Journal of Computer Science 8 (1): 107-120, 2012
ISSN 1549-3636

Engineering Privacy Revisited

Sabah Al-Fedaghi
Department of Computer Engineering,
Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

Abstract: Problem statement: Information Privacy Engineering (IPE) is the field that studies the protection of privacy in information and communication systems. The theoretical, technological and applications aspects of IPE require a framework that provides a general view and a systematic structure for the discipline's topics. This study discusses certain characteristics of such a framework and proposes enhancing and strengthening its structure. **Approach:** Several important problems and their solutions are presented through recasting of some current proposed approaches to Personal Identifiable Information (PII) definitions and handling. **Results:** A new framework is presented that is based on flow-based model, along with generic operations performed on PII. **Conclusion:** This study shows that the flow-based model can provide a structure that complements current efforts to develop a framework for IPE.

Key words: Conceptual model, information privacy, information security, privacy-enhancing technologies, Information Privacy Engineering (IPE), Personal Identifiable Information (PII), applications aspects

INTRODUCTION

Progress in Information and Communication Technology (ICT) is rapidly transforming society in ways that intensify interactions among citizens, businesses and government. ICT has witnessed fast developments in storage, processing and communication of information at unprecedented speed and volume. These developments have allowed for a greatly increased volume of collected personal data and the capacity to manipulate information. Governments and companies have been quick in applying ICT to enhance their functions and services. The capacity to assemble this information for commercial and government operations represents a great risk to privacy, as in the areas of handling of Personal Identifiable Information (PII) using data mining (Al-Saidi *et al*., 2009) (Masrom *et al*., 2011). Surveys indicate that privacy has persisted as an issue that causes concern among individuals and fear among consumers (HealthcareTechnologyNews, 2010), (Greene, 2009), (Tsai *et al*., 2011), (Roberts, 2005), (PAB, 2005). According to Marsh *et al*. (2008):

> Many information systems are designed to collect and store all available data, because filtering and selecting takes more effort and the benefits of investment in the system may grow if new ideas to extract value from data emerge in the future… But organisations that adopt this approach are doing more than storing up information they are also storing up problems for themselves as public concerns over privacy grow, privacy issues rise up the political agenda and legal sanctions for privacy violations increase. [Italics added]

Additionally, the introduction of stronger privacy laws and regulations and standards, reports of high-profile privacy failures and increasing public concerns have built a case for enterprises "to take privacy seriously" (Marsh *et al*., 2008).

Consequently, as could be expected, privacy has been developed over the years as a relevant field of study in engineering systems. According to Spiekermann and Cranor (2009):

> Privacy is a highly relevant issue in systems engineering today. Despite increasing consciousness about the need to consider privacy in technology design, engineers have barely recognized its importance

IPE is proposed as an approach that integrates privacy into the development of a project to ensure that privacy protection is taken into account. This integration covers all phases of the system: planning, design, testing, operations, maintenance and periodic reviews (Booz Allen, 2010). IPE is concerned with privacy policies, compliance mechanisms and technology.

The need for privacy engineering will be escalating substantially in the coming years as federal agencies increasingly turn to Internet-based cloud computing to manage vast databases more efficiently. These applications can potentially make personally identifiable information maintained by the government more accessible to unauthorized individuals. Without privacy engineering during the design, initiation, implementation and maintenance of cloud programs, data protection and accessibility standards will become increasingly challenging for agencies to properly control (Booz Allen, 2010). Marsh *et al*. (2008) declare that "Privacy requirements must be fed in at each of the four stages initiation, planning, execution and closure of a generic project lifecycle".

Several proposals have been published to build a framework for privacy engineering. Earp *et al*. (2002) proposed "a framework, for examining an organization's privacy management practices within the context of their respective privacy policies". Feigenbaum *et al*. (2002) studied digital-rights management technology with respect to compromising and protecting users' privacy. In a pioneering 2009 article in IEEE Transactions on Software Engineering, Spiekermann and Cranor (2009) presented a framework as:

> A view of the privacy field, situating each approach to privacy in a spectrum of system design options… [and] derive system requirements from accepted privacy definitions as well as from user concerns and propose a framework that integrates existing research to provide engineers a clear roadmap for building privacy friendly information systems

Spiekermann and Cranor (2009) use a three-layer model of user privacy concerns to be applied to system operations (data transfer, storage and processing) and develop guidelines for building privacy-friendly systems. They distinguish between two approaches: "privacy-by-policy" and "privacy-by-architecture". Since their approach applies to a wider variety of systems, we will focus on their study as a sample of current state-of-art in the field of IPE.

Spiekermann and Cranor (2009) also mention several privacy design frameworks such as those of Earp *et al*. (2002); Hong *et al*. (2004) and Feigenbaum *et al*. (2002) that are applied in specific applications such as commerce websites and ubiquitous computing applications. They propose that their approach "applies to a wider variety of systems including e-commerce websites and ubiquitous computing applications". Thus their study is first in developing an explicit framework for privacy engineering that "integrates existing research to provide engineers a clear roadmap for building privacy friendly information systems" (Spiekermann and Cranor, 2009).

This study is a sequel to Spiekermann and Cranor (2009) study published in IEEE Transactions on Software Engineering. This study includes new development in the field of privacy because Spiekermann and Cranor (2009) study focuses only on certain research study. The new materials complement other concepts, definitions and models that have appeared in recent publications. This will result in a firmer skeleton that can be utilized as a framework for IPE.

Motivations for revised framework: A framework is an abstract description of the underlying structure that supports something; in our case, it supports IPE. It includes logical formation of meaning that integrates and directs the growth of research in the field of study. Privacy frameworks can achieve a firmer foundation and more coherent structures for this purpose by incorporating diverse privacy research. This study gives sample justification for a revised framework. We concentrate on scrutinizing Spiekermann and Cranor (2009) study as the most comprehensive and recent study aiming to develop a framework for the field.

Issue of definition: A very important issue in the context of IPE is that of defining the elementary constituents or fundamental units of informational privacy. Spiekermann and Cranor (2009) use at least 11 terms to name the types of "data" involved in IPE: personal data, personally identifiable data, personal information, identifying data, identifiable personal data, privacy information, identifying information, personally identifiable information, identity information and privacy related information. They do not explicitly define these types of data. This is a serious issue because the data are the "things" around which (informational) privacy revolves. How would we develop a framework for numbers theory without defining numbers?

This study complements proposed frameworks for IPE with a reasonable definition of Personal Identifiable Information (PII).

The definition of PII carries the issue of the notion of identifiability. Spiekermann and Cranor (2009) define "identifiability" as "the degree to which data can be directly attributed to an individual". Defining identifiability in terms of "data attributed to an individual" is not suitable in the engineering context. The dictionary meaning of attribute includes "a characteristic or quality of a person or thing". In IPE, some construction-based definition is more suitable. In this study we clarify the relationship between privacy and identifiability in any privacy-related framework.

Issue of analysis: Spiekermann and Cranor (2009) also introduce "an analysis of privacy sensitive processes" in order to understand "what user privacy perceptions and expectations exist and how they might be compromised by IT processes … to understand the level of privacy protection that is required". Accordingly, they claim:

> All information systems typically perform one or more of the following tasks: data transfer, data storage and data processing. Each of these activities can raise privacy concerns. However, their impact on privacy varies depending on how they are performed, what type of data is involved, who uses the data and in which of the three spheres they occur. [Italics added]

"All information systems typically perform one or more of the following tasks: data transfer, data storage and data processing" seems to be a questionable claim. This is a problem that will be discussed in this study. What is a "task" in this context? Is it the operating system concept that refers to execution and bookkeeping of information? Is it synonymous with "process"? In developing a new framework for IPE, these issues of basic notions are essential ingredients in developing the field.

In the framework to be proposed in this study, there are six mutually exclusive processes built according to the condition of data: processing, creation, releasing, transferring, accepting and arriving, all interwoven in a flow system that specifies the transformation from one process to another.

Take the creation of new PII as an example. Spiekermann and Cranor (2009) lump two types of handling of PII under the term "processing": a mere operation upon data to change its form and a process that creates new data. Creating data, in the context of privacy, is a far more sensitive task than operating on data without creating new data. For example, using data mining and merging diverse databases to create the new information that John is a terrorist is far more sensitive than the process of searching for John in a list of tourists. In the first case, the process embeds a judgment that produces a conclusion that was not previously present in the system, while in the second case the system already has this information. There is a great difference between processing information about a person and issuing a judgment about him/her. It is possible to declare privacy rules that permit a mere process but prohibit processing that creates new information. Lumping these types of data handling causes ambiguity in understanding the objects in IPE: created PII and processed PII. We will discuss a set of operations that separate processing of data from creation of data.

Problems in definitions and analysis presented so far give a taste of many other problems in currently proposed frameworks. To provide opportunities for recasting and comparing the proposed approach with current approaches, then gives the new foundation by reviewing notions that have been introduced over several years in many publications, including, (Al-Fedaghi, 2006a; Al-Fedaghi, 2008; Al-Fedaghi, 2009; Al-Fedaghi, 2007a; Al-Fedaghi, 2007b; Al-Fedaghi, 2007c; Al-Fedaghi, 2007d; Al-Fedaghi, 2007e; Al-Fedaghi, 2007f; Al-Fedaghi, 2006b; Al-Fedaghi, 2006c; Al-Fedaghi, 2006d; Al-Fedaghi, 2005; Al-Fedaghi and Ahmad, 2006; Al-Fedaghi and Al-Haqan, 2009; Al-Fedaghi and Al-Turjman, 2007; Al-Fedaghi and Jeragh, 2011; Al-Fedaghi and Taha, 2006; Al-Fedaghi and Thalheim, 2008; Kangassalo, 1999; Zailani and Norjihan, 2009; Sato *et al*., 2009). Some materials are presented in new ways.

Foundation for new framework: We deal first with the problem of PII definition discussed previously. First, PII is defined along with a method to tie identity with PII. Second, the issue of analysis (data transfer, storage and processing) is recast through introduction of a flow-based model that specifies various processes involved in handling of PII.

Personal identifiable information**:** Davenport, (2007) claims that "personal information is the core of privacy"; thus, it is important to develop a workable definition of PII. This is a controversial issue. The most recent declaration in this context is an article in Communications of the ACM (2010) that labeled PII worthless in the context of privacy laws that "account for the possibility of deductive disclosure and… do not lay down a list of informational attributes that constitute PII". Narayanan and Shmatikov (2010) declares that

> For a concept that is so pervasive in both legal and technological discourse on data privacy, PII is surprisingly difficult to define. […] PII is meaningless, […] The term means next to nothing and must be greatly de-emphasized, if not abandoned, in order to have a meaningful discourse on data privacy

Narayanan and Shmatikov (2010) see that PII may present some difficulties “with respect to which there is a reasonable basis to believe the information can be used to identify the individual”. Nevertheless, we claim that these difficulties do not reach the level of making the notion worthless in the context of privacy laws that “account for the possibility of deductive disclosure and… do not lay down a list of informational attributes that constitute PII”.

Differentiating PII from personal information: Spiekermann and Cranor (2009), as we discussed before, used many terms for the type of “data” involved in IPE; thus it is difficult to determine whether their study targets PII. One of the terms they use is “personal data,” which is the term used in EU (1995):

> Article 2a: ‘personal data’ shall mean any information relating to an identified or identifiable natural person (‘data subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity

We claim that this is not suitable because “personal data” as in the sense of “personal property” may include PII and non-PII, as shown in Fig. 1.

Personal data (information): Non-personal information means ownership of information that is not necessarily PII (e.g., a personal recipe). PII means proprietorship and is not necessarily owned by the proprietor. A proprietor (the real person to whom PII refers) may not even know about his/her PII, as in the case of a disease not disclosed to a patient.

According to Spiekermann and Cranor (2009), “Personal data can be entered into a system anonymously (e.g., e-voting) or by identifying oneself (e.g., when conducting online banking transactions)”. Here a personal action (e.g., voting) is mixed with creation of data. Anonymous e-voting, e.g., that creates non-PII “candidate 17” is not personal information, analogous to a person who donates personal belongings to a charity (e.g., clothing); they are no longer his/her personal things.

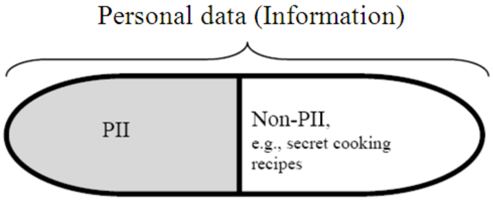

Fig. 1: Personal information subsets

PII definition: The definition we adopt is based on a diluted version of the correspondence theory that relates truth to reality: a statement is true if and only if the world it describes is real. In our case, a statement is PII if and only if the world it describes includes a singly identifiable real natural person. In logical terms, if the statement includes a referent to a singly identifiable person, called its proprietor, then it is PII. John is trustful is PII of John (assuming that John is a unique identification) in light of the fact that the statement John is trustful is about or can be mapped to a unique real person. Thus, John F. Kennedy was the 35th president of the USA is PII, whereas John F. Kennedy is a busy airport is not PII.

The referent is said to be the proprietor of PII. Proprietorship of PII is different from the concepts of possession and copyrighting. It is also different from the legal concept of ownership. The “referent” is recognized by mapping the word (logical name) in relation to the actual object (natural person) in reality. This mapping to a natural person limits possible extension to specific human beings.

Accordingly, there are two types of PII (Al-Fedaghi, 2005):

- Atomic PII (APII) that embeds a single proprietor
- Compound PII (CPII) that refers to more than one proprietor

Proprietorship of PII is nontransferable in the absolute sense. Others can possess or (legally) own it, but they are never its proprietors (i.e., it cannot become their proprietary information). APII of a proprietor is proprietary information of that proprietor, while others (e.g., other individuals, companies) can only possess it. CPII is proprietary information of its referents: all donors of pieces of atomic PII embedded in the compound PII.

Defining PII as “information identifiable to the individual” does not mean that the information is “especially sensitive, private, or embarrassing. Rather, it describes a relationship between the information and a person, namely that the information whether sensitive or trivial is somehow identifiable to an individual" (Kang, 1998). We don't discuss the issue of sensitivity of PII, which is hard to define, but difficulty in defining this type of PII does not make PII a worthless notion, the way difficulty in defining many notions such as indecent material, consent and so forth makes the notions worthless in privacy laws. Anwar (2008) criticized the definition of PII above since it "includes observation, reputation, or even public information in the realm of personal information and thereby, may introduce more ambiguity. For example, information referring to John in his professional capacity as mayor, for example, should not be considered as his personal information". This is analogous to criticizing the standard definition of integer set because it is infinite. The point is that this definition is inclusive/exclusive with regard to membership. Only after specifying the membership inclusion can we define other subsets, as with a primary integer in mathematics and sensitive PII in privacy.

Clearly, much of PII, as defined insignificant in terms of privacy. Insignificance does not imply lack of value. An insignificant amount of gold not worth the effort to mine is not worthless. Even though no criterion precisely divides significant from insignificant types of PII, it seems that, in most cases, the difference between them is apparent. Many works in the area of privacy have no difficulty identifying (significant) privacy in domains such as health information and financial information. "Significance" here refers to the threshold of an intrinsic value of PII.

PII and identifiers: The world of PII comprises pieces of information that refer to real natural persons. PII can be a composite of other PII or can be no composite.

Let us call the constituents (better, the minimum constituents) of PII that refer to proprietor signifiers. The correspondence relation collapses into a type of identity function that maps signifiers (domain) to proprietors (range). The proprietor is a real object in the world, so there is no need to get involved in a semantic dilemma such as "normal American," "the present king of France," or "unicorn".

The question now is how simple signifiers (e.g., an identifier) can be about proprietors. John A. Smith is something can be tested as PII by scrutinizing the mapping from the signifier "John A. Smith" to John A. Smith, separately from "is something" and its correspondence to reality. If there are two John A. Smiths, then John A. Smith is something is not PII. For example, on the technical side, a file containing information about a patient such as the signifier "examined in room 110 on Friday, 26 December" can be automatically classified as PII if there is no other entry (patient) in the file with the same information.

Consider the set of unique identifiers of persons. Ontologically, the Aristotelian entity/object is a single, specific existence (a particularity) in the world. In this study, the identity of an entity is its natural descriptors (e.g., tall, brown eyes, male, blood type A). These descriptors exist in the entity/object. Tallness, whiteness, location, etc. exist as aspects of the existence of the entity. We recognize the human entity from its natural descriptors. Some descriptors form identifiers. A natural identifier is a (minimum) set of natural descriptors that facilitate recognizing a person uniquely. Examples of identifiers include fingerprints, faces and DNA. No two persons have identical natural identifiers. An artificial descriptor is a descriptor mapped to a natural identifier. Attaching the number 123456 to a particular person is an example of an artificial descriptor in the sense that it is not recognizable in the (natural) person. An artificial identifier is a (minimum) set of descriptors mapped to a natural identifier of a person. By implication, no two persons have identical artificial identifiers. If two persons somehow have the same Social Security number, then this Social Security number is not an artificial identifier because it is not mapped uniquely to a natural identifier.

A basic principle in the definition of PII is as follows:

Identifiers of proprietors are PII. Such definition is reasonable since the mere act of identifying a proprietor is a reference to a unique entity. Every unique identifier of a person is a basic PII in the sense that this identifier cannot be decomposed into more basic PII

The second principle defines PII in general: Any personal identifier or piece of information that embeds identifiers is personal identifiable information.

Thus, identifiers are the basic PII that cannot be decomposed into more basic PII. Furthermore, every complex PII includes in its structure at least one basic identifier. Note that the concern here is not issues of flexibility or narrowness of PII definitions. This is a matter that can be settled after developing a precise definition encompassing all types of PII.

Handling PII: Spiekermann and Cranor (2009) claim that "all information systems typically perform one or more of the following tasks: data transfer, data storage and data processing. They do not provide any justification for such a declaration.

Basic flow model: A typical information system is defined as a system that transforms input into output. Storage is also added as an element in such a description. Alternatively, several study, as mentioned previously, have adopted a "flow-thing" machine, denoted a flow system that is an abstract machine with six components (or stages): arrival, acceptance, processing, creation, release and transfer (Al-Fedaghi and Al-Saleh, 2011). The component or stage here corresponds to the conditions of the things that flow, called flow things (e.g., PII), inside the machine. This is analogous to describing a chocolate plastique manufacturing system in terms of the transformations between various stages of the system: raw chocolate, melting , molding and packaging stage.

We adopt the conceptualization that all information systems perform the following actions: creation, processing, release, transfer, accepting and arriving, or a subset of these actions and we illustrate the transformations among these actions or processes (Fig. 2)

Figure 2 depicts an abstract machine (flow system) under the assumptions that arriving flow things are never rejected and released flow things are never returned. The transfer stage has a reflexive arrow, denoting flow to another transfer stage in another system. The creation stage indicates generation of new flow things (e.g., PII). Processing of flow things refers to changing flow things in form or action, but never to newly generating flow things.

A Flow Model (FM) involves modeling of enterprises using flow systems. A flow system is composed of three primitive concepts:

- Six stages, as mentioned previously
- Transformation among stages (arrows in Fig. 2)
- Flow things: things that flow in the flow system Flow things are in flow systems of different types of objects, i.e., physical objects or conceptual objects and in six stages representing a conceptual place that handles flow things of a certain type

In a flow system:

- The arrival stage handles arriving flow things from outside the system through an interface called the transfer component (the delivery/receiving component of the system)
- The acceptance stage handles flow things passing through the arrival stage. For the sake of brevity, arrival and acceptance may sometimes be merged into one state called receiving
- The processing stage handles processed flow things, e.g., in the chocolate plastique example, coloring, sweating, or smoothing the chocolate. Sources to the processing stage are flows from the creation stage and from the acceptance stage
- The creation stage creates flow things and handles these created flow things
- The release stage is an intermediate stage for outward-bound flow things
- The transfer stage is the interface with the outside such as ports in communication devices

An arrived flow thing cannot be in two of the stages simultaneously. Other conditions of flow things such as being stored, copied, destroyed, etc. are secondary conditions with respect to the six generic conditions. For example, stored flow things can be found at the arrival, acceptance, creation, processing, release and transfer stages.

Triggering: Flows in FM may trigger each other, represented as dashed arrows. To illustrate the FM-based representation and contrast it with a sample typical specification, consider a diagram such as a functional flowchart with "swim lanes" to represent who is responsible for or performs an activity (Fig. 3).

In scrutinizing this method of process specification, we observe how sketchy it is. The use of arrows is overdone, with control flow (the diamond decision shape), orders flow (e.g., from operator to department) and information flow (e.g., the manager needs information to check an order).

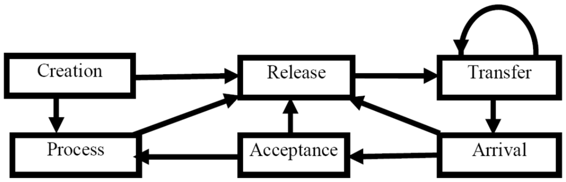

Fig. 2: The FM machine (assuming flow things are always accepted and transferred when released)

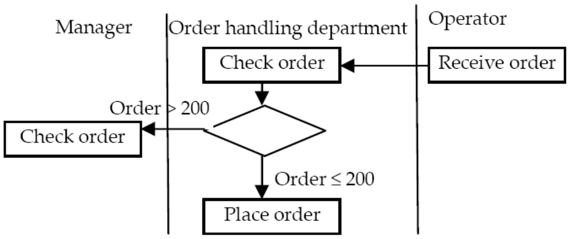

Fig. 3: Swim lanes (Mutschler 2006)

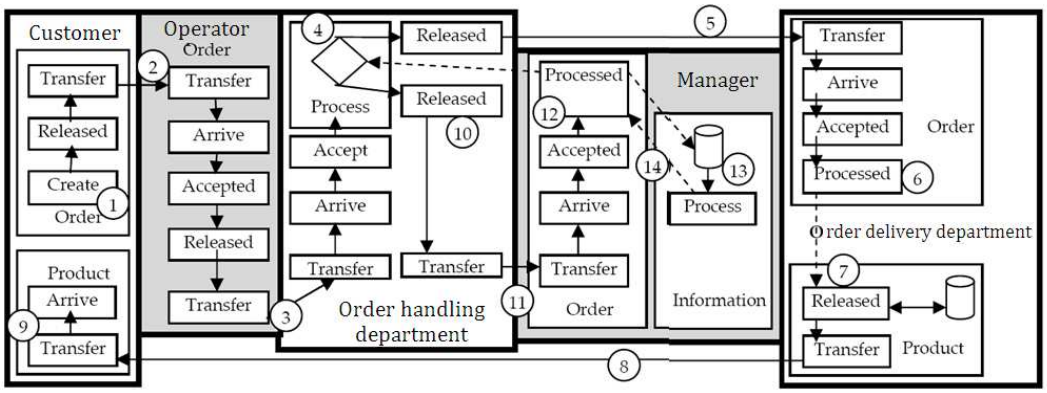

Fig. 4: FM description of the Fig. 3 flowchart diagram using swim lanes

The figure is also incomplete. From where does the operator get an order? What happens when an order is placed (e.g., before product can be sent)? These features become clear when the corresponding FM representation is developed, as in Fig. 4.

In the figure, the Customer creates an order (circle 1), sends it to the Operator (circle 2), who releases it to the Order Handling Department (circle 3). The meaning of the "place order" box in Fig. 3 is not clear because the order is already placed and being processed. It may mean that there is another administrative unit that delivers the order after checking. Accordingly, we add the order delivery department (it could be a subsection of the order handling department).

At circle 4 in the order handling department there are two possible flows of orders:

- Flow to delivery department (circle 5) where the approved order is received and processed (circle 6) to trigger delivery of product (circle 7). For delivery, the product is retrieved from storage and sent to the customer (circle 8), who receives it (circle 9)
- Alternatively, at circle 4 in the order handling department, the order is sent to a manager (circles 10 and 11), who processes it (circle 12)

Here, the original sketch in Fig. 3 seems to be incomplete in specifying what the manager needs to check for the order. We assume that he/she consults stored information (circle 13), which leads to a decision to trigger (circle 14) further processing by the Order Handling Department, from which the order goes to the Order Delivery Department or is denied (e.g., stopped). Notice that we have five spheres (the environment of the flow systems) in Fig. 4, Customer, Operator, Manager and the Handling and Delivery departments.

The details of the decision at circle 4 (diamond shape), which rests on a certain value of order (>200, ≤200) is an implementation issue that is not specified at this level of conceptual modeling. We are interested here in basic flows and the position of critical components, analogous to a blueprint for a high rise where, say, the exact voltage of a circuit breaker is decided at a later stage of details. Other descriptions such as OR and, … can easily be overlaid on the basic FM specification.

In contrast to Fig. 3, the description in Fig. 4 is complete and systematic. Flows are separated and put in sequences. Figure 4 appears more complicated than Fig. 3, but this is needed for completeness. In an analogy, the specifications of a building can be included in a mere sketch where flows of water, electricity and gas are not distinguished and interior floors are not specified, but this is not simplicity; it is incomplete and a poor differentiation of flows.

PII as flow things: PII can be created, processed, released, transferred, received and accepted into a system. This is an alternative conceptualization to Spiekermann and Cranor (2009) three tasks of data transfer, data storage and data processing. Besides being an incomplete representation, storage is shown by FM, theoretically, not to be a genuine task in information systems. Thus FM along with the definition of PII can form a foundation for any system that handles PII, creating a potential base for IPE.

The PII flow system is an alternative to Spiekermann and Cranor (2009) tasks of data transfer, data storage and data processing. These three tasks are incomplete in the sense that they do not represent all genuine "tasks" that can be performed on PII.

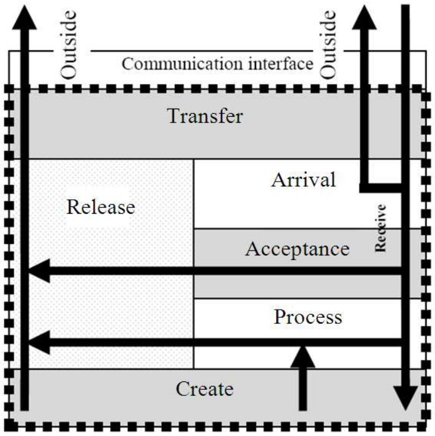

Fig. 5: Flow of PII to/in/from the hospital

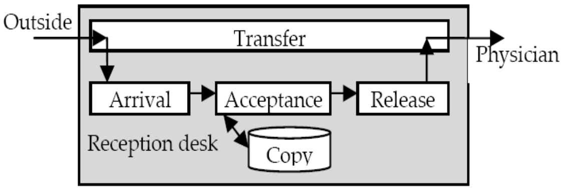

Fig. 6: Description of the secondary stage, storage

Consider the flow of PII in hospital H shown in Fig. 5. H has a communication interface (e.g., electronic or physical) that connects it with the outside. PII may originate from outside and enter H through its communication interface (e.g., H's connections to the Internet). PII then arrives (is received) in H.

Notice the difference between transferred and arrived states of PII in H. Suppose that PII is in the form of a physical DHL folder, denoted as x, to be delivered to a certain section in H. From the time x enters H until the time the receipt is signed by the section clerk, x is in H, but it is in the state of being transferred. In its transfer state, x may be carried by the DHL delivery person inside H to reach the destination section, or x may be deposited into a special mailbox in H, but x has not yet been acknowledged as arriving in the H information system. As soon as the H clerk signs the delivery receipt, x has arrived.

This difference between transfer and arrival components of a system is significant in all applications. For example, in the security area, if PII is compromised because an unauthorized person has a key to the H mailbox, this is a problem in the H transfer system (e.g., not in H, rather, for the contracted mail company). If PII is compromised because an unauthorized person in H signed the receipt, this is a problem in H's arrival system.

Similarly, for electronic PII, packets reach the communication fiber of H from, say, a public network and are transmitted as packets to the first device, e.g., a router. PII is in H, but in the transferred state. As soon as it settles in the first buffer, it has arrived at H. If a sniffing device is installed in the H communication line (which does not belong to the public network) before the router's buffer, this is a security problem at the transfer stage of H. If the data is compromised in the buffer of that router, it is an arrival stage problem.

Arrival does not guarantee acceptance; however, in Fig. 2, we assume that PII that arrives is accepted; otherwise the arrow between Arrival and Transfer would be bidirectional. After acceptance, PII can be processed. The meaning of processing here is more limited than the general usage of this term. It means any type of operation that changes PII (e.g., decimal to binary, compression, translation) without causing it to lose its identity.

Processing may be a cause of creation of new PII (e.g., data mining). PII may be created directly in H, as in the case of a laboratory that produces John is blood type O, for a patient John. John is blood type O has never reached H from outside H. It may be processed internally, or released to be transferred outside, as shown in Fig. 5.

Release is different from Transfer in H. PII may be released to be sent electronically to the outside; however, it can still be in H if the communication channel is busy or down. PII may stay in the released state for a while until it is transferred. There is a possibility that it will be returned to the sender after a certain period of waiting for transfer. In Fig. 2 and 5, we assume that released PII is eventually transferred.

The reader may wonder where such states of PII as being stored, copied, destroyed and so forth are located in the flow system. These are not primary states of PII that can occur anywhere in the system. For example, if the desk in H copies PII, then forwards it to the physician, this can be shown as in Fig. 6. In this case the copy is a different flow thing from the original.

In the remaining part of this study we apply our framework to certain issues raised by Spiekermann and Cranor (2009).

Spheres and responsibility: For the purpose of "framing privacy for engineering," Spiekermann and Cranor (2009) classify engineers' responsibilities into three distinct technical domains: the user sphere, the recipient sphere and a joint sphere:

> The “user sphere” encompasses a user’s device. From a privacy perspective, user devices should be fully controllable by the people who own them. Data should not flow in and out of them without their owners being able to intervene. Additionally, devices should respect their owners’ physical privacy, interrupting them only when needed and at appropriate times.

Such a description interweaves components of spheres: people (reality), which includes ownership (concept), together with devices (hardware). We revise this classification of engineers’ responsibilities in terms of FM.

Usually a sphere (of influence) denotes an area over which there is control or influence. It is usually represented by a map (e.g., city boundaries, governmental jurisdiction).

In information privacy, “usership” is a less important notion than proprietorship, which is a unique privacy trait. A proprietor has a special connection with his/her PII even after giving it up or selling it to someone, as in the case of ridiculing his/her name or picture, or misusing it. With regard to ownership, an owner does not care about what was in his/her possession (e.g., a chair) after selling it.

Consequently, a proprietor’s sphere ought to be of central importance. To illustrate the difference between use and proprietorship, consider the case of a handicapped proprietor who cannot use computers and gives instructions to an actual user (e.g., clerk) to use a certain software program to communicate his/her PII to a company that transfers it to a third party. FM allows for comprehensive conceptualization of spheres of proprietors, devices and software programs, a manual information system, organization, specified in terms of flow systems (Fig. 7); thus there are many spheres and a user may not actually be the end party in all transactions.

Spiekermann and Cranor (2009) utilize their description of “user sphere” in specifying an engineer’s responsibility and describing engineering issues related to stored data. They also develop classification of concerns; for example, in the user’s sphere the concerns are unauthorized collection, unauthorized execution, exposure and unwanted inflow of data. In FM, a “user” may be a proprietor or a nonproprietor. The privacy phenomenon is centered on the proprietor, not the user.

Spiekermann and Cranor (2009) also describe the “recipient sphere” as “a company-centric sphere of data control that involves backend infrastructure and data sharing networks”. This is a communication-infected view of PII exchange that involves senders and recipients. In FM, it is possible that the recipient is also the processor, the creator, the releaser and/or the transferor in the context of the flow system. A company is a flow system, a department in the company is a flow system and a section in the department is flow system … even a backend system and a network are flow systems (e.g., a communication channel is a flow system that receives, processes, creates (e.g., noise), releases and transfers PII).

As we see in both cases of user and recipient spheres, descriptions include such terms as devices, backend and networks. FM is a conceptual framework that does not involve hardware, software, or any implementation terminology. After all, the mere meaning of “framework” expresses an abstract description of the underlying structure that supports IPE.

Similar discussion can be applied to Spiekermann and Cranor’s “joint sphere of privacy control” that “encompasses companies that host peoples’ data and provide additional services” (2009). In FM, all people in a company are conceptualized uniformly as flow systems. This alternative conceptualization reflects a picture of the propagation of PII away from its proprietor. The proprietor is conceptualized as a flow system that transfers and receives PII to/from first-level nonproprietors, as depicted in Fig. 8. These first-level nonproprietors may transfer and receive information to/from other nonproprietors that do not communicate directly with proprietors.

Utilizing this framework in specifying privacy responsibility can be stipulated with respect to a proprietor’s sphere as PII arrives and is accepted, processed, created, released and transferred. This encompasses PII transferred by hardware (e.g., wired, wireless), PII processed by software (e.g., database applications and server) and many other angles such as how PII is created (e.g., manually or by data mining), how PII is stored at different stages (e.g., disks, buffers, hard copies) and so forth. For example, the engineers may decide on priorities of storage, with created information more valuable than information received from outside. An exhaustive list of all possible types of responsibilities can be developed for a proprietor’s sphere or other spheres. Similarly, a proprietor’s concerns can be categorized with respect to arrival, acceptance, processing, creation, release and transfer of information. For example, transfer concerns include security transfer with hardware, software and system considerations.

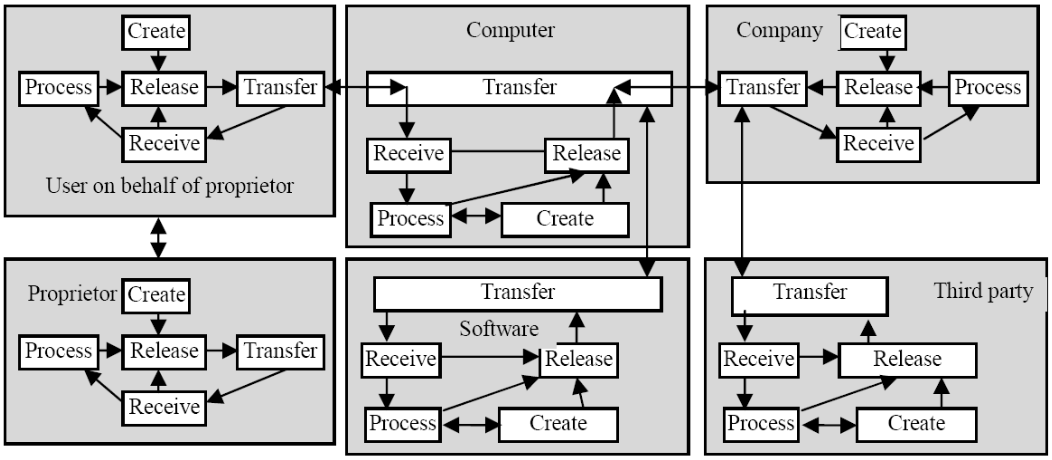

Fig. 7: In FM, proprietors, users, devices, software and companies are all flow systems

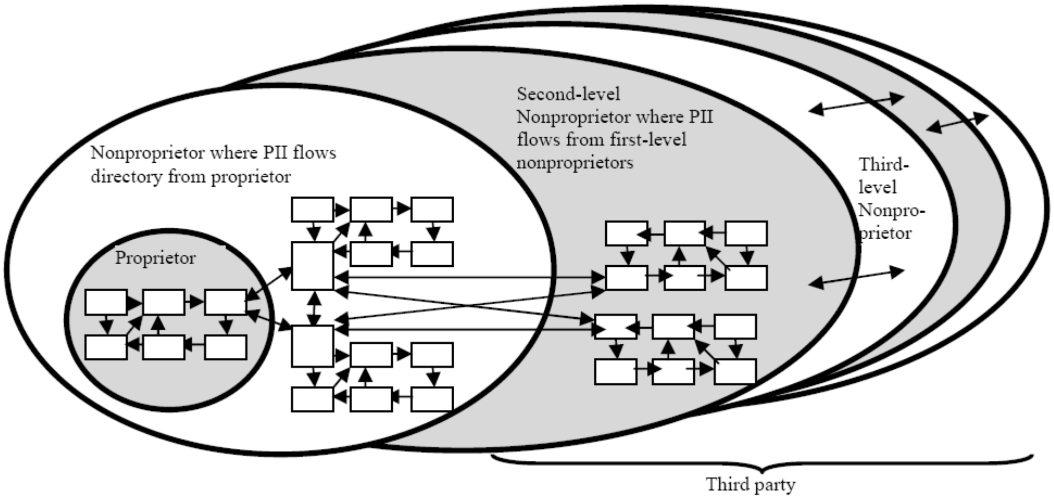

Fig. 8: A proprietor-collector-third party framework for “spheres” related to privacy

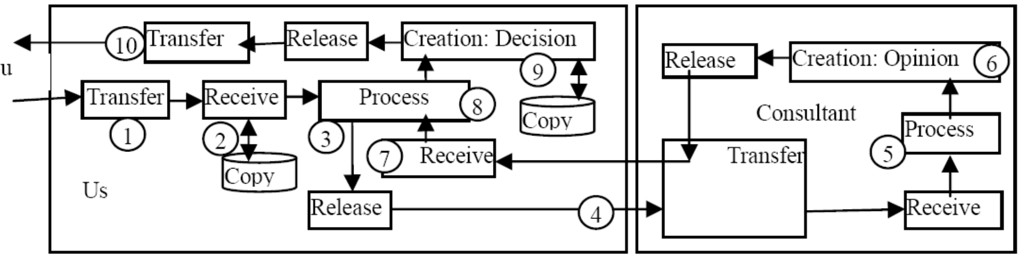

Fig. 9: FM specification of a sample notice

This framing is far more comprehensive than Spiekermann and Cranor’s (2009) arbitrary classification of concerns in the user’s sphere that include the four elements of unauthorized collection, unauthorized execution, exposure and unwanted inflow of data.

Notice as a map: As a “methodology for systematically engineering privacy friendliness,” Spiekermann and Cranor (2009) introduce the “notice and choice” approach based on the principles of Fair Information Practice (FIP) (EU, 1995). They discuss how this can be supported through "privacy-by-policy". According to them:

> US regulatory and self-regulatory efforts supported by the Federal Trade Commission (FTC) have over the past decade focused on a subset of these principles, tailored to the e-commerce context notice, choice, access and security… While the notice and choice approach is useful, it is not clear that it should serve as the golden rule for privacy design since notice, choice, access and security only come into play when a system collects personal data.

The US USFTC (2000) gives the following description of the notice requirement:

> Web sites would be required to provide consumers clear and conspicuous notice of their information practices, including what information they collect, how they collect it (e.g., directly or through non-obvious means such as cookies), how they use it, how they provide Choice, Access and Security to consumers, whether they disclose the information collected to other entities and whether other entities are collecting information through the site.

Again we notice the absence of a complete and generic list of possible types of handling of PII, l*et al*one the absence of a nonverbal definition of PII. Why mention only collection, use and disclosure of consumers' information? Suppose that an organization or government agency is generating PII about citizens through monitoring of their activities for some season. For example, someone produces Mary is a violent person, which is an observation or judgment. Technically, it is not collected information.

The point here is that listing random operations without basing them on formal or at least semiformal apparent definitions (e.g., PII) and systematic recognition of all possible operations is not a good methodology for establishing a framework for any field of study. In FM, a notice can be realized as a conceptual map of PII. Instead of verbal descriptions, a flow map can be constructed (automatically!) to give a complete conceptual specification. For example, an account of informing the user about his/her PII is shown in Fig. 9.

The notice can be described as follows:

> We take your PII (circle 1) and save a copy (circle 2) for one year, process it (circle 3) and then send it (circle 4) to our consultant (e.g., Intonation XYZ) (implicitly in the diagram, the consultant does not save a copy), who processes it (circle 5) to produce an opinion (circle 6). Upon receiving this opinion (circle 7), we process your case (circle 8) to make a decision that is stored (circle 9) for one year and send you the results (circle 10)

Interface design in the implementation of this type of "notice" is an important factor in keeping the proprietor informed. This interface can also be used as a map that includes all types of policy rules, such as at circle 1, "no PII is received from nonproprietor," and at circle 3, "processing does not apply data mining," etc. Decisions about when to interrupt users with privacy-related information can be agreed on according to this map.

The interface can also be used to inform the proprietor and/or the administration about the progress of the flow, or to generate alerts. This transparency is an important principle of fair information practices (Richardson, 1972) and would certainly increase trust in a company.

Fair information practice: According to Feigenbaum *et al*. (2002), the goals for practical privacy engineering are best presented by "fair information practices" (Richardson and Weinberger, 1973). "Although the FIPs are well understood, the technological literature has said relatively little on how to translate them into engineering principles" (Feigenbaum *et al*., 2002). Nevertheless, these principles are not well stated (Al-Fedaghi, 2007b). Consider the Collection Limitation Principle stated in the (Organisation for Economic Co-operation and Development, 2002):

> There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject

The (Organisation for Economic Co-operation and Development, 2002) use of the term "collection" can be interpreted to refer to the mere act of collecting PI. This is exactly the same as the first stage, receiving, in FM. The term is also used in another OECD principle: "the purposes for which personal data are collected should be specified not later than at the time of data collection".

The Canadian Standards Association Model Code for the Protection of Personal Information (2004), which are standards based on the OECD guidelines (1980), defines "collection" as "the act of gathering, acquiring, or obtaining personal information from any source, including third parties, by any means".

If this interpretation is appropriate, then we can ask, what about limits on other PII handling stages, namely, the processing, creating, receiving, releasing and transferring stages in FM? For example, data mining techniques can produce new PII that is not collected. Why don't we specify explicitly as in the case of collection that there are limits to the mining of PII and that any such data mining should be performed by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject? The same question can be raised with regard to manual internal examination of data (e.g., psychological character analysis) that also can produce new PII.

The point here is that limits can be specified on all stages of handling PII, not just on the collecting stage.

According to Al-Fedaghi, two [reinvented] notions are embedded in the (Organisation for Economic Co-operation and Development, 2002):

- Limiting the gathering of PII: e.g., lawful and fair
- Limiting the use and storage of PII with the knowledge or consent of the proprietor

These notions should be separated because they are sometimes not directly related. Obtaining PII by lawful and fair methods does not imply any consequences for methods of handling PII. For example, an agent can collect PI lawfully and still misuse it. An agent can limit its use of PII (to avoid being noticed) even though it might have collected this PII unlawfully.

The limitation principle can be rewritten to include seven rules to be observed by agents according to the types of PII. The first two rules, based on the above discussion, concern a gathering agent and limit the handling of PII at the collecting stage only:

- A gathering (collecting) agent should gather PII by lawful and fair means
- A gathering agent should use and store PII with the knowledge or consent of the proprietor

We can conclude that goals for practical IPE, mentioned previously as best presented by the fair information practice, are more comprehensively framed within FM.

Privacy enhanced systems: design: Fair information practices (USFTC, 2000) require avoiding collection of excessive information. An important question in this context is how to incorporate this principle into the design of the system. For Marsh *et al*. (2008), "It is important to avoid collecting excessive information. ... Anonymizing personal information, or permitting pseudonyms, can be an important privacy enhancing technique, though limitations of this approach need to be understood". Spiekermann and Cranor (2009) distinguish between "privacy-by-policy" and "privacy-by-architecture". "The privacy-by-architecture approach minimizes the collection of identifiable personal data and emphasizes anonymization and client-side data storage and processing".

In FM, "collection" in fair information practices can be viewed as the totality of privacy-related information in possession of the enterprise. A better term is "minimizing handling" of PII since handling in FM includes receiving (collecting), processing, creation, releasing and transferring of PII. This can be achieved only by separating PII from non-PII. Privacy-by-architecture and privacy-by-policy complement each other. Bad design leads to difficulties in declaring and mapping policies and incomplete policies do not take advantage of good design. Consequently, from the engineering point of view, the minimization principle can be achieved at two levels.

The design level: This level is the level of decisions regarding what to include in the system and how to minimize what is to be included. According to Marsh *et al*. (2008), "if the 'minimization' principle is not applied strongly enough in designing data collection, much more effort may need to go into privacy protection measures during data processing and storage". This requires minimizing meta-PII and organizing it such that it is recognizable. Meta-PII is information that describes PII such as NAME describes "John D. Smith". For example, the relational schema in a database specifies the set of attributes of EMPLOYEES, the required fields collected about each CUSTOMER. Recognition of meta-PII requires isolating it from meta-non-PII.

For example, Al-Fedaghi and Thalheim (2008) built databases for PII by developing relational schemas (relational tables) for both types of information separately.

Implementation level: This should not be mixed with the design level and is concerned with how to handle PII while applying the same minimal principle. Such tools as anonymization and pseudonyms are used.

CONCLUSION

The theoretical, technological and applications aspects of IPE require a framework that provides a holistic view and a systematic structure for the discipline's topics. This study has shown that the flow-based model where PII comprises the flow things can provide a structure that complements current efforts to develop such a framework. The study has presented several important problems and their solutions through recasting of some current proposed approaches to PII definitions and handling.

Many issues in this area are not discussed because the objective is to demonstrate the potential of this FM-based approach in complementing or providing a skeleton to develop the required framework. Future research in this direction can expand subjects that can be described based on our methodology, e.g., privacy preserving (Nardal and Sahin, 2011), privacy protection (Al-Fayoumi and Aboud, 2005) and application areas such as urban design (Ramezani, Hamidi, 2010).

REFERENCES

Al-Fayoumi, M. and S. Aboud, 2005. Blind decryption and privacy protection. Am. J. Applied Sci., 2: 873-876. DOI: 10.3844/ajassp.2005.873.876

Al-Fedaghi S. and A. Jeragh 2011. A flow-based model to assess privacy impact. J. Inform. Technol. Impact. 11: 101-120.

Al-Fedaghi S. and F. Al-Haqan, 2009. Privacy sensitivity: Application in arabic. Proceeding of the International Conference on Asian Language Processing, Dec. 7-9, IEEE Xploor, Singapore, pp: 156-161. DOI: 10.1109/IALP.2009.40

Al-Fedaghi S. and M. Taha, 2006. Personal information eWallet. Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, Oct. 8-11, IEEE Xploor, Taipei, pp: 2855-2862. ODI: 10.1109/ICSMC.2006.385307

Al-Fedaghi S.A. and B. Thalheim, 2008. Databases of personal identifiable information. Proceeding of the IEEE International Conference on Workshop on Security and Privacy in Telecommunications and Information Systems, Nov. 30-Dec. 3, IEEE Xploor, Bali, pp: 617-624. DOI: 10.1109/SITIS.2008.49

Al-Fedaghi S.S. and F. Al-Turjman, 2007. Conceptual modelling: A privacy perspective. Proceeding of the IEEE International Conference on Digital Ecosystems and Technologies, Feb. 21-23, IEEE Xploor, Cairns, pp: 416-421. DOI: 10.1109/DEST.2007.372009

Al-Fedaghi, S., 2005. How to calculate the information privacy. Kuwait University.

Al-Fedaghi, S., 2007a. Dismantling the twelve privacy purposes. Trust Manage., 238: 207-222. DOI: 10.1007/978-0-387-73655-6_14

Al-Fedaghi, S., 2007b. Wresting informational privacy from free speech. Int. J. Liability Scientific Enquiry, 1: 319-334. DOI: 10.1504/IJLSE.2008.018275

Al-Fedaghi, S., 2007e. How sensitive is your personal information? Proceeding of the 22nd ACM Symposium on Applied Computing, Mar. 11-15, ACM, New York, pp: 1688. DOI: 10.1145/1244002.1244046

Al-Fedaghi, S., 2008. Scrutinizing the rule: Privacy realization in HIPAA. Int. J. Healthcare Inform. Syst. Inform. 3: 16. DOI: 10.4018/jhisi.2008040104

Al-Fedaghi, S., 2009a. Drafting informational privacy laws: Information science perspective. Issues Inform. Syst., 10: 165-174.

Al-Fedaghi, S., Information privacy and its value. Kuwait University.

Al-Fedaghi, S.S. and M.Y. Ahmad, 2006. Personal Information Modeling in Semantic Web. Semantic Web ASWC., 4185: 668-681. DOI: 10.1007/11836025_65

Al-Fedaghi, S.S., 2006a. Anatomy of personal information processing: Application to the EU privacy directive. Inte. J. Liability Scientific Enquiry, 1: 129-138. DOI: 10.1504/IJLSE.2007.014586

Al-Fedaghi, S.S., 2006b. Personal management of private information. Proceeding of the IEEE International Conference on Innovations in Information Technology, Nov. 2006, IEEE Xploor, Dubai, pp: 1-5. DOI: 10.1109/INNOVATIONS.2006.301955

Al-Fedaghi, S.S., 2006c. Personal information flow model for P3P Kuwait University.

Al-Fedaghi, S.S., 2006d. Aspects of personal information theory. Proceeding of the IEEE Information Assurance Workshop, June 21-23, IEEE Xploor, West Point, pp: 155-162. DOI: 10.1109/IAW.2006.1652090

Al-Fedaghi, S.S., 2007d. Incorporating Personal Information into RDF. 1st Edn., Idea Group Inc. USA., pp: 4. DOI: 10.4018/978-1-59904-929-8.ch029

Al-Fedaghi, S.S., 2007f. Beyond purpose-based privacy access control. Proceedings of the 18th Conference on Australasian Database, (CAD' 7), Australian Computer Society, Inc. Darlinghurst, Australia, pp: 158. ISBN:1-920-68244-9

Al-Saidi, N.M.G. and M.R.M. Said, 2009. A new approach in cryptographic systems using fractal image coding. J. Math. Stat., 5: 183-189. DOI: 10.3844/jmssp.2009.183.189

Anwar, M.M., 2008. An Identity-and Trust-based Computational Model for Privacy. PhD Thesis, Department of Computer Science, University of Saskatchewan, Saskatoon, Canada.

Booz Allen Ideas Festival, 2010. Privacy Engineering Development and Integration.

Davenport, T.H., M. Leibold and S.C. Voelpel, 2007. Strategic Management in the Innovation Economy: Strategic Approaches and Tools for Dynamic Innovation Capabilities. 1st Edn., John Wiley and Sons, Erlangen, ISBN: 10: 3895786039, pp: 441.

Earp, J.B., A.I. Anton and O. Jarvinen, 2002. A Social, Technical and Legal Framework for Privacy Management and Policies. 1st Edn., Americas Conference Information Systems, America, pp: 612.

EU, 1995. Directive 95/46/EC of the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. European Union.

Feigenbaum, J., M.J. Freedman, T. Sander and A. Shotack, 2002. Privacy engineering for digital rights management systems. Security Privacy Digital Rights Manage., 2320: 153-163. DOI: 10.1007/3-540-47870-1_6

Greene, T., 2009. Trust the Cloud? Americans Say No Way. PCWorld Communications, Inc.

Hong, J.I., J.D. Ng, S. Lederer and J.A. Landay, 2004. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. Proceeding of the 5th Conference of Designing Interactive Systems, Processes, Practices, Methods and Techniques, Aug. 1-4, ACM, New York, USA., pp: 91-100. DOI: 10.1145/1013115.1013129

Kang, J., 1998. Information privacy in cyberspace transactions. Stanford Law Rev., 50: 1193-1294. DOI: 10.2307/1229286

Kangassalo, H., 1999. Information Modelling and Knowledge Bases. 1st Edn., IOS Press, Amsterdam, Washington, ISBN: 9051994532, pp: 351.

Marsh, S., I. Brown and F. Khaki, 2008. Privacy engineering whitestudy: A report from a special interest group of the cyber security KTN. University of Oxford.

Masrom, M., Z. Ismail, R.N. Anuar, R. Hussein and N. Mohamed, 2011. Analyzing accuracy and accessibility in information and communication technology ethical scenario context. Am. J. Econ. Bus. Admin., 3: 370-376. DOI: 10.3844/ajebasp.2011.370.376

Mutschler, B., J. Bumiller and M. Reichert, 2006. Why Process-orientation is scarce: An empirical study of process-oriented information systems in the automotive industry. Proc. EDOC, pp: 440. KEG, Tsinghua.

Narayanan, A. and V. Shmatikov, 2010. Myths and fallacies of ‘personally identifiable information. Commun. ACM., 53: 24-26. DOI: 10.1145/1743546.1743558

Nardal, S. and A. Sahin, 2011. Ethical issues in e-commerce on the basis of online retailing. J. Soc. Sci., 7: 190-198. DOI: 10.3844/jssp.2011.190.198

Organisation for Economic Co-operation and Development, 2002. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. 1st Edn., OECD Publishing, Paris, pp: 62. DOI: 10.1787/9789264196391-en

PAB, 2005. New survey reports an increase in ID theft and Decrease in consumer confidence. Privacy and American Business.

Ramezani, S. and S. Hamidi, 2010. Privacy and social interaction in traditional towns to contemporary urban design in Iran. Am. J. Eng. Applied Sci., 3: 501-508. DOI: 10.3844/ajeassp.2010.501.508

Richardson, E.L. and C.W. Weinberger, 1973. Records, Computers and the Rights of Citizens. Advisory Committee on Automated Personal Data Systems. MIT Press. ISBN: 0262080702, 9780262080705

Richardson, E.L., 1972. Advisory Committee on Automated Personal Data Systems. U.S. Department of Health, Education and Welfare.

Roberts, J., 2005. Poll: Privacy Rights under Attack. CBS Interactive Inc.

Sato, K., S. Izumi, Y. Kato and K. Takahashi, 2009. A privacy-based personal and group information modeling in semantic web. Proceeding of the 13th IASTED International Conference on Internet and Multimedia Systems and Applications, Honolulu, Hawaii, Aug. 17-19, Honolulu, Hawaii, pp: 18-25.

Spiekermann S. and L.F. Cranor, 2009. Engineering privacy. IEEE Trans. Software Eng., 35: 67-82. DOI: 10.1109/TSE.2008.88

Tsai, J.Y., S. Egelman, L. Cranor and A. Acquisti, 2011. The effect of online privacy information on purchasing behavior: An experimental study. Inform. Syst. Res., 22: 254-268. DOI: 10.1287/isre.1090.0260

USFTC, 2000. Privacy Online: Fair Information Practices in the Electronic Marketplace. 1st Edn., U.S. FTC., USA.

Zailani, M.S. and A.G. Norjihan, 2009. Controlling and Disclosing Your Personal Information. WSEAS Trans. Inform. Sci. Applied, 6: 397-406.