated characteristics of humanity's current epoch. Many elements of Giddens' characterisation still resonates with the conditions of modernity faced by globalised society at the time of writing. What Giddens considered key elements of a late-modern society included (among other things) pluralism, social, cultural and political flux, perpetual and intractable *risks* (perceived and actual), interdependency via globalisation, and a return to positivist epistemologies as the dominant intellectual paradigm [17] [18].

Although Giddens' 'runaway engine' was intended as a metaphorical device to conceptualise and describe his perception of the capitalist treadmill of production and consumption, it is also how he characterised late-modernity's 'high-point'; evidently, that high-point has not yet been reached, and our runaway engine has yet to 'derail'. Thus, we propose updating Gidden's perceptions of late-modernity to incorporate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

the practical realities of a society that now relies on an exponentially increasing number of interdependent digitally networked technologies, of which each are inherently *algorithmic* in the sense that they rely on precise and predefined, but also complex, hidden and/or inaccessible sets of *reasoning rules* to function. There may never be such symmetry or synergy between the founding technologies of society and their respective processes of rationalisation as elaborate as those featured in society today.

In the context of algorithmic modernity, the victims as well as those responsible for data breaches are both overwhelmed, outnumbered and outgunned. For example, a plumber might have overheard enough of a conversation taking place in an adjacent bathroom stall to learn the information required that would authorise *administrator*-level access to the employing organisation's payroll system. Similarly, politicians, as well as just about anyone who has set foot in the West Wing, are more than likely to be exposed to sensitive or even classified information, as well as its digital location. However, we cannot assume that in either of these cases that users will automatically react by attempting to access and exploit that data. We cannot expect all users to have the energy after a day of hyper-rationalised work to actually attempt data-exploitation, but history and recent events have underscored the point that we must expect this behaviour from *some*.

Just the same as it was when Simmel wrote *Metropolis* [19], our societies are indeed overwhelmingly complex, and they are certainly emotionally 'draining' to say the least. The rise in urban populations and population density in cities is yet another key aspect of late modernity that not only cultivates alienation, but also promotes societal fragmentation (individualisation). The highly-specialised labour markets of late-modern societies creates new jobs dedicated to the maintenance, sales, and distribution of ICTs. Once the technological paradigms of technology giants like Microsoft and Apple become normalised, each company hired their own armies of specially trained technicians and sales staff, the confluence of which can easily make the already individualised and alienated residents of algorithmic modernity feel that their opinions about and perceptions of technology represent a *minority* and *inferior* view relative to that of the overwhelming majority. The consumer is thus made to feel, but also perceive themselves as not only outnumbered, but also as *outgunned* by innumerable and individualised sets of assumptions about the expertise of ICT workers, of whom few are incentivised to do anything but *encourage* such perceptions for the sake of maintaining compliant and deferential consumers.

Overwhelmed and outnumbered by a perceived majority, users become situated in an increasingly severe position of vulnerability attuned to self-imposed assumptions of inadequacy - an uncooperative PC or laptop is no longer understood as something that the consumer is capable of fixing without the assistance of expert knowledge reserved for ICT specialists. These assumptions about the public's inability to learn how to manage its own technology stems from the additional assumption that the public is not up to date with the most recent developments in computer engineering. The hyper-rationalisation of algorithmic modernity leaves no scope for the individual agent to perform even the most basic diagnostics on their own computers, which we must remember are now configured discursively to be perceived as mystical devices of complexity well beyond the comprehension of a lay public. Instead, the management and maintenance of ICT systems, which are now inclusive of the operational trivialities and the mundane technical difficulties encountered in everyday life, are transformed into job 'tickets' that are 'resolved' rather frequently with the delivery of a brand new machine.

Of course, this is partly the result of "planned obsolescence", but the point is to underscore some of the entrenched techno-social predilections in algorithmic modernity that make users more vulnerable to cyber attacks that rely on data exploitation than ever before. Consider the user who believes that s/he is unable to comprehend how their device works, imagine the scope for manipulation and the ease with which elicited self-disclosures of personal and otherwise valuable data can be collected. In the absence of any knowledge about computers at all, then *why should a user be surprised when they are told by a criminal masquerading as technical support that their laptop now requires them to type their social security number and mother's maiden name as 'routine' steps towards the provision of a technical solution to the end-user's ostensibly otherwise intractable problem?*

Put simply, as users become less-equipped to manage their own data, vulnerabilities for exploitation and data breaches increase. One of many such vulnerabilities stem from new opportunities for technologically inclined political operatives (or those who know someone who is technologically skilled and that they can control by legal or extralegal means, which would include practically all major politicians around the world). Politicians (and their unsuspecting, coerced, or naive proxies) now engage in the same business-as-usual and ethically unsavoury, mercenary,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

and legally ambiguous (but nevertheless lucrative) activities such as opposition research and redistricting tactics, but with all the affordances of convergence culture. Centralised databases of census data, financial records as well as law-enforcement activities and reports are all easily accessed and manipulated to suit the purposes of a suitably motivated individual.

1.3. Paper Outline

The rest of the paper is organized as follows. Section 2 presents the motivation of this research along with a relevant state-of-the-art research work. Section 3 presents the proposed ethical access control approach, including the responsibility and cost models. Section 4 demonstrates a walkthrough of the proposed theoretical access control approach via the policy ontology, a prototype and the evaluation of our proposal. Finally, Section 5 concludes the paper and outlines future research.

2. Research Motivation and the State-of-the-Art

In order to motivate this research, we briefly present the research issues and challenges related to current state-of-the-art access and privacy control models and frameworks - ethical access control, data breach, responsibility and accountability.

2.1. Lack of Security and Privacy Policies

Digital and social networking platforms continue to pose many issues and challenges due to peer-to-peer interactions. One of the serious issues is related to security and privacy threats that are usually associated with data breaches (e.g., disclosure of unauthorized data and sensitive information) [20]. The users store their massive data in the data stores utilizing the recent advancement of cloud computing [5]. However, security and privacy aspects of such cloud servers draw much attention due to several vulnerabilities, like data stored in the public cloud is insecure [21].

One important issue is, *how to protect such data and information resources from an unauthorized access using relevant security (access control) and privacy control policies*. In addition, *how the organizations or service providers can take proper countermeasures (i.e., proactive steps) to control data by estimating relevant costs against data breaches?*

2.2. Data Breaches and Our Dependence on Brittle Information Systems

Privacy violations on social media platforms, including data breaches on Facebook [1] [2], have become increasingly frequent. Revealing private information held by social media organisations as well as any instance of unauthorized data access is constitutive of a data breach. Privacy breaches can also occur through third-party applications that can be accessed on any social networks [22]. There is a need to protect privacy while users are online, and this raises many data management challenges, and underscores the need for appropriate access and privacy control agreements between the service providers and other parties.

It is important not to lose sight of behavioural tendencies that all too often undermine even the finest and most comprehensive technological solutions to cyber security problems. In this case, we must be mindful of the fact that all humans are walking information systems; organic databases with innumerable potentially exploited attack vectors arranged across an immense 'attack surface'. For example, political candidates on campaign can be found buried under layers of information management and data security systems, and if they can be reached and compromised by an attacker, the information security architecture of the candidates are rendered useless. We have seen this demonstrated repeatedly, such as in the example of John Podesta, whose failure to recognise a phishing email resulted in a tremendous data breach that severely damaged Hilary Clinton's campaign [23].

Podesta is demonstrative of the fragility of the information systems upon which we rely to protect our data, all that was required of Podesta to significantly alter the trajectory of a political campaign of supreme importance was an uncritical and perhaps unexamined decision to click on a falsified Gmail credentials reset link [23]. The key insight to take from this example is less about incompetence, insufficient training, incident response plans, or even technological countermeasures. It is the simple reality that in all likelihood, Podesta, like anyone else living under the hyper-rationalised conditions of algorithmic modernity, could not risk losing access to critical campaign communications infrastructure. Podesta was obviously under substantial pressure already, and was likely attempting to manage multiple networked devices and countless online accounts simultaneously (in much the same way as a substantial proportion of globalised society). Put simply, Podesta presented an easy target and was exploited using basic social engineering

techniques, but his vulnerability stemmed from the practical impossibility of sustaining a consistent critical focus on each and every device, application and interface with which he interacted.

Even if Podesta had evaded that phishing attempt, our observation of algorithmic modernity tells us we can reasonably expect that the same outcome would have been achieved eventually. Podesta's technological attack surface was so vast that attackers would have found an exploit sooner or later, and it is at this juncture in the case of Podesta that we see the necessity of recognising and critically examining the intersecting roles of human and technological factors that are involved in data breaches. The techno-social structural conditioning of algorithmic modernity has cultivated a culture and a socio-political environment in which the management of security control has become a task that exceeds the abilities of the vast majority of residents.

When hackers breached Apple's iCloud service and successfully distributed hundreds of compromising photographs of celebrities via various social media platforms in 2014, globalised publics briefly shifted their attention to matters of data-security and the vulnerabilities of cloud-based storage [24]. In this watershed moment, the technological approach of brute force password cracking converged with cultural practices that have continued to socialise globalised publics into a state of semi-ambivalence: easy-to-crack simplistic passwords are certainly not encouraged, but to the same extent, their creation is hardly discouraged [24]. The structural forces that undergird algorithmic modernity requires users maintain control over their exponentially increasing number of user-accounts, and at least some measure of agency is exercised by users whenever they participate in their selection of access credentials.

However, in the wake of a seemingly endless stream of catastrophic data breaches that are only becoming more frequent and spectacular in scope, we have not seen an equivalent increase in public outrage, and nor have we seen the subject of data-security as a matter of public discourse achieve proportionate prominence in the public sphere. Instead, what we have seen is a kind of culturally inscribed ambivalence over data-security. When lifeworlds become rationalised to such extremes, as is the case in algorithmic modernity, even the time taken to type a single keyboard character has a price. Thus, we have seen a rise in structural incentives to avoid selecting strong passwords that, once combined with the defeatism cultivated by self-imposed assumptions of computer literacy inadequacies, has placed globalised public in a position of considerable vulnerability.

Every user revisits the question of how much they are willing to invest in the security of their data on a regular basis. A discourse of the layman's inadequate computer literacy is a cultural artefact that is routinely refined and reconstructed through everyday conversations in the workplace, in films and TV shows, in the online blogs and forums, and other social media commonly consumed, and through participation in deliberation with peers and family.

2.3. Lack of Accountability

Research aimed at modeling responsibility and accountability is scarce in extant literature (e.g., [10] [25]), and the majority of these models are (with a handful of promising exceptions - the Context-Aware Access Control (CAAC) frameworks [12] [15] [8]), mostly ineffectual in today's dynamic network society. With assets distributed globally, social networks of involved actors just as likely to be readily identified as they are deliberately concealed, as well as the considerable variables raised by the socio-culturally diverse populations intrinsic to a globalised pluralist society, we can see how algorithmic modernity is a structural impediment that confounds attempted estimations of the cost of damages incurred as a result of data breaches.

True to the demands of algorithmic modernity, relentless, recent and ongoing advancements in digital and social platforms have rendered many current models of accountability attribution [10] [25] and empirical approaches to estimating the risks of data breaches [26] [27] obsolete. Layton and Watters [11] introduced a simple model for efficient and effective estimations of tangible and intangible costs involved in data breaches.

Although some have already estimated the costs of data collection, processing (e.g., localization, encoding) and quality assurance (e.g., [28]), the available literature indicates that there is still a scarcity of research into new and suitably holistic strategies for estimating the associated costs that are involved in data breaches. To this body of research, we propose further developing an ethical CAAC framework by integrating a new model for responsibility attribution for decision making (i.e., who can access what data for what purposes) that will be paired with a cost model for estimating a data breach.

2.4. The State-of-the-Art Access Control Systems

Access Control Systems (ACSs) are computer programs that rely on 'policies' to automatically award (or refuse) access authorisation credentials to data and information resources. Policies are 'rules', principles

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

and protocols that are created, indexed, and organised by system administrators, and subsequently permit ACSs the ability to autonomously execute various processes such as monitoring the identities of users and recording who has (and previously had) access to the data associated with the ACS, along with when, for what purposes and for how long the data were accessed.

It is necessary to identify who is responsible in order for a culpable party to be held accountable [29] for making unethical (see our proposed ethical access control in *Section 3*) access control decisions about sensitive personal data. In this respect, we propose taking full advantage of the ACS that are embedded throughout most information systems as well as the privately owned computing hardware of globalised public in order to implement an ACS strategy based on advanced end-to-end "nonrepudiation techniques" drawn from the "audit and accountability" family of security controls [30]. The viability of this aspect of our research is supported by its remarkable symmetries with past and successfully implemented ACS strategies designed for the management of Internet of things [31].

A family of Context-Aware Access Control (CAAC) models suitable for managing data from multiple types of data sources [12] [15] [8] [32] [33] has been introduced in the last few years. However, in the face of algorithmic modernity's dynamism, we have found that the majority of existing ACSs are following similar trajectories towards incompatibility (e.g., vulnerability scanning policies that impede productivity and contradict the founding political-economy upon which algorithmic modernity relies) and/or obsolescence (e.g., removable media control policies may become pointless if paradigmatic shifts in information system hardware design no longer requires the inclusion of access ports).

3. The Proposed Ethical Access Control Approach to Countermeasure Data Breaches

In this section, we model a broader notion of responsibility with regards to protecting against data breaches - *Ethical Access Control Approach*.

3.1. The Responsibility Model

Figure 1 illustrates the digital and social platforms and the associated business models, comprising different parties - such as service provider, data owner, data analyst, data sharer, data scientist, marketing people, and law enforcement people. We consider the basic responsibilities of different individuals/parties. For example, a data owner, who can be either a user, service provider or both, is a party who normally owns

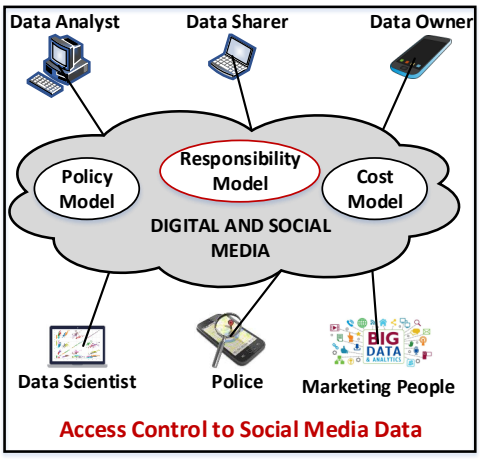

Figure 1. Access Control to Social Media Data

the data.

Before going to formally express the responsibility model, we can go back to a relevant earlier research on situation-aware access control policy model [33] [34], in which,

- '*a situation is defined as the states of the relevant entities and the specific purpose*', and
- '*a purpose is defined as the user's intention in accessing data or information resources*'.

For example, a paramedic should have the right to access a patient's emergency medical records when there is a potential life-threatening situation is detected.

In another earlier work [32],

- '*a context is defined as any dynamic conditions that are used to characterize the states of relevant entities*'.

When the context changes (e.g., the patient's health condition becomes normal as soon as the critical situation is an under-control), the paramedic should not have the right to access the patient's health records for treatment purpose.

Following earlier research on context/situation-awareness [33] [34] and access control [32], we define the broader notion of responsibility.

> **(Definition 1 - Responsibility Attribution)**. *The responsibility is defined as the role of any involved party and the associated actions for particular purposes, and the consequences of certain actions is defined as accountability.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

The notion of responsibility includes mainly two parts. On the one hand, it includes who is responsible for accessing (e.g., publishing and sharing) data for any relevant purposes. On the other hand, it also includes who is accountable for any consequences of certain actions. Using this broader notation of responsibility attribution, our goal is to control data against authorized versus unauthorized parties, by restricting such parties to access relevant data for relevant purposes.

> **(Example 1)**. *Let us consider a data access policy from a relevant earlier work [32]. A paramedic is allowed and responsible to access emergency medical records of a patient for giving treatment and saving his life from a critical heart attack, when they are co-located at the same place.*

3.2. The Cost Model for Estimating Data Breaches

In our ethical access control approach, on the one hand, we have taken into account a broader notion of responsibility attribution for controlling 'who can access what data for what purposes and for how long' (e.g., considering the contexts - temporal, spatial, relationship, and purpose-based). On the other hand, the accountability also has been considered with respect to estimating relevant costs that are involved in data breaches.

For the sake of flexibility in application and to prevent any confusion over an imposed normative agenda, we emphasise that the only ethical expectation that we have aimed to enforce is the near universal premise that suffering should be avoided wherever possible, and that inflicting suffering on others is always morally suspect.

We adapt an earlier cost estimation treatment [11] to this research in determining the accountability of a data breach. A data breach can occur while there is a data loss, theft or unauthorized access to users' personal information [35] [36].

> **(Definition 2 - Data Breach)**. *A data breach can be defined as an unauthorized access to users' confidential or sensitive information without any consent from them.*

> **(Example 2)**. *As discussed before, a paramedic is not allowed to access the medical history of the patients. However, such records can only be accessed for an emergency treatment purpose when there is a critical situation has been detected.*

According to a relevant earlier research [11], the cost estimation model should have two basic parts.

- The *tangible costs* are directly associated against a budget (e.g., costs of involving security administrators for specifying relevant access control policies).
- The *intangible costs* are associated with any future consequences of a data breach (e.g., the stolen credentials are seen to be used for other purposes).

Some examples of **tangible costs** include.

- $InvestCost$ - Direct cost of investigating the case of an unauthorized data access.
- $PolicyCost$ - Direct cost of specifying relevant access control policies that are associated with such an unauthorized case, including all the teams involved, such as security administrator, application developer, and so on.
- $NotifyCost$ - Direct cost of notifying users/customers that their data has been accessed purposelessly, without a permission.

According to Gibson [30], **intangible costs** are those which are not directly related to the actual cost of a physical or otherwise material asset (i.e., capital budget). Examples provided include projections of lost revenue, damaged reputations and similar issues such as the consequences of defamation, all of which are again linked by Gibson to yet more intangible assets such as "client confidence" and "customer influence". Notably, **intangible costs** are those most readily recognised as 'abstract', and the conceptual connotations suggest a lack of substance or meaning. We feel it is worth underscoring that the most damage wrought by a data breach will typically stem from the intangible costs rather than the tangible. For example, data breaches that eventually links a brand with stigmatised criminal activities will result in an irretrievable trajectory of devaluation. However, we contest the claim that so-called 'intangible' costs such as a brand name are not 'directly related' to economic capital.

Logical inconsistencies in conceptualisations of computer-mediated activities as "virtual" [37] provides a precedent for our argument that every data breach involves at least some degree of abstraction. This is because every case of data breach revolves around unauthorized user interactions with data, which means that even the most tangible data breaches, such as the retrieval and illegal sale of archived government email correspondence by a system administrator to a journalist will inevitably require the translation of data from empirically falsifiable (i.e., tangible) strings of binary

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

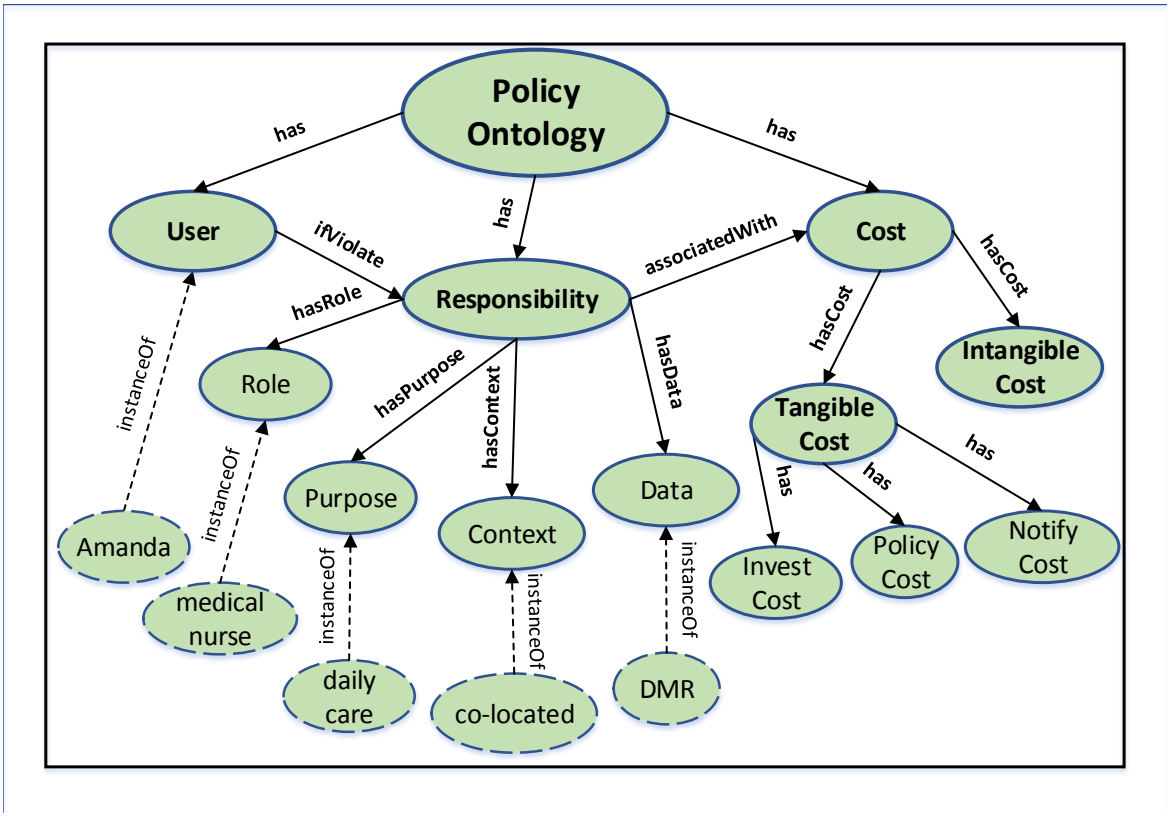

Figure 2. The Ethical CAAC Policy Ontology

code into another codified symbolic, intangible system. The assignation of pecuniary value to an email requires that all participating parties first accept as valid the abstract idea of currency [37] before they may engage with all the abstractions involved in the information economy and news-media industries to which journalists are bound and, in this hypothetical example, motivated the data breach in the first place.

4. Policy Ontology and Evaluation

In this section, we present an ontology-based ethical CAAC policy framework to provide the practical basis for realizing the proposed broader notion of responsibility attribution. In addition, we demonstrate the practicality of the proposed solution through revisiting a case study from a relevant earlier research [8]. Experience from earlier research [8] [32] [33], in this work we use ontology languages OWL and SWRL, to specify ethical access control policies for representing responsibility and to codify reasoning rules for estimating costs of a data breach.

4.1. Policy Ontology

Figure 2 shows our policy ontology extending a relevant earlier CAAC framework [15] [16]. In the policy ontology the responsibility of all involved parties, based on the implications of their roles and purposes of data access (i.e., individual portions of ethical responsibility), is incorporated. Accountability for the consequences of data breaches as a result of unauthorized data access is also incorporated through a flexible cost model based on assessments of net losses measured.

The *Policy Ontology*, as depicted in Figure 2, has the following basic concepts or classes: $User$, $Responsibility$ and $Cost$. An object-type property is used to connect between two classes, and a data-type property is used to connect between a class and its attributes. For example, a core class $Responsibility$ is defined by four other classes, and an object property, named has, is used to connect the domain concept $Responsibility$ and its four range concepts $Role$, $Purpose$, $Context$ and $Data$. The $Role$ is represented by the role identity ($roleIdentity$), which is a data-type property. For simplicity, we do not show all data-type relationships in Figure 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

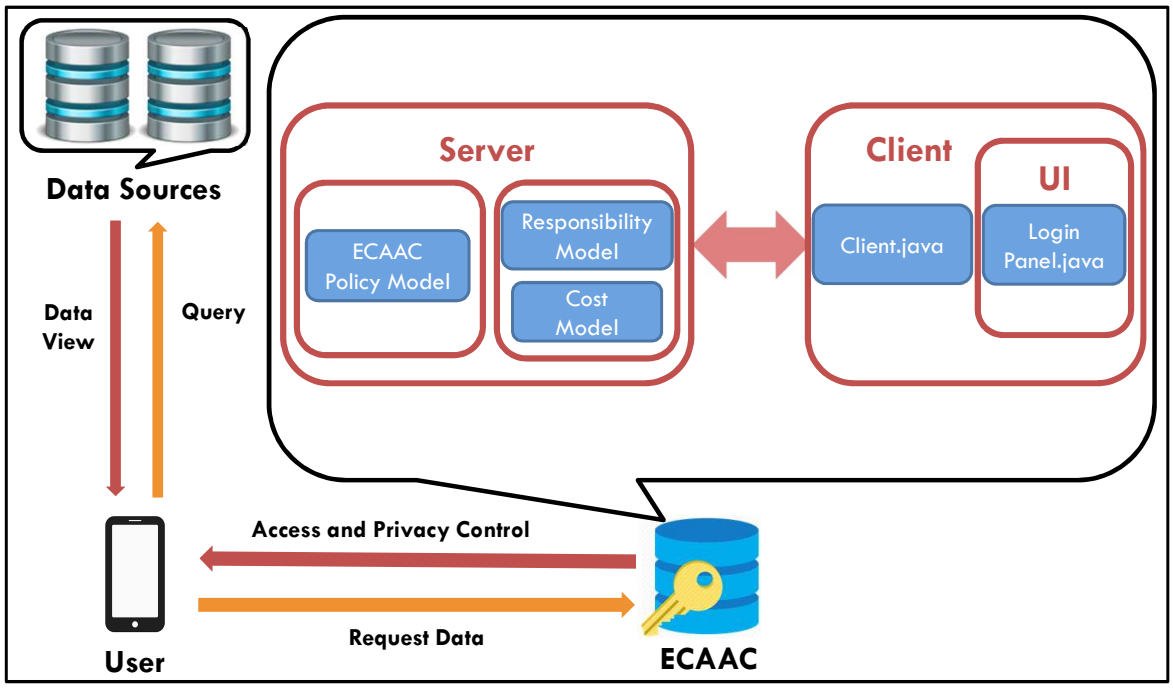

Figure 3. The Development Environment of the ECAAC Prototype

A $User$ can access $Data$, and when a $User$ violates a $Responsibility$, then it is associated with a $Cost$. In the policy ontology, the $Cost$ is calculated based on the relevant $TangibleCost$ and $IntangibleCost$. However, in this paper, we only focus on estimating tangible costs for a data breach. A $TangibleCost$ is calculated according to $InvestCost$, $PolicyCost$, and $NotifyCost$, which we have already discussed in the previous section.

4.2. Prototype and Case Study

In this section, we demonstrate our proposal via a prototype testing, extending our earlier development environment [8]. In addition, we demonstrate the practicality of the proposed ethical access control through revisiting a case study from our earlier research [8].

Figure 3 illustrates the development environment that includes the main components of the prototype. When a data access request comes from the $LoginPanel$ of the $Client$ part through UI, the $Server$ part checks the applicable responsibility against any data breach (i.e., an unauthorized data access) according to the applicable/relevant ECAAC (ethical context-aware access control) policies. In this research, we have used Java language and two Oracle databases to build our prototype application. We limit the users to access data for any relevant purpose from these databases according to their roles and by satisfying the relevant contexts. If any user violates the responsibility (i.e., the violation of any relevant access control policy), that might result in a fine. The fine amount is charged based on the cost that is calculated by the reasoning rule.

Following the policy ontology, let us demonstrate a policy case for a hospital nurse. In this case, *Amanda*, who is a *medical nurse*, has right to access a patient Bob's *daily medical records* - *DMR* - for *daily care* purpose when she is *co-located* with Bob. However, Amanda has no right to access inpatient records (e.g., demographic data of all the hospitalized patients). Figure 2 shows the instances for this example policy case. A relevant reasoning rule-set is specified in SWRL language and the associated cost is estimated accordingly, which is presented as follows.

- $User(u_1) \wedge Purpose(p_1) \wedge canPerform(u_1, p_1) \wedge Responsibility(res_1) \wedge Purpose(p_2) \wedge has(res_1, p_2) \wedge sqwrl{:}notEqual(p_1, p_2) \wedge ifViolate(u_1, res_1) \wedge Cost(c) \wedge TangibleCost(tc) \wedge hasCost(c, tc) \wedge associatedWith(res_1, tc) \Rightarrow charges(u_1,tc)$

Based on our proposed ECAAC prototype, we have tested the hospital nurse case. In this case study, Amanda, who is a medical nurse, has no right to access inpatient records. She is only permitted to access the daily medical records of a patient for a daily care

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 August 2019 doi:10.20944/preprints201908.0239.v1

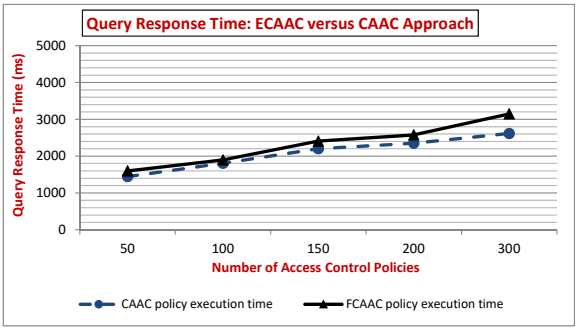

Figure 4. Query Response Time w.r.t. Number of Policies

purpose, when they both are co-located in the general ward of the hospital. The total cost, that can be charged for an unauthorized data access, i.e., a data breach, can be calculated as,

- *X$ for 'InvestCost' + Y$ for 'PolicyCost' + Z$ for 'NotifyCost'.*

This is because Amanda is a nurse, it is not her responsibility to deal with and she has no right to access inpatient medical records, such as demographic data, date of admission, date of referral, reason for referral, other illness data, and so on.

4.3. Evaluation of the ECAAC Framework

In this section, we demonstrate an empirical study on the performance of the proposed ethical access control framework compared to our earlier framework [8]. In addition, we analyse the possible benefits of our proposed responsibility attribution.

We measure the query response time with respect to providing data access permissions to users. Due to the space limitation, we have demonstrated a single set of experimental results with our ECAAC prototype framework. The test focuses on measuring data query response in the light of increasing the number of access control policies. Like the relevant experimental setup from our earlier work on CAAC framework [8], in our current setup, we first specify 50 ECAAC policies and later vary them from 50 to 200 for an increment of 50. In Figure 4, we can see that the query response time is varied from 1.5 sec (for 50 ECAAC policies) to 3.1 sec (for 300 ECAAC policies). However, in our earlier setup, the response time is measured as 2.6 sec with respect to 300 CAAC policies, whereas in our current setup, it measures as 3.1 sec. This is due to the increasing size of the ontology knowledge-base, while increasing the number of access control policies along with an extra reasoning task for assessing risk/responsibility in our ECAAC framework.

Our view of responsibility, in the context of controlling data along with protecting them against data breaches, is a novel research area. According to our proposed responsibility attribution model, the following general access and privacy control features can be measured:

- what are the data that should be controlled and can be accessed at different granularity levels by authorized parties, and
- who are the unauthorized parties that should not have right to access data.

Our ethical access control model also can be used:

- to identify the relevant responsibilities of involved parties, and
- to know about who are responsible for what actions and accountable for the consequences of certain actions, i.e., data breach countermeasures.

Using this model, the social media platforms can:

- facilitate their users by controlling data access, and
- publish to a specific set of authorized parties.

This research appears to be leading to responsibility and accountability attributions, and is capable of estimating countermeasures against data breaches. Overall, the main benefits of our proposed responsibility model is to protect and control data from unauthorized accesses/entities and provide the users with the full control of their own data, along with a proper countermeasure while a data breach, like an unauthorized data access, is detected.

5. Conclusion and Future Research

While researchers have explored various security and privacy control solutions in the past [38], it seems that there is not yet any adequate literature suitably equipped to define responsibility and accountability that are associated with data breaches. This paper provides a retrospective view on the research of responsibility attribution for controlling and protecting relevant data with two major characteristics, *both the roles and their associated actions are measured as responsibility* and *the consequences of certain actions are measured as accountability*. We have investigated different parties involved in data access scenarios and their

responsibilities accordingly. We have discussed and reviewed the data privacy issues and their implications for publishing and accessing data, as well as existing access and privacy control models and frameworks in relation to tackle these issues.

In this paper, we have introduced an ethical access control approach in realizing our formal notion of responsibility and the associated cost model. We have only considered the estimation of the associated tangible costs against a data breach. We specify a theoretical means of developing ethical access control policies and the associated responsibility and cost models. One of the main contributions of this proposal is its ability to provide countermeasures against data breaches. We have shown the practicality of our proposed approach via a case study and a walkthrough of our theoretical approach including a policy ontology and a prototype testing. In addition, we have demonstrated the query response time in the light of increasing the number of access control policies, including an empirical comparative study of the performance of our current ECAAC solution with respect to a more generic CAAC solution. Experimental results demonstrate an acceptable performance of our proposed approach. The performance also can be improved further using more powerful machines.

Overall, the feasibility of our new ethical access control framework is explored and evaluated through detailed comparisons by drawing on experience from our earlier research [12] [8] with existing and emerging CAAC technologies.

Future research work is required to cope with the challenges mentioned in this article, as current access and privacy control solutions are not adequate to improve the access control decisions and restrict users while accessing and protecting data over digital and social platforms. Much research work, especially on data privacy along with corresponding service agreements, is still required so that different parties can have appropriate control of their data on the Internet. Further investigation will be important to automate the proposed responsibility model so that appropriate privacy control policies may be translated according to relevant business models from service providers.

Conceptualising that which constitutes a 'data breach' in terms of political indiscretions remains a challenging task that requires further research beyond the scope of this paper. Future researchers will need to find suitably creative techniques suitable for identifying a baseline critical mass of archetypal, demonstrative and recurring exemplars of data exploitation which might be 'tangible' and/or 'intangible'. All data breaches can be located on one spectrum of abstraction that ranges from the empirically falsifiable to the purely theoretical. The more abstract or 'intangible' the data breach, the more complex the case.

Indeed, defining and managing tangible and intangible cost estimations is a considerable task that will require further research to establish. In particular, intangible costs are particularly important but also especially difficult to identify. Future dedicated investigations will need to involve the development of appropriate methodologies for tracing, mapping, and building taxonomies of flow-on effects/consequences of data breaches (e.g., loss of reputation, increased risk of future attacks due to data breaches and subsequent vulnerabilities, sale of hidden backdoors leftover from the original data breach etc.). Establishing a consolidated threat matrix from the victims of data breach could provide the basis of recovery protocols.

We may need the implementation of suitably dynamic access control policies in order to curb hemorrhaging public trust in news and social media and government institutions, and to improve the quality of decision making capabilities when confronted by the need to integrate users' data from constantly multiplying sources [8]. However, even with highly developed policy ontologies (e.g., maximised concepts coverage) and explicit data models, it is unlikely that humans alone will be capable of managing the access control decisions that must be integrated into next generation ethical access control systems. Such a system also requires ethical checks and balances to ensure that data is treated appropriately and to always make access control decisions in the best interests of the client. A matter of data breaches, once combined with the properties of machine learning, may have major implications as a countermeasure to data and information flows.

References

[1] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach," *The Guardian*, 2018.

[2] J. Isaak and M. J. Hanna, "User data privacy: Facebook, cambridge analytica, and privacy protection," *Computer*, vol. 51, no. 8, pp. 56–59, 2018.

[3] W. H. Inmon, *Building the data warehouse*. John wiley & sons, 2005.

[4] H. Fang, "Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem," in *CYBER*, pp. 820–824, IEEE, 2015.

[5] H. Wang, D. He, and S. Tang, "Identity-based proxy-oriented data uploading and remote data integrity checking in public cloud," *Trans. on Info. Forensics and Sec.*, vol. 11, pp. 1165–1176, 2016.

[6] J. E. Kim, X. Fan, and D. Mosse, "Empowering end users for social internet of things," in *IoTDI*, pp. 71–82, ACM, 2017.

[7] Y. Saleem, N. Crespi, M. H. Rehmani, R. Copeland, D. Hussein, and E. Bertin, "Exploitation of social iot for recommendation services," in *WF-IoT*, pp. 359–364, IEEE, 2016.

[8] A. Kayes, W. Rahayu, T. Dillon, and E. Chang, "Accessing data from multiple sources through context-aware access control," in *TrustCom*, IEEE, 2018.

[9] Q. Zhu and M. M. Skoric, "The role of icts in adaptive and persistent authoritarianism: A study of china at the administrative division level," in *2014 47th Hawaii International Conference on System Sciences*, pp. 2168–2177, IEEE, 2014.

[10] S. Frunza, "Ethical responsibility and social responsibility of organizations involved in the public health system," *Revista de cercetare si interventie sociala*, vol. 32, p. 155, 2011.

[11] R. Layton and P. A. Watters, "A methodology for estimating the tangible cost of data breaches," *Journal of Information Security and Applications*, vol. 19, no. 6, pp. 321–330, 2014.

[12] A. Kayes, W. Rahayu, T. Dillon, E. Chang, and J. Han, "Context-aware access control with imprecise context characterization for cloud-based data resources," *Future Generation Computer Systems*, vol. 93, pp. 237–255, 2019.

[13] Q. Ni, E. Bertino, J. Lobo, C. Brodie, C.-M. Karat, J. Karat, and A. Trombeta, "Privacy-aware role-based access control," *TISSEC*, vol. 13, no. 3, p. 24, 2010.

[14] T. Striphas, "Algorithmic culture," *European Journal of Cultural Studies*, vol. 18, no. 4-5, pp. 395–412, 2015.

[15] A. Kayes, J. Han, W. Rahayu, T. Dillon, M. S. Islam, and A. Colman, "A policy model and framework for context-aware access control to information resources," *The Computer Journal*, vol. 62, no. 5, pp. 670–705, 2018.

[16] A. Kayes, J. Han, and A. Colman, "Ontcaac: an ontology-based approach to context-aware access control for software services," *The Computer Journal*, vol. 58, no. 11, pp. 3000–3034, 2015.

[17] A. Giddens, "The consequences of modernity. 1990," 2007.

[18] A. Giddens and P. W. Sutton, *Essential concepts in sociology*. John Wiley & Sons, 2017.

[19] G. Simmel, "The metropolis and mental life," in *The urban sociology reader*, pp. 37–45, Routledge, 2012.

[20] J. H. Ziegeldorf, M. Henze, R. Hummen, and K. Wehrle, "Comparison-based privacy: nudging privacy in social media," in *Data Privacy Management, and Security Assurance*, pp. 226–234, 2015.

[21] S. Srinivasan, "Cloud computing security," in *Cloud Computing Basics*, pp. 81–100, Springer, 2014.

[22] S. Tomy and E. Pardede, "Controlling privacy disclosure of third party applications in online social networks," *IJWIS*, vol. 12, no. 2, pp. 215–241, 2016.

[23] B. B. Gupta, N. A. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," *Telecommunication Systems*, vol. 67, no. 2, pp. 247–267, 2018.

[24] M. Ahmed and A. T. Litchfield, "Taxonomy for identification of security issues in cloud computing environments," *Journal of Computer Information Systems*, vol. 58, no. 1, pp. 79–88, 2018.

[25] P. A. Watters, M. F. Watters, and J. Ziegler, "Maximising eyeballs but facilitating cybercrime? ethical challenges for online advertising in new zealand," in *2015 48th Hawaii International Conference on System Sciences*, pp. 1742–1749, IEEE, 2015.

[26] R. Sen and S. Borle, "Estimating the contextual risk of data breach: An empirical approach," *Journal of Management Information Systems*, vol. 32, no. 2, pp. 314–341, 2015.

[27] S. Romanosky, D. Hoffman, and A. Acquisti, "Empirical analysis of data breach litigation," *Journal of Empirical Legal Studies*, vol. 11, no. 1, pp. 74–104, 2014.

[28] I. Pavlović, T. Kern, and D. Miklavčič, "Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study," *Contemporary clinical trials*, vol. 30, no. 4, pp. 300–316, 2009.

[29] M. Jackson, "Responsibility versus accountability in the friedrich-finer debate," *Journal of Management History*, vol. 15, no. 1, pp. 66–77, 2009.

[30] D. Gibson, *Managing risk in information systems*. Jones & Bartlett Publishers, 2014.

[31] J. B. Bernabe, J. L. H. Ramos, and A. F. S. Gomez, "Taciot: multidimensional trust-aware access control system for the internet of things," *Soft Computing*, vol. 20, no. 5, pp. 1763–1779, 2016.

[32] A. Kayes, W. Rahayu, T. Dillon, S. Mahbub, E. Pardede, and E. Chang, "Dynamic transitions of states for context-sensitive access control decision," in *International Conference on Web Information Systems Engineering*, pp. 127–142, Springer, 2018.

[33] A. Kayes, J. Han, and A. Colman, "An ontological framework for situation-aware access control of software services," *Information Systems*, vol. 53, pp. 253–277, 2015.

[34] A. Kayes, J. Han, and A. Colman, "Po-saac: A purpose-oriented situation-aware access control framework for software services," in *International Conference on Advanced Information Systems Engineering*, pp. 58–74, Springer, 2014.

[35] C. Wilka, "The effects of clapper v. amnesty international usa: an improper tightening of the requirement for article iii standing in medical data breach litigation," *Creighton L. Rev.*, vol. 49, p. 467, 2015.

[36] A. Johnston *et al.*, "Prevention is better than the cure: Getting privacy compliance right is essential practice management," *Australian Journal of General Practice*, vol. 48, no. 1/2, p. 17, 2019.

[37] D. Miller and H. A. Horst, "The digital and the human: A prospectus for digital anthropology," *Digital anthropology*, pp. 3–35, 2012.

[38] A. Senarath and N. A. Arachchilage, "Why developers cannot embed privacy into software systems?: An empirical investigation," in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, pp. 211–216, ACM, 2018.