Identity And Privacy Services

Harry Katzan, Jr., Savannah State University, USA

ABSTRACT

Personal identity and privacy are important topics in information systems in general and data analytics in particular. Normally associated with digital security, the scope of identity and privacy is much greater and affects most aspects of everyday life. Related subjects are behavioral tracking, personal-identifiable information (PII), privacy data relevance, data repurposing, identity theft, and homeland security. ***Identity and Privacy Services*** *is an admixture of the major issues in the area of personal identity and privacy and the security of individual rights in a complex societal environment. This is a general paper on this important subject, intended to give exposure to the constituent topics.*

Keywords: Identity; privacy; security; combination of evidence; abductive inference

INTRODUCTION

Identity is a major issue in the security of modern information systems and the privacy of data stored in those systems. Identity and privacy concerns are commonly associated with behavioral tracking, personal-identifiable information (PII), the relevance of private data, data repurposing, and identity theft. (Windley 2005) We are going to approach the subject from an information systems perspective, recognizing that the inherent problems also apply to societal systems. Information systems are a good conceptual vehicle for the underlying security, identity, and privacy models, because data is typically stored off-premises and is under the control of a third-party service provider. When a third party gets your data, who knows what is going to happen to it? Management of information has historically been with the organization that creates or maintains it. From a personal perspective, on the other hand, persons should have the wherewithal to control their identity and the release of information about themselves, and in the latter case, a precise determination of to whom it is released and for what reason. Privacy issues are not fundamentally caused by technology, but they are exacerbated by employing the technology for economic benefit. After a brief review of identity and privacy to set the stage, we are going to cover identity theory, privacy theory, and identity requirements.

Identity

Identity is a means of denoting an entity in a particular namespace and is the basis of security and privacy – regardless if the context is digital identification or non-digital identification. We are going to refer to an identity object as a *subject*. A subject may have several identities and belong to more than one namespace. An identity denotation is based on attributes as suggested by Figure 1.

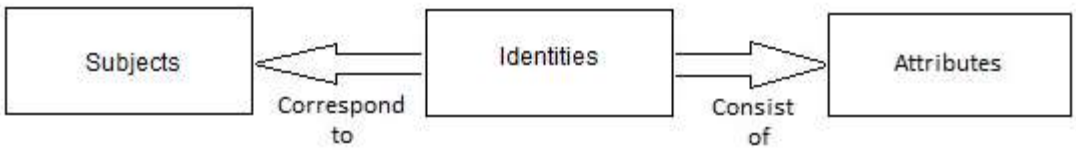

Figure 1: Conceptual relationship between subjects, identities, and attributes

A pure identity denotation is independent of a specific context, and a federated identity reflects a process that is shared between identity management systems. When one identity management system accepts the certification of

another, a phenomenon known as "trust" is established. The execution of trust is often facilitated by a third party that is acknowledged by both parties and serves as the basis of digital identity in information systems.

Access to computing facilities is achieved through a process known as authentication, whereby an entity makes a claim to its identity by presenting an identity symbol for verification and control. Authentication is usually paired with a related specification known as authorization to obtain the right to address a given service.

It is generally felt that a framework for understanding a technology should reflect the underlying concepts required for its development and subsequent acceptance as an operational modality. A technology should enable the delivery of value rather than constrain it, and that is our objective with this paper.

Privacy

Information systems typically process and store information about which privacy is of paramount concern. The main issue is identity, which serves as the basis of privacy or lack of it, and undermines the trust of individuals and organizations in other information-handling entities. The key consideration may turn out to be the integrity that organizations display when handling personal information and how accountable they are about their information practices. From an organizational perspective, control over information should remain with the end user or the data's creator with adequate controls over repurposing. From a personal perspective, the person should have the wherewithal to control his or her identity as well as the release of socially sensitive identity attributes. One of the beneficial aspects of the present concern over information privacy is that it places the person about whom data are recorded in proper perspective. Whereas such a person may be the object in an information system, he or she is regarded as the subject in privacy protection – as mentioned earlier. This usage of the word *subject* is intended to imply that a person should, in fact, have some control over the storage of personal information.

More specifically, the *subject* is the person, natural or legal, about whom data is stored. The *beneficial user* is the organization or individual for whom processing is performed, and the *agency* is the computing system in which the processing is performed and information is stored. In many cases, the beneficial user and the subject are members of the same organization.

The heart of the issue is *privacy protection*, which normally refers to the protection of rights of individuals. While the concept may also apply to groups of individuals, the individual aspect of the issue is that which raises questions of privacy and liberty

Privacy Assessment

The Federal Bureau of Investigation (U.S.A.) lists several criteria for evaluating privacy concerns for individuals and for designing computer applications: (FBI 2004)

- *What information is being collected?*
- *Why is the information being collected?*
- *What is the intended use of the information?*
- *With whom will the information be shared?*
- *What opportunities will individuals have to decline to provide information or to consent to particular uses of the information?*
- *How will the information be secure?*
- *Is this a system of records?*

Since privacy is a fundamental right in the United States, the above considerations obviously resulted from extant concerns by individuals and privacy rights groups. In a 2009 Legislative Primer, the following concerns are expressed by the Center for Digital Democracy: (CDD 2009, p. 2)

- Tracking people's every move online is an invasion of privacy.
- Online behavioral tracking and targeting can be used to take advantage of vulnerable consumers.
- Online behavioral tracking and targeting can be used to unfairly discriminate against consumers.
- Online behavioral profiles may be used for purposes beyond commercial purposes.

We are going to add to the list that the very fact that personal data is stored online is a matter of concern and should be given serious attention. Based on these issues, this paper is going to take a comprehensive look at the subject of identity in computer and human systems.

IDENTITY THEORY

The notion of identity is an important subject in philosophy, mathematics, and computer information systems. In its most general sense, identity refers to the set of characteristics that makes a subject definable. Each characteristic can be viewed as a single point in a three-dimensional Cartesian coordinate system where the axis are *subject*, *attribute*, and *value*. (Katzan 1975) Thus, the fact that George is twenty-five years old could be denoted by the triple <George, age, 25>. A set of characteristics over a given domain can uniquely identify a subject. This simple concept is the basis of privacy and identity in information systems and everyday life. The notion of identity applies to organizational subjects as well as to person subjects.

Knowledge, Attributes, and Identity

Identity is primarily used to establish a relationship between an attribute or set of attributes and a person, object, event, concept, or theory. The relationship can be direct, based on physical evidence, and in other cases, the relationship is indirect and based on a reference to other entities. In a similar vein, the relationship can be certain or uncertain, and in the latter case, based in deduction or inference. The relationship determines an element of knowledge. For example, the knowledge element "you are in your car" is a statement in which "you" and "your car" are things that exist and the "in" is a relationship. Direct knowledge is known by *acquaintance* and is evidenced by a physical connection. Indirect knowledge is determined through a reference to a particular with which the analyst is acquainted. The form is known as knowledge by *description*. (Russell 1912) *Direct knowledge* is determined through sense data, memory, or introspection. *Indirect knowledge* is determined through a reference to another particular, as in "the person who ran for Congress in 2004" or through a form of self-awareness where what goes on in subject's mind, for example, is estimated by an analyst's interpretation based on experience or self-evaluation.

Synthetic knowledge reflects certainty based on evidence inherent in the attribute values at hand. *Analytic knowledge* reflects a degree of uncertainty and is determined by deduction, as in "he is the only person with that 'attribute value'," or by inference based on known particulars, such as "all terrorists have beards." Inference, in this case, could be regarded as a form of derivative knowledge. The value of analytic knowledge is that it enables the analyst to exceed his or her limit of private experience.

Numerical and Qualitative Identity

Identity refers to the characteristics that make a subject the same or different. We are going to establish two forms of identity: numerical and qualitative. Two subjects are *numerically identical* if they are the same entity, such that there is only one instance. Two subject (or objects in this case) are *qualitatively identical* if they are copies or duplicates. In the popular movie *The Bourne Identity*, for example, the characters *Jason Bourne* and *David Web* are numerically identical, and the number of subjects is one. So it is with *Superman* and *Clark Kent* in another domain. On the other hand, a set of animals with the same biological characteristics – e.g., a species – are regarded as being qualitatively identical. The notion of qualitative identity is remarkably similar to the modern definition of a *category* informally defined as a collection of entities with the same characteristics, having the same values for the same attributes.

Theory of the Indiscernibles

An important aspect of identity theory is that subjects exhibit features of permanence and change, analogous to sameness and difference mentioned previously. We are going to discuss the concept of temporal identity in the next section. The notion of change implies that a subject undergoes transformation and also has a property that remains unchanged. Both Locke and Hume have proclaimed that change reflects the idea of unity and not of identity. Leibnitz proposed the *Theory of Indiscernibles* suggesting that subjects (i.e., objects or entities) that are indiscernible are identical. (Stroll 1967) The subject of indiscernibles has implications for information systems and attribute change. To what extent a change in a characteristic denotes a change in identity is an open item at this time and implies that there is a probabilistic aspect to identity.

Russell approaches the subject of identity from an alternate viewpoint, analogous to definite and indefinite articles. Russell proposes that a description may be of two sorts: definite and indefinite. A definite description is a name, and an indefinite description is a collection of objects x that have the property Ø, such that the proposition Øx is true. (Russell 1919) In the phrase *Dan Brown is a famous author*, for example, 'Dan Brown" is a name and the indefinite description is obvious, leading to the probabilistic link between a subject and a characteristic.

Temporal Identity

There is a rich quantity of philosophical literature on the change of identity over time. Are you the same person you were yesterday? Are there persistent attributes that allow for positive identity between time periods? As alluded to previously, entities in everyday life exhibit features of permanence and change. In the domain of personal identity, address attribute is a primary candidate for change. For example, John Smith lives at 123 Main Street. He moves out and another John Smith moves in. This is a distinct possibility in a crowded city. In there a concept in identity theory for this phenomena? Should an identity system take this eventuality into consideration?

There is a form of *attribute duality* between a person subject and an object subject. A subject – an object, such as a residence, in this case – is characterized by who lives there. For example, rich people live on Sutton Place in New York. The discussion leads to four related concepts: endurant identity, perdurant identity, endurant attribute, and perdurant attribute. Clearly, the term *endurant* refers to a noun that does not change, where perdurant refers to one that does. Thus, the identity problem is essentially translated to an operant problem of "recognizing identity."

PRIVACY THEORY

It has long been recognized that privacy is a two-edged sword, not only for individuals, but also for groups and organizations. Subjects have First and Fourth Amendment rights designed to protect against unwarranted disclosure of information with unlimited scope to unwanted parties without proper authorization by the subject. However, privacy considerations protect criminals and terrorists, in addition to ordinary citizens, groups, and organizations. Protections and other conventions used to safeguard trade secrets can also be employed to enable non-disclosure of design and manufacturing flaws from consumers and regulatory bodies.

Privacy has been in the news for at least forty years originating with Alan Westin's seminal book on the subject entitled *Privacy and Freedom*, published in 1967. Others have joined the struggle, namely (Westin 1977, Miller 1971, Katzan 1980, and Givens 2009) to reference only a few of many, with apologies to those not mentioned. One of the toughest problems facing the computer industry is data protection, summarized very well in 1971 by Arthur R. Miller: (Miller 1971, p.37)

The new information technologies seem to have given birth to a new social virus – "data-mania." Its symptoms are shortness of breath and heart palpitations when contemplating a new computer application, a feeling of possessiveness about information and a deep resentment toward those who won't yield it, a delusion that all information handlers can walk on water, and a highly advanced case of astigmatism that prevents the affected victim from perceiving anything but the intrinsic value of data. Fortunately, only some members of the information-handling fraternity have been stricken by the disease.

This quote was written over 39 years ago; what would the author think about today's environment?

Privacy and Data Protection

Data protection is given the most attention when the privacy of an individual or an organization is jeopardized. According to Alan F. Westin: (Westin 1967)

Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.

Privacy is related to data protection, because it is an integral part of society and affects the behavior of its citizens. *Privacy is a service that a subject should expect from and be provided by society.* The physical state of being private has four primary attributes: solitude, intimacy, anonymity, and reserve, which supply group separation, group participation, group freedom, and personal protection, respectively. These states collectively provide the confidentiality required to participate in a civilized society. Concerns for privacy should be an integral part of a data protection program.

An organization requires privacy to achieve its basic objective – whether it is business, education, or government. The disclosure of private internal affairs affects "brand equity" and is detrimental to success.

Another consideration is personal surveillance – even though it may be socially or legally accepted. When a subject does not have control over its informational profile, there is no safeguard over its authenticity. Therefore, a double barreled approach, consisting of technology and regulation, is required for operating in a global economy. (Katzan, 1980, p. 44)

Information Control

Because of the widespread application of computer and communications technology, there has been a gradual trend among private institutions and government agencies to ignore the individual's need for privacy. Privacy safeguards are the individual's sole line of defense against the exercise of power through information control. Individuals can lose control of information about themselves in three ways:

1. Information obtained against the subject's wishes.
2. Information obtained from an agency against the wishes of the agency and of the subject.
3. Information willingly disclosed by the beneficial user or agency but against the subject's wishes.

Information obtained against a subject's wishes is an area in which privacy is normally expected. This category includes explicit attempts to obtain information and implicit methods where a subject is forced to disclose personal information. Typical actions are:

1. Searches and seizure
2. Compelled self-disclosure
3. Informers and secret agents
4. Participant monitoring
5. Public observation and recording of information
6. Consent for fear of reprisal
7. Disclosure for privilege

Some benefits are commonly associated with disclosure of private information, so the fine line between willing and unauthorized disclosure is frequently blurred. In the case of *Information obtained from an agency against the wishes of the agency and of the subject*, the conditions of privacy should apply to the agency as they do to the subject and are normally of concern because of computer security deficiencies and unauthorized access. In the case of *Information willingly disclosed by the beneficial user or agency but against the subject's wishes,* as in interagency transfers, accuracy and context are normally of concern. This is the prototypical *repurposing of information* that lies at the heart of most subjects' concerns over the disclosure of personal information.

Recordkeeping

Records typically fall into four classes: administrative, operational, intelligence, and statistical. In theory, *administrative records* are maintained by governmental agencies and give subjects their identity. For individuals, administrative records normally include birth certificates, diplomas, military discharge papers, driver's licenses, and immigration papers. For organizations, administrative records include certificates of incorporation and related documents. *Operational records* reflect tax and other certificates. *Intelligence records* are maintained by government agencies and represent security permissions and legal investigations. *Statistical records* can be obtained through an official questionnaire, as with the census, or from any of the other records that have been "cleansed" so as not to reflect personal information. Privacy safeguards are summarized in a far-reaching report by the Department of Health, Education, and Welfare (HEW 1973, pp. xx-xxi.):

1. There must be no personal data record-keeping systems whose very existence is secret.
2. There must be a way for an individual to find out what information about him is in a record and how it is used.
3. There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.
4. There must be a way for an individual to correct or amend a record of identifiable information about him.
5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

The five principles are regarded as a Code of Fair Information Practice, emphasizing that privacy is a service that should be afforded to all citizens by other citizens, organizations, and the government in a free and open society.

Privacy Issues

The subject of privacy in all of its "multi-faceted dimensions" is of concern to many persons. Some individuals only wake up to the subject when their privacy is invaded and then quickly go back to sleep when the situation subsides, or they get tired of worrying about it. In the present context, Internet computing would seem to constitute a privacy threat to many persons and also organizations, because sensitive information is held by third-party service providers. However, having a third-party service provider is not a necessary condition for privacy invasion. The gang-of-three (government, employers, and education) would appear to be doing a good job with that. What are the specific issues about which we should be concerned? The topic has been addressed by the Privacy Rights Clearinghouse (PRC) in a document entitled "Privacy Today: A Review of Current Issues" developed by its director Dr. Beth Givens. (Givens 2009) The report lists twenty-three issues in privacy rights with a substantial description of each issue. The report highlights and summarizes the key issues and also contains links to special interest groups working on particular topics in that domain. We are going to concentrate on five subjects deemed relevant to the mission of this compendium:

- Biometrics
- Video surveillance and workplace monitoring
- Data profiling
- Behavioral tracking and targeting
- Records on the Internet

A selection from the PRC list is also necessary because every privacy subject has its privacy point and twenty-three primary issues are more than we can usefully cover in this paper. Here is a simple straight-forward case of an individual personal privacy concern. "Joe Smith is a good runner and ran a local marathon in 3 hours and 20 minutes. The marathon organizer lists the name, age, finishing time, finisher's place, and home city and state of all finishers of the race on the Web. Joe has two concerns. He is a bit embarrassed, because a couple of years ago, he ran the same race in less than 3 hours. So, in this case Joe would prefer not to have the results published online for everyone to see – that is, if anyone besides runners would be interested. Joe's friend Al has a different opinion. Al says, 'That is a great time Joe. My father, who is about your age, ran it in 3 hours and 10 minutes.' The second

concern is more serious. Joe is 57 years old and is looking for a good position, since he was recently laid off. He is concerned that a prospective employer can Google him and determine his age from the online list of finishers, since age discrimination is a major concern for many employers in this country." If the race were run in Canada or Europe, on the other hand, the same information would not be available to outside persons, because of privacy laws.

We are going to present a descriptive technique that will apply the five selected dimensions, placing each dimension in a privacy-identity continuum.

Biometrics

The term *biometrics* refers to the use of bodily characteristics for identification, which can be exact or probabilistic. If you have been in the ROTC, the military, law enforcement, possess a government security clearance, or have been born recently, you have an exact biometric identity consisting of your set of fingerprints on file in an official place. A person's DNA and retinal scan are also supposedly exact biometric identifiers. Clearly, an exact biometric marking does in fact identify a particular individual. However, the assignment of a name from an appropriate namespace is quite another thing. If the task is to link an individual with a specific name, then there is some probability involved. The picture on an official passport, driver's license, or government issued identification is also regarded as an exact identifier. But, how exact is exact? As mentioned before, there is some risk in linking name identification between two or more types of identity.

Less exact biometrics, such as facial recognition, has been employed in social situations to identify persons of interest – such as at sporting events. Using facial geometry and other visual clues, facial recognition technology has been very successful in criminal investigation. But, what about the identification and recording of persons in a lawful demonstration, guaranteed as a First Amendment right? Everyone knows there are at least two kind of demonstrators: those persons participating in the physical part of a demonstration because they genuinely believe in the cause, and those persons with nothing else to do on a Saturday afternoon. As Dr. Givens writes, "As a result, innocent people can be wrongly identified as criminal (false-positives), and known criminal and suspected terrorists can fail to be detected altogether (false-negatives).

Video Surveillance and Workplace Monitoring

Low-cost video surveillance systems are prevalent in modern society, and their use ranges from convenience stores to day-care centers. In fact, video surveillance is so pervasive that most people think nothing about being under the eye of the camera. In criminal investigation, video surveillance is a useful identifier, albeit within some probabilistic limits, and also as an investigative tool.

Collectively, video surveillance and workplace monitoring can provide information related to the following phenomena:

- Facial recognition
- Unproductive employment activity
- Improper use of resources
- Violation of conditions of service

Use of an employer's computer or other resources is a good case in point. There are other forms of surveillance, such as Radio Frequency Identification (RFID) chips embedded in employee identification cards that can be used as an employee locator by recording when he or she leaves one room and enters another.

Keystroke monitors are sometimes used to determine ineffective use of equipment. Most employees do not seem to mind employee monitoring when on premises – but what can they do about it? Off premises and off hours surveillance and monitoring are quite another thing and exist as an open issue in privacy.

Data Profiling

Most of us are well represented in a multitude of gang-of-three databases, such as the tax bureau, social security administration, state motor vehicles office, education records, employment files, insurance, and health records. Information of this type can be regarded as the operational part of the fabric of life. We can temper the intrusion but not totally eliminate it, because it is paramount to identity determination and service management. Identities are linked by numbers, such as the social-security number, name, date of birth, telephone number, address and ZIP code, mother's maiden name, and even mother's birthday. It is even possible to find the social security number of an unrelated deceased person on the Web. Immigration records are also easily obtainable. The Privacy Act of 1974 and its amendments generally cover governmental data protection and profiling.

There is another form of data that is involuntarily collected about individuals where there is some choice involved, such as personal expenditures, lifestyle, Internet activity, political activity and donations, and so forth. Supermarkets, bookstores, department stores, health stores, fitness centers, libraries, toll booths, big-ticket retailers, travel agencies, magazine publishers, and airlines – all contain personal data on individuals. An idea of interests, activities, and expenditures are available from credit card purchases, bank records, and operational files of business, governmental, and educational institutions. Thus, it is quite easy for an interested party to create a *data profile* of a person.

Pundits claim that profile data determines who or what we are. However, there is a tendency to interpret data based on the psychological perspective of the profiler. If you subscribe to "guns and ammo," does that indicate that you are a terrorist, member of the local shooting club, an Olympic athlete, worker in a sporting goods store that sells guns, or a medical professional who uses a service to provide magazines for the waiting room.

It has been reported that search providers turn over search queries of individuals to the agencies of the government. (Conti 2009, pp.259-298) This is a modern form of data profiling. A method, termed *chaffing*, is mentioned to widen the search domain and provide some protection. So, if you are going to search for a controversial person, you might also want to search for some non-threatening person to widen the search area.

Behavioral Tracking and Targeting

Behavioral tracking and targeting is an area of privacy concern related to data profiling with emphasis on what a subject does. Here is a typical scenario. A subject rents a car and drives that vehicle out of state or out of the country by accident or by intention. When the car is turned in to the agency, the renter is charged an enormous penalty. The fine print in the contract was not read, because the renter is usually out of his or her element or just in a hurry. How did the agency know of the unfortunate travel? The car rental agency used a global positioning system (GPS) device to track the path of the vehicle. In addition to GPS tracking, license plate tracking, implemented through highway cameras, is also widely used by state and local law enforcement officials. There is always a stated reason why organizations do things, but in the case of privacy, the main problem is the repurposing of collected data. Through data mining technology, computers can identify patterns based on happenstance, rather than purposeful activity. Here is another example: At the time this paragraph was written, the state of Arizona decided to take border control into its own hands. The federal government could do it and can do it, but we live in a large country with enough problems to go around. Getting the right person's or an organization's attention at the right time usually takes some up front planning. Demonstrations ensue for varying reasons, including the possibility that certain outside people want to stir up trouble. Proper officials are looking into persons flying into Arizona with recently booked tickets for travel lasting only a few days and are doing some data mining to identify those persons. Are identified persons demonstration instigators or grandparents attending a graduation ceremony. Regardless, they are prime candidates for behavioral tracking. In an era of supercomputing, piecing together a travel itinerary is not a major task. All that is needed is a subject to track.

The subject of behavioral tracking also includes the practice of collecting and compiling consumers' online activities, interests, preferences, and/or communications over time. (Givens 2009, 18 of 23) This form of behavioral targeting serves as the basis for advertising and other forms of marketing. Web browsing is a primary source of information in this regard.

There is also a growing trend by Internet service providers (ISPs) to use deep packet investigation (DPI) to look at email, Web sites visited, music, video sharing, and downloads by inspecting the data packets that constitute Internet traffic. This form of privacy intrusion is a major challenge to privacy advocates.

Records on the Internet

There is a tendency in society for persons in a political or geographical jurisdiction to be generally the same. This refers to attitudes, culture, psychological properties, and so forth. Between countries, however, there tend to be some differences between the two groups of people. People from Switzerland are different from people from England. The same idea holds true for people from Minnesota and Georgia, for example. We are referring to what is acceptable data, from a cultural perspective.

The disclosure of public records in an open government is not sensitive to cultural differences, since the context for the information in government-managed files does not travel with the information. Citizens in one area may be more or less sensitive to the content of public information than persons from another – especially in a large country. The "one size fits all" mentality of public disclosure is a subject that frustrates privacy advocates.

Nevertheless, divorce records, criminal records, under-age convictions, bankruptcy proceedings, DIU convictions, motor vehicle records, and so forth, are all publically available through mailing list and information brokers. All an identity thief or stalker needs is a Social Security number and $19.95. The motivation for many, if not most, automobile break-ins in modern times is an attempt to obtain personal information in the glove box, even thought the thief may also take a camera from the rear seat.

IDENTITY REQUIREMENTS

It would appear that there are two essential problems in identity theory: protection of identity and recognition of identity. Protection refers to the safeguarding of one's identity from unwanted intrusion into personal affairs, and is reflected in the identity principles that follow. Recognition refers to the use of identity measures to classify certain persons, based on the combination of evidence and abductive inference. This characterization of the identity problem reflects two edges of the same sword.

Identity Principles

It is generally regarded that effective identity governance should be based on a set of principles to guide the professional activities of IT managers, security officers, privacy officers, and risk management. (Salido 2010, OECD 2010) As delineated, the principles would be based on efficacy in governance, risk management, and compliance with the following objectives:

Governance. Assurance that the organization focuses on basic issues and who is responsible for actions and outcomes.

Risk Management. Assurance that procedures are in place for identifying**,** analyzing, evaluating, remedying, and monitoring risk.

Compliance. Assurance that actions are within the scope of social and legal provisions.

In accordance with the stated objectives, we can delineate the eight core principles of effective and efficient identity management. (OECD op cit., p.3)

Principle #1. Collection Limitation Principle – there should be prudent limits on the collection of personal data with the knowledge or consent of the subject.

Principle #2. Data Quality Principle – personal data should be relevant to stated purposes and be accurate, complete, and up-to-date.

Principle #3. Purpose Specification Principle – the purpose of the data collection should be specified beforehand.

Principle #4. Use Limitation Principle – data should be used only for the use specified and not be repurposed.
Principle #5. Security Safeguards Principle – personal data should be safeguarded by reasonable and state-of-the art security facilities.
Principle #6. Openness Principle – the technical infrastructure for protecting personal data should be open as to development, practices, and policies.
Principle #7. Individual Participation Principle – the subject should have the right to definitive information concerning the personal data collected, methods used, and safeguards employed and have the right to challenge the procedures employed.
Principle #8. Accountability Principle – social, business, educational, and governmental data controllers should be required by legal or regularity means to abide by principles 1-8 and be accountable for violations of their provisions.

The eight principles of identity agree in part and parcel to Cavoukian's "7 Laws of Identity, listed as follows: personal control and consent; minimal disclosure for limited use; need to know access; user-directed identity; universal monitoring of the use of identification technology; human understanding and involvement; and consistent access and interface to personal data. (Cavoukian 2010)

Identity Analytics

An important aspect of identity theory concerns whether a certain subject is a member of a group of interest. The basis for this form of identity determination is that identity is a function of the subject's namespace and attributes. A subject belongs to a category if it possesses the attributes that define the category. Another approach is to employ a knowledge source to determine a subject's group membership. This is the method we are going to use in this section. A popular characterization of the problem would be, "Is suspect A a member of group T?" or in short form, "Is A a T?" Clearly, the methods would apply to most diagnostic systems, such as medical diagnosis, auto repair, and the analysis of aircraft failures. We are going to propose two methods of analysis: the combination of evidence (Shafer 1976, Katzan 2006, Katzan 2010c) and abductive inference (Josephson 1996).

With the *combination of evidence*, a certain level of belief is afforded a knowledge source, as in the following scenario:

We are trying to identify subjects that belong to a certain group G. We know about the group G and its attributes. We have a paid knowledge source K_1 that informs us that subject A is a member of G. However, K_1 is not always correct, and we know that. We have used K_1 enough to know that he provides us with information when he needs money. We have an intuitive belief of how often he is correct. Fortunately, we have another source K_2 that can supply similar information. K_2 is not as hungry for money as K_1, and his opinion frequently runs contrary to K_1's. We would like to use analytics to combine the information from K_1 and K_2 so as to obtain a composite picture of the situation.

The relations between the knowledge sources and the subject are represented by the following mappings:

$K_1 \rightarrow$ A
$K_2 \rightarrow$ A

and the characteristics of the relationships are given as:

A = {m, n}
K_1 = {r, u}
K_2 = {c, i}

The question is whether A is a member of G, denoted by m, or not a member of G, denoted by n. As far as K_1 is concerned, he might be telling us what he thinks we want to hear, so his judgment is classed as reliable, denoted by r, or unreliable, denoted by u. K_2 is simply correct or incorrect, denoted by c or i, respectively. Through a method known as belief propagation (Katzan 2010a), the knowledge is transferred from the problem space to the solution space, resulting in the following representation:

Source	*Representation*
K_1	{[(m), p]. [(m, n), 1-p]}
K_2	{[(n), q]. [(m, n), 1-q]}

The results of belief propagation assign the mass (p) of the information received from K_1 to (m) and the remainder of the belief is assigned to (m, n). A similar argument applies to K_2 such that the mass (q) of that belief is assigned to (n) and the remainder to (m, n). Using Dempster's rules of combination (Dempster 1967), the resulting forms can be combined yielding the following assessment in the solution space:

$$\left[(m), \frac{p(1-q)}{1-pq}\right], \left[(n), \frac{(1-p)q}{1-pq}\right]. \left[(m,n), \frac{(1-p)(1-q)}{1-pq}\right]$$

using symbolic math from calculations in *Mathematica*™. Applying the expression to several values of p and q yields the following results:

K_1 (p)	K_2 (q)	$K_1 \oplus K_2$
.6	.7	{[(m), 0.310], [(n), 0.483], [(m, n), 0.207]}
.7	.5	{[(m), 0.538], [(n), 0.231], [(m, n), 0.231]}

This is what we wanted to show. QED.

Abductive Inference

An alternate methodology, known as *abductive inference*, is used to determine the probable cause of group membership. (Josephson 1996) As with many forms of diagnosis, we have an event or condition and wish to determine the probable cause of the occurrence. A person may have a condition, such as liver disease, or a physical system may fail, such as a fighter aircraft. The list of probable causes in each instance is called the *differential*. Abductive inference is often referred to as, "turning modus ponens induction on its head." Abduction takes the following pattern:

E is an event or a collection of data
C explains E
No other hypothesis can explain E and well as C

Therefore, C is probably true

For example, J attends a training camp that is associated with membership in a militant group. There could be several possible causes for this phenomenon, two of which are that J is a militant or wants to be one. J could also be a journalist wanting to find out about things, but that is definitely less probable. The use of subjective probabilities, assigned through abductive inference, can be an analytic technique in its own right, or it can be used as input to a "combination of evidence" methodology.

SUMMARY

Personal identity and privacy are important topics in the modern world of communications and the Internet. Most citizens are not aware of the major issues or do not realize the serious nature of identity theft and privacy invasion. The academic community is needed to foster attention to this subject, and this paper attempts to spotlight the major concerns. Accordingly, this paper is an admixture of topics that include identity, identity theory, privacy, and privacy theory, along with a summary of the major aspects of each domain. There is a due consideration given to identity requirements through a set of identity principles and some proposals for identity analytics. In the latter instance, evidential methods are presented as promising research topics.

AUTHOR INFORMATION

Dr. Harry Katzan is the author of books and papers on computer science, decision science, and service science and is the founding editor of the *Journal of Service Science.* His current research interests are in the area of a strategy for trusted identity in cyberspace.

REFERENCES

1. ACLU of Northern California. 2010. *Cloud Computing: Storm Warning for Privacy?* www.dotrights.org, (downloaded 3/11/2010).
2. Black, M. 1952. Identity of Indiscernibles. *Mind* 61:153. (Secondary reference.)
3. Cavoukian, A. 2009. *Privacy in the Clouds*. Toronto: Information and Privacy Commission of Ontario (www.ipc.on.ca).
4. Cavoukian, A. 2010. 7 Laws of Identity: The Case for Privacy-Embedded Laws of Identity I the Digital Age." Toronto: Information and Privacy Commission of Ontario (www.ipc.on.ca).
5. Center for Digital Democracy (CDD). 2009. Online Behavioral Tracking and Targeting: Legislative Primer September 2009. www.democraticmedia.org/privacy-legislative-primer. (downloaded 3/11/2010).
6. Conti, G. 2009. *Googling Security*. Upper Saddle River, NJ: Addison-Wesley.
7. Federal Bureau of Investigation. 2004. Privacy Impact Assessment. www.fbi.gov/biometrics.htm. (downloaded 2/20/2010).
8. Dempster, A.P. 1967, "Upper and Lower Probabilities Induced by a Multivalued Mapping," *The Annals of Statistics* 28:325-339.
9. Givens, B. 2009. Privacy Today: A Review of Current Issues. Privacy Rights Clearinghouse. www.privacyrights.org/ar/Privacy-IssuesList.htm, (downloaded on 2/15/2010).
10. Josephson, J. and S. Josephson 1996. *Abductive Inference: Computation, Philosophy, Technology.* Cambridge: Cambridge University Press.
11. Katzan, H. 1975. *Computer Data Management and Data Base Technology*, New York: Van Nostrand Reinhold Co.
12. Katzan, H. 1980. *Multinational Computer Systems: An Introduction of Transnational Data Flow and Data Regulation*. New York: Van Nostrand Reinhold Co.
13. Katzan, H. 2006. Consensus. Proceedings of the Decision Science Institute Annual Meeting, (San Antonio, TX, November 17-21, 2006).
14. Katzan, H. 2010a. Identity Analytics and Belief Structures. *Journal of Business & Economics Research,* 8(6):31-39.
15. Katzan, H. 2010b. On the Privacy of Cloud Computing. *International Journal of Management and Information Systems,* 14(2):1-12.
16. Katzan, H. 2010c. *Privacy, Identity, and Cloud computing*, New York: iUniverse, Inc.
17. Miller, A. 1971. *The Assault on Privacy: Computers, Data Banks, and Dossiers*. Ann Arbor: The University of Michigan Press.
18. Nelson, M. 2009. Cloud Computing and Public Policy. Briefing Paper for the ICCP Technology Foresight Forum. JT03270509, DATI/ICP(2009)17.
19. OECD 2010. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. www.oecd.org. (downloaded 3/23/2010).
20. Russell, B. 1912. *The Problems of Philosophy*. (Republished by Barnes & Noble, New York, 2004).
21. Russell, B. 1919. *Introduction to Mathematical Philosophy*. (Republished by Barnes & Noble, New York, 2005).
22. Salido, J. and P. Voon. 2010. A Guide to Data Governance for Privacy, Confidentiality, and Compliance: Part 1. The Case for Data Governance. Microsoft Corporation.
23. Shafer, G. 1976. *A Mathematical Theory of Evidence,* Princeton: Princeton University Press.
24. Stroll, A. 1967. *Identity*. (Entry in *The Encyclopedia of Philosophy*, Volume 4, Paul Edwards, Editor in Chief, New York: Macmillan Publishing Co., 1967).
25. U.S. Department of Health, Education, and Welfare (HEW) 1973. *Records, Computers, and the Rights of Citizens*, Cambridge: The M.I.T. Press.
26. Westin, A.E. 1977. *Computers, Health Records, and Citizen's Rights*, Princeton: Petrocelli Books, Inc.
27. Westin, A.F. 1967. *Privacy and Freedom*. New York: Atheneum.
28. Windley, P. 2005. *Digital Identity,* Sebastopol: O'Reilly Media, Inc.