Measuring Privacy

Tracy Ann Kosa[*]
UOIT, Oshawa, Canada
TracyAnn.Kosa@uoit.ca

Khalil El-Khatib
UOIT, Oshawa, Canada
Khalil.El-Khatib@uoit.ca

Stephen Marsh
Communications Research Centre Canada
Ottawa
Canada
steve.marsh@crc.gc.ca

Abstract

There is no unified theory of privacy. Law, political science, economics, sociology and psychology have thoroughly explored the concepts of privacy, while computer science has attempted to apply these concepts with varying degrees of success. The study of privacy is often lost in a debate over values, whether privacy itself is a good thing or a bad thing, and how / when it may be reasonably invaded. This paper ignores that debate, reasoning that privacy is legislated so the values issue is no longer relevant, and proposes a theoretical mechanism for measuring privacy using trust as a model based on the need (briefly examined in Section 3) that knowledge about an individuals state of privacy is necessary. Presenting 3 different sets of factors (human, computer and data) derived from multiple disciplines, this work identifies the list of considerations from which a state of privacy may be derived in any given situation; physical or virtual world. This work proposes an original model of the states of privacy based on the identifiability of an individual. Representation is a finite state machine, while the same list of factors can be used to calculate transitions in the machine.

Keywords: Privacy, Trust, Finite State Machine

1 Introduction

The invention of the sphygmomanometer - the blood pressure meter - offered medical professionals the opportunity to identify diseases and treatments that had a correlation effect on systolic and diastolic blood pressure. A standardized tool for measurement resulted in advancement in research on the treatment of heart disease that saved lives. Measurement tools like these allow people, processes and technology to advance and share research using a common language and functionality. Conversely, privacy research has historically focused on dissecting the meaning of the word in different legal and cultural environments, beginning in the West with Warren and Brandeis's claim of privacy as the right to be alone. There are two scholars that set out the foundational theories that inform modern research approaches to privacy. The first, and arguably most popular, is Alan Westin's (1967) that describes privacy as the ability of the individual to determine when, how and what information is communicated about themselves. The second by Altman (1975) as *the selective control of access to the self*. While Westin's definition focuses on information that is personal, Altman's contextualizes privacy in social interaction. This relatively inclusive theory of privacy emphasis process over classification, while Westin's necessitates defining state and functions over the disclosure process [7]. The significant difference in approach indicates that agreement on the theory of privacy is a long way off.

Journal of Internet Services and Information Security (JISIS), volume: 1, number: 4, pp. 60-73

[*] Corresponding Author: Faculty of Science, University of Ontario Institute of Technology, 2000 Simcoe Street, Oshawa, Ontario, L1H7K4, Canada. This work was partially supported by NSERC.

The literature review in Section 2 demonstrates that research in privacy continues to be fragmented within the social sciences and humanities, which has led to complexities in the applied sciences, in the virtual world in particular. A brief discussion of the why knowledge about a data subject's preferred state of privacy is critical, particularly in today's increasingly networked environment, in presented in Section 3.

Section 4 of this paper builds on [5]. which demonstrated that the formalization of trust can be leveraged to formalize privacy, suggesting the future work should expand on the states of privacy and how privacy decision-making can inform the applicability of an individual's threshold for privacy. It postulates a measurement for privacy that uses a finite state machine to identify an individual person's state of privacy in a computational environment (off or online).

Section 5 proposes that there are a set criteria of factors that can be used to identify privacy states: one set derived from sociological and psychological research on privacy, another derived from the information management services provided by computer systems that exist in a networked environment that further impact privacy (with or without the individual's knowledge and consent), and a third that represents the type of data [6]

In Section 6, transitions between states using factors is presented. Section 7 proposes a framework for testing the model, and areas for future work are identified in Section 8.

This work has one goal: to provide a mechanism for measuring privacy. In the event this is not possible, it may still be beneficial in its contribution towards the creation of a unified theory of privacy. It also provides a defined list of functions for computers in the management of information flows, and integrates the research of independent disciplines to demonstrate how they work together in the study of privacy.

2 Literature Review

Privacy research is multi-disciplinary. Like trust, the concept of privacy originates in the field of psychology and social psychology. Hirshleifer writes about privacy as an evolutionary trait, were sociality is based on the principles of sharing, private rights and dominance. Privacy evolved in the social structure known as territoriality, where people had personal space. The privacy ethic, as Hirshleifer calls it, is predicated on the 'reluctance to intrude and willingness to retreat on the part of potential challengers.'

That duality may be a selfish claim, but it is a participatory ethic that not only requires the individual to resist the temptation to violate another's privacy, but also participate in the enforcement of privacy where privacy is not being threatened [3]

This duality is expressed in earlier related literature. For example, Sears talks about privacy as a conflict of values where the right to privacy must be balanced against the right of society to know about whatever subject; including man himself, and how he functions as an individual and in a group. This demand for knowledge typically plays out in scientific research, where the problem becomes that anyone can say what research will and will not benefit society (and therefore justify an invasion of privacy) [12].

Sociological research focuses on the individual (versus group) side of privacy, and how it can be a tool for healthy social interaction. Schwartz identifies privacy as a critical ingredient for group preservation and a stable social system. Privacy is used to support the maintenance of status divisions and protect personal relationships by allowing the expression of non-public postures. Individuals have use privacy to maintain information control; which prevents the ego from identifying too closely with public roles by maintaining a secret self. That said, there are benefits to giving up privacy, Schwartz notes, including mutual gratification and mutual revelation; if I tell you a secret, you are more likely to reveal one yourself. Deception plays an important role in privacy, as individuals must be aware of the masquerades they and others put on in public. It becomes important to maintain a sense of self by resisting the pull into a wider social unit. To exercise privacy appropriately, Schwartz notes three consistent tensions: sincerity and guile, self release and self containment, and the impulse to embrace the public and drive to escape the discomfort of group demands [11].

Acquisti and Grossklags (economics) have examined how individual psychology techniques can be applied to privacy decision-making. The authors combine theoretical and empirical approaches to privacy to investigate the drivers and apparent inconsistencies of privacy decision-making behaviour. Challenges in privacy decision making include incomplete information provided to data subjects and psychological factors involved in decision-making processes (bounded rationality and deviations from rationality). The authors constructed a survey based on e-commerce preferences to avoid the self-selection criteria demonstrable in the privacy field, and observed a number of correlations. For example, privacy attitudes are linked with income: the lowest personal income groups tend to be less concerned about privacy than all other income groups. Privacy attitudes were clustered with k-means multivariate techniques, yielding another correlation between the importance of privacy and concern for privacy. In terms of privacy-related behavior, the study results indicated multifaceted behavior: not all individuals protect privacy all the time in the same way or for the same reasons. In their analysis, the authors noted that respondents frequently over estimated or under estimated privacy threats; and failed to realize how powerful the links are between different pieces of personal information. Acquisti and Grossklags concludes that educated privacy decision making requires a enormous amount of information and analytical processing; the capability of human beings to undertake such a process is limited by bounded rationality and subject to a number of established psychological deviations from rationality (as proven in economic and psychological literature) [1].

The disciplines of law and computer science include research on privacy taxonomies and ontologies. Solove [15] reviews the traditional conceptualizations for privacy (limited access to self, secrecy, control over personal information, intimacy etc) and proposes a pragmatic bottom-up approach based on Wittgenstein's notion of family resemblances using Dewey's approach to philosophical study. The new approach does not distinguish between public and private, but considers individual practices in the space of the family, body and home. He concludes by applying this approach to specific cases in American jurisprudence that could be clarified by this method of analysis. To some extent, Solove [16] includes consideration of historical sociological research and human evolutionary practices in his conceptualization. He later sets out to examine the mechanisms by which privacy may be invaded in respect of these practices.

Tang and Meersman [18] set out to apply ontological technology directly to regulated privacy requirements, by linking case law and legislation. In this environment, the proposed ontology would be represented by fact lexons (extracted from case law) and the directive commitments (that tailor fact lexons to ascribing real life application requirements). The authors are some of a very few researchers in ontological field of privacy that propose a development environment: DOGMA (Developing Ontology-Guid ed Mediation for Agents), as it separates concepts and relations from constraint s, derivation rules and procedures. However, no testing has been done to date on the application of such a model.

Hassan and Logrippo [2] present a computational model-based approach to extracting privacy requirements from legislation and incorporating them into a technical framework. The authors note this is merely the first part of a three step process to detect compliance faults in system architecture. Using American financial text from the Sarbanes-Oxley, instruments examples from Canada's financial compliance law, and accountability requirements in Canada's privacy law, the authors illustrate how their method uses a two-stage process to identify high-level requirements and classify the refined requirements once identified. The proposed UML-based governance extraction model would operate as part of an implemented legal compliance framework in a given organization. The article notes that the novelty of the model lies in the classification of legal requirements and the abstraction of the governance model, as well as the potential for translating both to a logic-based language for validation. While this may assist the organization in compliance frameworks, particularly in the administration of privacy programs, it neglects to expressly consider the individual's preferences.

Jacobs and Abowd) [4] propose a new framework for technologists to consider privacy requirements, which makes a unique contribution in computer science by positioning the individual (as opposed to the system) front and centre. The critical points are: consideration of the physical nature of the input stimulus, location origin, sensing location and granularity of information produced. This combination of hardware, software and usage factors are the basis of the proposed framework, which is developed based on the Terrell's legal ethics work, itself rooted in meta ethics (how values are expressed rather than what they are).

3 Knowledge About Privacy

Historically, privacy was a luxury for the upper class. The ability to invade an individual's privacy reflected ones own status in society; for example, a doctor was expected to know every patient's secrets. Obligations were definite, and defined in the context of information collection, use and disclosure. Privacy violations were evaluated on a precise (if not widely consciously exercised scale) that considered the nature of the transgression, the interpersonal bond between two people sharing personal information, individual expectations and the capacity for secrecy [11].

Modern ideas about privacy are generally addressed by legal theorists and philosophers. This work is largely conceptual, and generally confusing. While there is a recognition of the importance of the concept, its relation to freedom and democracy, the increasing number of privacy laws (even within the same jurisdictions) has added to the confusion. Case law depends on the interpretation of privacy as a value; using the term *reasonable expectation* to set guidelines about what an organization should do in managing personal information [15].

As a result, privacy is now about legislation and compliance. Knowledge of a data subject's privacy preferences is necessary for organizations and Government to comply with various laws in providing services. To do so in an virtual world - where there are multiple service delivery partners - makes the management of these preferences even more complex.

Knowledge of privacy preferences are necessary in both social, governmental and commercial transactions. Violating privacy in a social setting has a direct impact on the legitimacy, duty and privileges of the invader. For the invaded, personal relationships are at risk, there is a loss of information control and the - depending on the nature of the violation - the maintenance of the status divisions in society are threatened [11].

When Government is the invader, the risk becomes significantly higher as law enforcement in particular is given the ability to invade privacy directly and indirectly [12]. In other transactions, Government is required to (at minimum) inform and notify the individual of the collection, use and disclosure of their personal information, and be aware of their preferences for secondary use.[1]

In commercial transactions, companies are required (under most Western legislative regimes) to at minimum acknowledge the collection of an individual's personal information; and quite possibly obtain consent as well. The need exists for both individuals and organizations to know what a given persons state of privacy is.

For individuals, the benefit is more educational, and supports decision making. The majority of Western legislative regimes enforce privacy rights by complaint. In order to proceed through the process,

[1] For example, the Personal Information Protection and Electronic Documents Act, Freedom of Information and Protection of Privacy Act, Personal Health Information Protection Act, and the Municipal Freedom of Information and Protection of Privacy Act. Each statue requires consent for secondary use, consistent with the requirement s under the Canadian Standards Association Model Code for the Protection of Privacy, and the Organisation for Economic Co-operation and Development Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data.

individuals should be able to determine when to complain, and on what basis. Identifying their own state of privacy assists in making decisions about moving from state to state, and determining tolerance for privacy risk.

4 The States of Privacy

The authors purpose that an individual will be in one of nine possible states of privacy in any given computational environment. The privacy states represent, among other other things, different level of identifiability.

1. State 1 is *Private*: total privacy, and existence is unknown. It is unknown whether a person is present or not.

2. State 2 is *Unidentified*: only existence is known, for example, a shadowy figure can be seen but no identifying features can be determined.

3. State 3 is *Anonymous*: limited object and information may be known, but there is no link to a specific identity. Consider an anonymous donor to a named organization; there are a number of known data elements (donation amount, time, organization name), but a specific person cannot be derived from this data.

4. State 4 is *Masked*: the object and information is known but linkages to an identity are concealed. The intent of the concealment may be deliberate or accidental.

5. State 5 is *De-identified*: the object or information itself does not directly identify a person, but when linked with other objects or information the person may become known.

6. State 6 is *Pseudonymous*: defined objects and information are revealed but identified or associated by an assumed (incorrect) name. The level of identity assurance is the key distinction between this state and others.

7. State 7 is *Confidential*: defined objects and information revealed to a defined person or organization acting in a certain role in a defined setting.

8. State 8 is *Identified*: the defined objects and information are capable of being distinguished, and named. Characteristics (objects and information) known to almost everybody with few or no control.

9. State 9 is *Public*: no private objects or information, complete openness. This is the least amount of privacy possible for an individual, where everything about them is available and assigned.

These states are represented in figure 1.

5 Common Factors in Privacy

The state of privacy can be derived from a finite set of factors for a given situation.

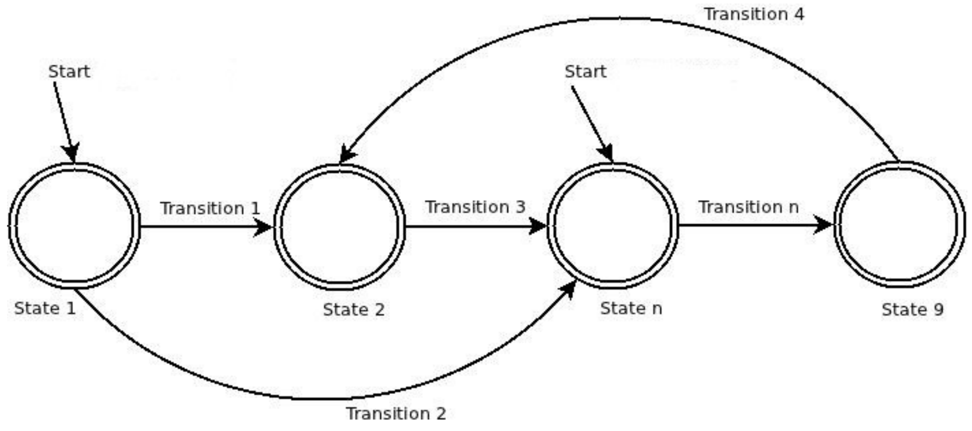


Figure 1: States of Privacy

5.1 Theoretical Calculation

Together, these factors can be added up to identify the state of privacy in a given environment. Where Alice's state of privacy is P, the sum of all factors are represented as human H, services factors are represented as S, data factors are represented as D, Alice's state of privacy for a given environment is: $P = \int(H,D,S)$. Where F represents any one of the set of factors H,D,S then the total set of factors can be represented by $F = \sum(H,D,S)$. The higher the result of the total factors set $\sum F$ the more likely the individual will move to a lower state of privacy; represented as $P_n < P_m$.

5.2 Human

Principles about privacy are not rules; they are observed phenomena from the psychological and sociological literature about privacy. Most are logical enough to be included as rules for computing states of privacy. Human privacy rules are specific to the establishment; they are reflected in the physical structure and properties of society.

There are a number of factors that a data subject will consider when making decisions on voluntary information disclosure (reducing their own state of privacy). In any given situation, one of the key decision criteria is the existence of a common bond; people are more likely to share information with someone they know, and have established a social relationship with. That relationship is based on role specificity: I am a patient, you are my doctor, and may change depending on the social status of the observer to whom information is disclosed [11].

The group environment should be determined by certain structural properties, or expectations. The existence of a ritual changes information disclosure decisions; one is expected to share information at a social gathering. The social condition itself is also a factor, and should have defined properties: am I at a dinner party or a police station? The situation itself should have a defined beginning and end, and the visibility of event also matters: is the room videotaped, and I am specifically being recorded? Considerations about expectations from the public are also a factor, and knowledge of the existence of environmental discretion matters [11].

Individuals will consider the object of the secret, making decisions about themselves and their appearance as they do. They will also consider their own abilities prior to disclosing information. Specifically, the ability to regulate access to the information disclosed, and control over that information once it is disclosed. Control over the audience is also an important factor; can the individual identify all the parties around the space? Is it possible to identify everyone who will have access to the information I disclose?

Do I have any ability to exercise decisions in this situation? Do I have a choice about what information to disclose? [11]

Table 1 below summarizes the human factor list.

Number	Factor	Consideration
H_1	Object	Subject matter?
H_2	Appearance	Of self, others
H_3	Choice	Is choice possible?
H_4	Control: Info	What info is disclosed?
H_5	Control: Audience	Who is present?
H_6	Control: Access	Who may have access?
H_7	Discretion	Is discretion possible?
H_8	Roles Established	Each party has a role?
H_9	Status	Social status of invader?
H_{10}	Common Bonds	Existing relationship?
H_{11}	Social Structure	Existing social relationship?
H_{12}	Social Condition	What kind of situation?
H_{13}	Ritual Type	What are the social rules?
H_{14}	Authority	Is there an authority figure?
H_{15}	Visibility	Defined physical space?
H_{16}	Public Expectation	Absence of expectations?

Table 1: Human Factor Set

Table 1 presents the list of known factors that could be considered by individuals making privacy decisions. Not all of these factors will apply to each decision a data subject makes about their privacy. In some cases, they will overlap or be irrelevant. It is possible, if not likely, that the majority of these choices and decisions are made sub-consciously [11].

5.3 Data Types

The notion of privacy as information protection is well represented in legislation and regulation across the world. Less widely used is the notion of identifiability; that data exists that may or may include traditional identifiers (e.g. name) but may still uniquely identify a data subject. Within the notion of identifiability, some types of data can reveal more identifiable information then others. With few notable exceptions a phone number is less privacy invasive then a unique identifier.[2] Table 2 summarizes the data type factors.

Some of this data is generated by a data subject, other types are created by systems about a data subject, and yet other identifiable information is generated by the technical systems we use in an electronic (off line or online) environment. Consideration of what type of data is moving throughout the system could move the individual from one state of privacy to another, independent of what their own preferences are or what the technical function of the system is.

[2] For a data subject in the witness protection program, a telephone number can be significantly more invasive then a unique identifier associated with a new identity. As narrow and specific as this example is, any privacy measurement tool has to be flexible enough to consider each possible combination.

Number	Data Type	Examples
D_1	Biographic	name, age, birth date, race, ethnicity, religion, martial status
D_2	Demographic	mailing address, telephone number, location history
D_3	Health	diagnosis, care plans, genetic information, blood type, test results
D_4	Education	schools, places of work, employment history, education history
D_5	Financial	credit card numbers, account information, credit history
D_6	Criminal	victim, witness or perpetrator of crime
D_7	Identifiers	driver's license, health care numbers, medical record numbers
D_8	Behavioural	preferences and choices, current physical location

Table 2: Data Factor Set

5.4 Services

Computers are generally accepted to be effective tools for information management; used to organize, retrieve, acquire and maintain information. As technology evolves it becomes cheaper and more convenient to store information for longer periods of time. Increasingly, machines can read information without human intervention. Yet for all the new technologies that continue to evolve, when it comes to managing information about a data subject, there are a discrete number of functions that computers can provide. Alongside of those discrete functions are specific privacy considerations, some of which are more significant than others.

Computers can be networked together for the for the purposes of transmitting information. They serve as storage devices, providing hosting services to hold information. Some computer devices are used strictly to maintain directories to enable access to information, or provide registration services. Other devices are used to build messaging systems, including email, instant messaging and chat functionality. Applications are used to manipulate information, backup tapes are used to retrieve information in the event of system failure. Archives are used to store information for long term use. Finally, portals and web services are used as a mechanism to display information in a networked environment. These 8 functions that computers perform when it comes to managing information have varying impacts on the privacy of individuals, simply by the way the must be built and designed at the system level in order to function. Since the functional list is discrete, a list of factors can also be derived related to a given computer system that will impact a data subject's state of privacy in the electronic world.

Networking technology involves digital messages that are transmitted by a data source across the physical, data and presentation layers over wired or wireless connections. These messages include not only the content created by the individual sending a specific message, but also information about the devices used (e.g. what kind of operating system, or type of phone), and other machine generated data (logs) that contain information required to move the information along the network. From a privacy perspective, networks provide the opportunity to conduct real time tracking surveillance of individuals. The risk of disclosure is high because the function of the service is to transmit (disclose) but also because identifiability is high.

Information kept in databases, directories and even the cloud environment are managed by hosting services, either self hosted or Internet hosting. Internet hosting includes dedicated hosting services, or managed hosting services. It may also include virtual private servers, or co-location facilities. These types of computers hold every kind of information a data subject or organization can create, use and maintain, in addition to machine generated data required by the given hosting environment to keep it functional and reliable. With the delivery of these kinds of services, privacy issues tend to be related to secondary use of data (for a purpose other than originally intended) , records management (data is kept, for example, when it should have been destroyed), and security (data is not secured appropriately). The relative risk of disclosure is limited, since the primary function relates to storage and use, although identifiability of hosted data can vary considerably depending on what the users are putting in the environment, and what information is captured about them as they use the service.

Registration processes are used to enable individuals to become users of the system; they provide credentials (such as a user name and password), where non-registered users are referred to as guest. Individuals who are required to provide proper credentials for a website sign in to a service or system to see customized content. In such systems, the key privacy issue relates to secure and appropriate access controls surrounding the use of unique user names and password. However, there is also an issue of behavioural tracking and surveillance of the registered user once they have signed in to a specific service as specialized actions related to identity are enabled. The function of a registration system is to enable access, which necessitates a certain amount of information disclosure. Depending on the system, however, the identifiability of the individual may vary.

Messaging services include all human-readable messages in formatted or unformatted text, sent and received directly from one data subject to another. Information is created, cached and stored to support message creation, delivery and integrity of services. Typical content includes the actual text of the message the individual created, identity related data about the individuals who are sending and receiving the data , and machine generated data (for example, message headers). Along with this type of service and content comes privacy issues like non-repudiation, surveillance and access controls. The purpose of messaging services is to transmit (disclose) data, and while identifiability of the sender and receiver may vary, there is additional (usually unknown) identifiable data generated and transmitted by machines.

Software applications are tools operated by the computer directly to perform a task at the direction of the individual. These types of tools can include office software, media players, and database applications. Essentially, applications are the tools used to create, manage, store and access content by individuals in any kind of platform. The type of identifiable content is largely user generated, but also includes meta data (user and computer generated) and machine generated information about that user generated data. Privacy considerations tend to relate to collection, use and disclosure of the data used in these systems; e.g. emails sent to the right people with the right amount of data.

Backups refer to making copies of data so it may be used to restore the original after a data loss, and therefore must include at least one copy of all data worth saving, which necessitates particularly large data storage requirements. This function is useful for two purposes: restoring an entire state following a disaster, or restoring a small number of files after they are accidentally deleted or corrupted. By its function, this service includes all data derived from the other services and includes user and system generated data, including content and meta data. The use of this service often changes the physical location of the data, which results in a specific privacy consideration regarding access, accuracy and correction. Identifiability may vary depending on the services and data, but the function of the service is storage.

A file archiver is a computer program that combines a number of files into one archive file, or a series of archive files. The archive must include some information about names and lengths of originals files for reconstruction. File meta data is also stored, typically including operating system information , timestamps, ownership and access controls. Like backup services, by its function, this service includes all data and changes the format and structure of the data. As well, there are a number of specific privacy considerations regarding access, accuracy and correction; and the function of the service is storage, while identifiability varies depending on the data backed up although it will most definitely include additional machine generated data. The key difference between archiving and backup is the medium by which the data is moved: in most cases, archiving will change the electronic (and physical) location of the data.

A website displays information virtual, whereas a portal can also provide the functions and features to authenticate users and provide them with access to information and services that are of relevance and interests. On the back end, it pulls together information from multiple sources. Website and portals both capture information about the user, hidden and known, and content may include preferences set by users, system tracking of user actions and preferences. As a result, the key privacy considerations are behavioural tracking, data mining, and accountability. While the function is to display information, identifiable data is often tracked about the user through known and unknown mechanisms.

These use of one or more of eight system functions in combination or alone can move a data subject from one state of privacy to another, independent of the human factors. Each function has a relative weight associated with it, assigned based on how much information can potentially be disclosed through the use of the service about an identifiable individual.

Table 3 summarizes the systems factor set.

Number	Factor	Sources of Identifiable Information
S_1	Network	User, Machine
S_2	Hosting	User, Machine
S_3	Registration	User, Machine
S_4	Messaging	User, Machine
S_5	Backup	User, Machine, Metadata (both)
S_6	Software	User, Machine, Metadata (both)
S_7	Archiving	User, Machine, Metadata (both)
S_8	Websites / Portals	User, Machine, Metadata (both)

Table 3: Services Factor Set

6 Transitions

An individual may move in different ways in along the finite state machine representing privacy; either forward or backward, and jumping ahead by multiple states depending on the weight of the factors.

6.1 Moving Forward

Forward transitions are represented by R, where an individual advances along the state machine (losing privacy) as the sum of the factors increases. Where P_n is the starting privacy state $P_n + R = P_m, P_m > P_n$.

The forward transition is triggered by a information disclosure. There are four different types of disclosures that can occur [11].

1. I disclose information about myself.
2. A third party discloses information about me.
3. I disclose information about my property or objects.
4. A third party discloses information about my property or my objects.

The intention behind this disclosure is not of interest for the purposes of this paper; it may be well-intentioned, accidental or malicious. Regardless, the disclosure reveals information about an identifiable individual.

Where ID represents a given information disclosure, V represents information about an individual, I represents information about an individual's property, and x, y are two different people or organizations, the four different disclosures may be represented as:

1. $IDx(x,v)$: I disclose information about myself
2. $IDy(x,v)$: A third party discloses information about me
3. $IDx(x,i)$: I disclose information about my property or my objects
4. $IDy(x,i)$: A third party discloses information about my property or my objects

6.2 Moving Backward

A data subject can move backward along the state machine (gaining privacy) as the sum of the factors decreases, where the transition is represented as a negative value $|-B|$. Where P_n is the starting privacy state, this movement can be represented as $P_n + |-B| = P_m, P_n < P_m$.

The backward transition is the result of implementing a reactive information protection; there must be some action that protects privacy resulting in a mitigation of the sum of the factors. There are two possible actions represented in table 4, where E represents establishing a privacy protection, a backward transition may be represented as $|-B| = \sum E$.[3]

Number	Protection	Data Type	Options
E_1	Redact Data	Binary	Yes / No
E_2	Establish Agreement	Binary	Yes / No

Table 4: Information Protection Mechanisms

A detailed backward transition may be represented as $|-B| = E_1 + E_2$.[4]

7 Testing

Privacy is currently evaluated through the use of a privacy impact assessment (PIA), a qualitative methodology that is not standardized. Results of these assessments are not effective for comparison purposes.[5]

Organizational privacy enforcement is the domain of jurisdictional regulators and the courts. This mechanistic tool is intended for individual use based on sociological research adapted for test purposes to the online environment for the purposes of extending the study of computer science. This presents an issue of what discipline to test. Sociological testing generally occurs through focus groups, which are used to narrow down and define the hypothesis more accurately. Statistics are used to follow the groups, record predefined incidents and events using statistical analysis software. Once a set of observable data is obtained, the a specific individual and / or group can be identified for a given research module. Person trials typically last anywhere from 3 to 5 years in order to determine if the hypothesis is valid. Upon validation, a larger study may commence. On the other hand, mechanistic testing is based on trials, usually blind or double blind (in which neither the scientists nor the participants know which group is given the 'real' tool). Results are analyzed first by eliminating for other errors, then a tool is given a rating based on the established success criteria. Neither is appropriate for testing privacy states.

To start, computational testing methods will be used to obtain an evaluation of whether the model is viable, and answer the question of whether privacy is measurable. The general test outline is:

1. The model will run in the background of a defined online action, e.g. sending an email.

[3] Weightings will be established from regulatory orders and industry practices.

[4] Another factor that may be considered for inclusion is time. In most legislation, personal information is reclassified after a certain period of time and is no longer subject to any legal privacy protection mechanism.

[5] The use of the PIA as a privacy metric was evaluated in [5].

2. Based on the define factor set, the model will calculate and display a specific privacy state for the online action.
3. A query will run asking the user if the state seems reasonable.
4. A second query will run asking the user to respond to a short series of questions intended to assess decision-making criteria, based on H factors.
5. Human decision making factor weights represented by $H_1, H_2, H_3 ... H_n$ will be adjusted (if necessary) based on the user feedback.

7.1 Anticipated Results

The output of the numerical privacy state achieves two objectives:

1. Informing the individual about their own state of privacy in a given online environment, and,
2. Standardizing the metric for evaluating privacy.

The resulting state can be integrated as part a software interface can be used to determine the value on a measurable, repeatable scale. After the state is displayed, the individual can use the information to decide whether to make a choice to move from one state to another. The completed model will use agents to represent the individual's desired privacy state, alerting when some threshold established by the individual prior to running the original query is reached.

8 Future Work

The delineation of states is closely related to the risk of identifiability in a given environment. Research on privacy and calculating identifiability (as well as associated risks in de-identified data) will be included to advance the empirical work to support the model, in particular [17, 10]. The theoretical work to support the model will be expanded based on [13].

8.1 Is Measuring Privacy a Good Idea?

There are scholars who do not perceive privacy as a positive value; Altman (social psychologist) and Westin (law) study privacy as a neutral value, while Etzioni (sociology) sees it as a mitigated good [7].

This debate of the value of privacy in evident in applied sciences, where privacy is seen as an impediment to research, particularly in the medical field [9].

It is also apparent in industry, where a number of high profile industry leaders of search engine and social networking products have made public statements about the death of privacy (Mark Zuckerberg 2010, Eric Schmidt 2009, Scott McNealy 2000). Building a mechanism for measuring privacy is a value neutral action, not intended to prove or disprove claims about the death of privacy but to act as a mechanism to measure an individual's privacy.

In doing so, it is necessary to acknowledge that the act of measurement may itself provide some harm. Consider the use of Robert D. Hare's Psychopath Checklist-Revised (PCL-R). The checklist, originally used in the lab to identify recidivism rates, is now used across the United States justice system as a key factor in making decisions about granting parole to offenders [14].

This type of mental health research is often a threat itself; with the spoken (or unspoken) goal of identifying what may be considered deviant behavior in any given political environment [12]

Tools for measurement, once developed and published, lose the precise control available in a test environment. For further consideration, in an virtual world the data subject (whether educated about their current state of privacy or not) may in fact have little control or choice about disclosure. McDonald reviews decisions made about the threats from website data collection; privacy policies are the mechanism by which users are expected to educate themselves about virtual data practices. In summarizing the options for policy development in privacy, McDonald notes many inconsistencies with the current legislative framework. For example, notice and choice about information disclosures only work if users have informed consent, which studies have shown is not reflective of the current reality [8].

It is possible that a measurement tool may pose more a risk of more harm than good; further consideration of which should be elicited during the test phase.

References

[1] A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1):26–33, January 2005.

[2] W. Hassan and L. Logrippo. Governance requirements extraction model for legal compliance validation. In *Proc. of the 2nd International Workshop on Requirements Engineering and Law (RELAW'09), Atlanta, Georgia, USA*, pages 7–12. IEEE, September 2009.

[3] J. Hirshleifer. Privacy: Its origin, function, and future. *The Journal of Legal Studies*, 9(4):649–664, December 1980.

[4] A. R. Jacobs and G. D. Abowd. A framework for comparing perspectives on privacy and pervasive technologies. *IEEE Pervasive Computing*, 2:78–84, October 2003.

[5] T. A. Kosa. Vampire bats: Trust in privacy. In *Proc. of the 8th Annual International Conference on Privacy Security and Trust (PST'10), Ottawa, Ontario, Canada*, pages 96–102. IEEE, August 2010.

[6] T. A. Kosa. Privacy and gsm. In *Prof. of the 2011 IASTED International Conference on Wireless Communications (WC), Vancouver, British Columbia, Canada*, June 2011.

[7] S. T. Margulis. Privacy as a social issue and behavioral concept. *Journal of Social Issues*, 59(2):243–261, July 2003.

[8] A. M. McDonald. *Footprints Near the Surf: Individual Privacy Decisions in Online Contexts*. PhD thesis, Carnegie Mellon University, 2010.

[9] H. B. Newcombe. Cohorts and privacy. *Cancer Causes & Control*, 5(3):287–291, 1994.

[10] A. Pfitzmann and M. Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf, August 2010. v0.34.

[11] B. Schwartz. The social psychology of privacy. *The American Journal of Sociology*, 73(6):741–752, May 1968.

[12] R. R. Sears. In defense of privacy. *The School Review*, 76(1):pp. 23–33, March 1968.

[13] C. E. Shannon. A mathematical theory of communications. *Bell System Technical Journal*, 27:379–423, 623–656, July-October 1948.

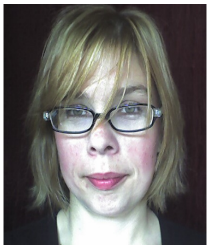
[14] J. Skeem and D. Cooke. Is criminal behaviour a central component of psychopathy? conceptual directions for resolving the debate. *Psychological assessment*, 22(2):433, 2010.

[15] D. J. Solove. Conceptualizing privacy. *California Law Review*, 90(4):1087–1155, 2002.

[16] D. J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–560, January 2006.

[17] L. Sweeney. Simple demographics often identify people uniquely. Carnegie Mellon University, Data Privacy Working Paper 3, 2000.

[18] Y. Tang and R. Meersman. Judicial support systems: Ideas for a privacy ontology-based case analyzer. In *Proc. of On the Move to Meaningful Internet Systems 2005: OTM Workshops (OTM'05), Agia Napa, Cyprus, LNCS*, volume 3762, pages 800–807. Springer-Verlag, October-November 2005.



Tracy Ann Kosa is a doctoral candidate in the Faculty of Science Computer Science program at the University of Ontario Institute of Technology. Her university studies began in political science. Ms. Kosa continued her studies with a Masters degree in political leadership and her second graduate degree focussed on health privacy. She is now employed by the provincial Government as Team Lead for the province's privacy Centre of Excellence, which provide advice, assessment and training services to 600 different Government programs. Her most recent paper examines the use of artificial intelligence techniques for predicting privacy outcomes, and her dissertation work proposes a unit of measurement for privacy.

Khalil El-Khatib was an assistant professor at the University of Western Ontario prior to joining the Faculty of Business and Information Technology, University of Ontario Institute of Technology, in July 2006. He received a bachelor degree in computer science from the American University of Beirut (AUB) in 1992, a master degree in computer science from McGill University in 96, and a Ph.D. degree from the University of Ottawa in 2005. Between the years of 1992 and 1994, he worked as a research assistant in the computer science Dept. at AUB. In 1996, he joined the High Capacity Division at Nortel Networks as a software designer. From Feb. 2002, he worked of Research Officer in the Network Computing Group (lately renamed the Information Security Group) at the National Research Council of Canada for two years, and continued to be affiliated with the group for another two years.

Steve Marsh is a Research Scientist in the Network Security Group at in the Communications Research Centre in Ottawa, Ontario, Canada. His PhD (University of Stirling, 1994) was a seminal work that introduced the first formalisation of the phenomenon of trust (the concept of 'Computational Trust'), and applied it to Multi Agent Systems. As a milestone in trust research, it brought together disparate disciplines and attempted to make sense of a vital phenomenon in human and artificial societies, and is still widely referenced today. Steve's current work builds extensively on this model, applying it to network security, MANETs, and mobile device security. His research interests include computational trust, device comfort, critical infrastructure interdependencies, regret management, and socially adept technologies. He is the Canadian delegate to IFIP Technical Committee 11: Security and Privacy Protection in Information Processing Systems. He is an adjunct professor at UNB (Computer Science), UOIT (Business and IT) and Carleton University (Systems and Computer Engineering and Cognitive Science).