Personal Information Flow Model for P3P

Sabah S. Al-Fedaghi
Computer Engineering Department
Kuwait University
PO Box 5969 Safat 13060 Kuwait
sabah@eng.kuniv.edu.kw

Introduction

We claim that P3P lends itself to a stronger theoretical foundation based on a personal information flow model (PIFM). The model specifies 16 acts on personal information (PI) that can used to replace intended purpose. The paper investigates the notion of PI and provides fine-grained sub-processes of PI handling, and an opportunity to tailor privacy policies according to the requirements of each sub-process.

As a background for the problem, the extended version of this paper reviews Thibadeau's critique of P3P, especially the discussion related to "more specificity in declaring the purpose behind taking information" (Thibadeau, 2000). "We could have hundreds of very specific purposes. For people who know about the science of human intentionality, it makes sense to be able to list many specific purposes" (Thibadeau, 2000, 2004). Also, the review includes the limitations of P3P discussed in Hogben (2002, 2003) and in Schunter et al. (2002), who argued that the current P3P specification lacks clear guidelines for policy interpretation.

We will first introduce our definition of personal information, then our personal information flow model.

What is Personal Information?

Personal information (PI) is any linguistic expression that has referent(s) of type person (Al-Fedaghi, (2005)). Assuming that p is a sentence such that X is the set of its *referents*, then there are two types of PI:

(1) p is atomic personal information if it is an assertion that has a single human referent (e.g., *John is 25 years old*). "Referent," here, implies an identifiable (natural) person.

(2) p is compound personal information if it is an expression that has more than one human referent (e.g., *John loves Mary*).

The relationship between individuals and their own atomic PI is called *proprietorship*. If p is a piece of atomic PI of a person (i.e., the referent) then he/she is its *proprietor*.

Personal Information Flow Model (PIFM)

The personal information flow model (PIFM) divides functionality into four modules or phases: creating, collecting, processing, and disclosing PI, as shown in Figure 1 (Al-Fedaghi, 2006a, b). New PI is created at Points 1, 2, and 3 by proprietors or non-proprietors (e.g., medical diagnostics by physicians), or is deduced by someone (e.g., data mining that generates new information from existing information). The created information is used either at Point 4 (e.g., use: decision making), Point 5 (stored), or Point 6, where it is immediately disclosed. Processing the PI phase involves acting (e.g., anonymization, data mining, summarizing, translating) on the PI. The disclosure phase involves releasing the PI to insiders or outsiders. The "disposal" or disappearance of PI can happen anywhere in the model, such as in the transformation to an anonymous form in the processing phase. "Store" in Figure 1 denotes both storing and retrieving operations.

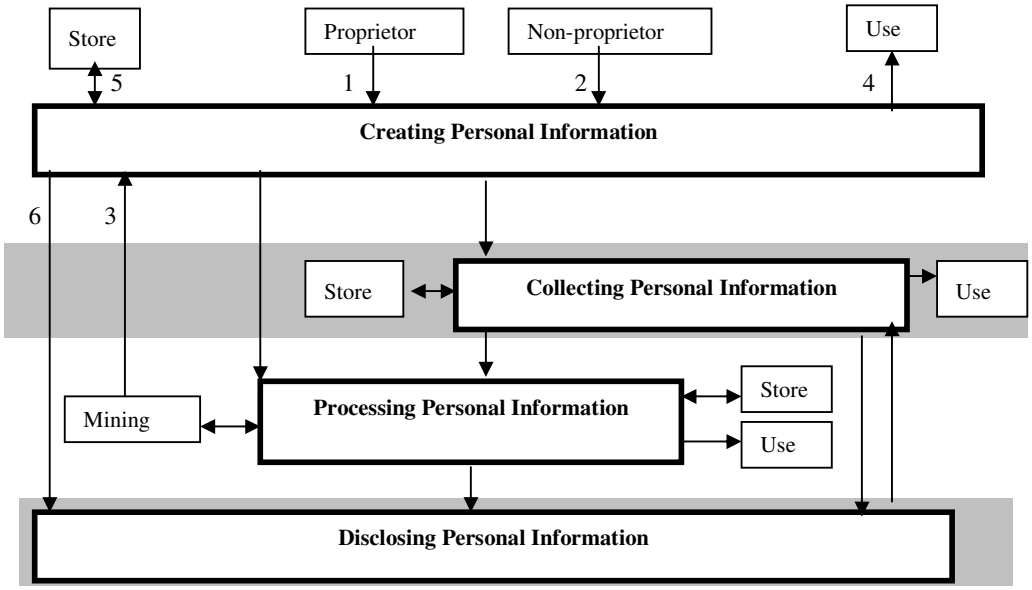

Figure 1: Personal information flow model.

Example: Consider a situation where we have one proprietor and two PI agents (e.g., companies, departments, agencies, other individuals). Suppose that the roles of these three actors are defined as follows:

Proprietor: *Creates*, *stores*, and *discloses* PI to Agent 1.

Agent 1: *Collects*, *stores*, *uses*, and *discloses* PI to Agent 2.

Agent 2: *Collects* and *processes* PI through a *mining* technique that *creates* new PI that is *stored* and *used* in some applications (e.g., decision making).

The PIFM for this simple environment can be drawn based on Figure 1, as shown in Figure 2. Irrelevant details have been omitted.

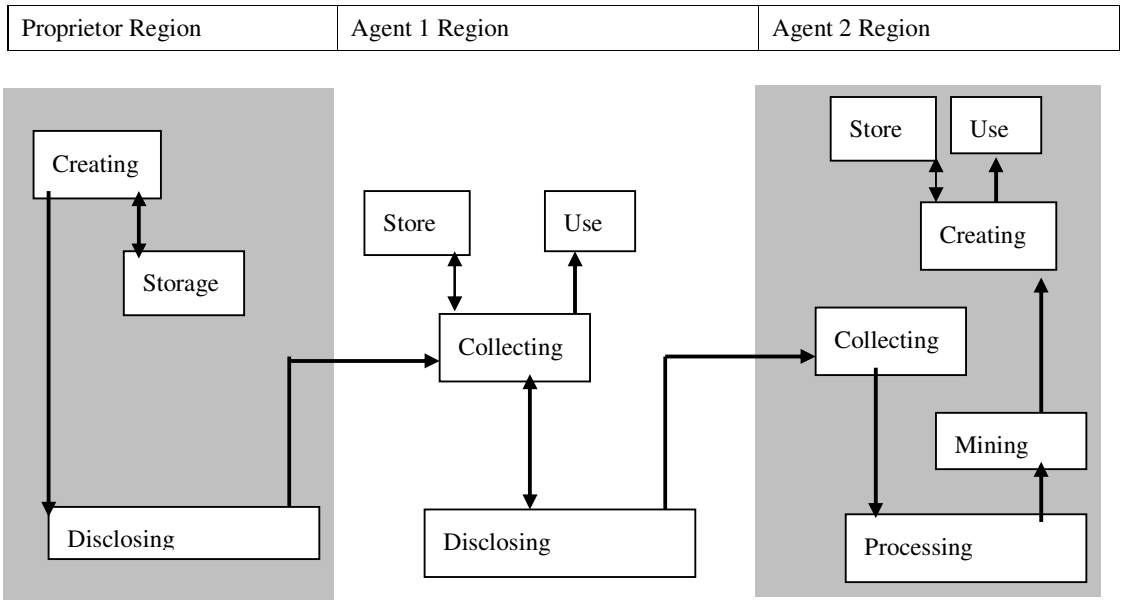

Figure 2: A proprietor and two agents model.

Using the PIFM, we can build a system that involves a proprietor on one side and those who perform different types of activities (other persons, agencies, companies, etc.) in the PI transformations among the four phases of flow of PI on the other side. We will refer to these as PI agents. PI agents may include anyone who participates in activities over PI. The proprietor is not accepted as an agent with respect to his/her own PI. Notice that this type of system is basically a "binary" system that involves the proprietor on one side and all agents, represented in the EU Privacy Directive by the "controller", on the other.

As a result, we need two *types* of PIFMs: one for the proprietors and one for the agents. We construct this proprietor/agent PI flow architecture with two regions: the proprietor's region of activities on his/her PI and the others' region of activities on the proprietor's PI, as shown in Figure 3.

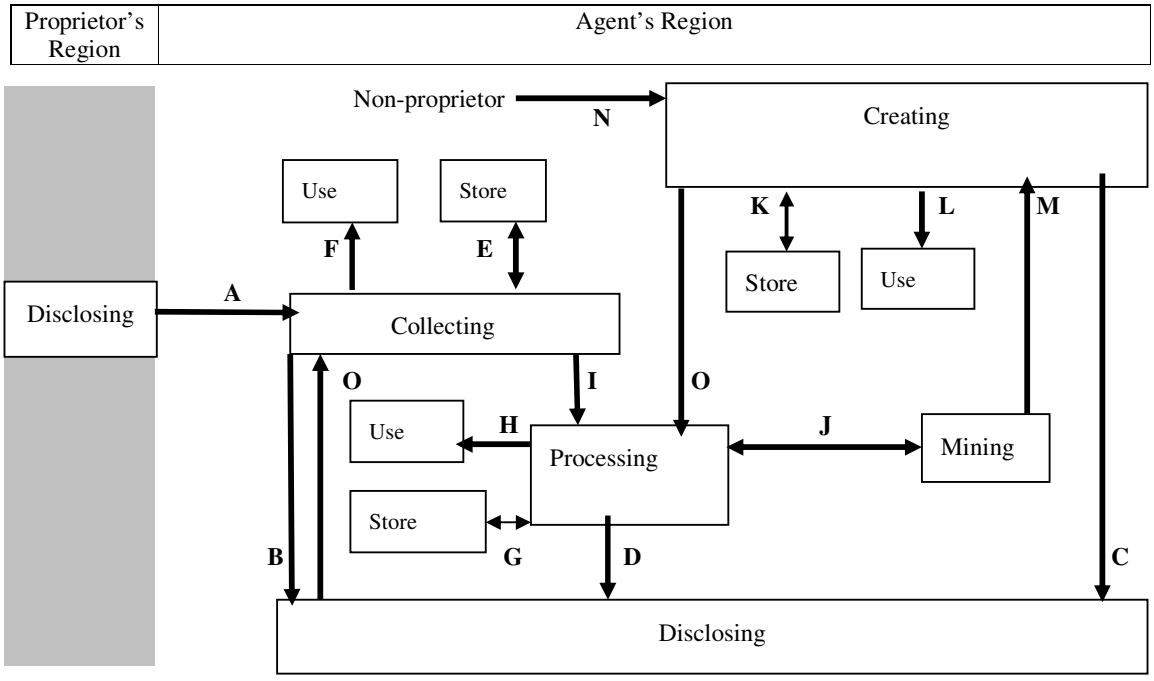

Figure 3: Architecture of Proprietor/Agent PI flow

We distinguish 16 types of acts on PI (labelled A through O), as shown in Figure 3 and described in Table 1. These acts form ordered sequences or chains, as will be discussed later.

Table 1: 16 Generic Acts on Personal Information

Act	Description of the Act	Act	Description of the Act
A	Disclosing PI by a proprietor	I	Processing PI by an agent
B	Disclosing PI by a collecting agent	J	Mining PI by a mining agent
C	Disclosing PI by a creating agent	K	Storing PI by a creating agent
D	Disclosing PI by a processing agent	L	Using PI by a creating agent
E	Storing PI by a collecting agent	M	Creating PI by a mining agent
F	Using PI by a collecting agent	N	Creating PI by a non-proprietor
G	Storing PI by a processing agent	O	Collecting PI from a non-proprietor
H	Using PI by a processing agent	P	Processing created PI

Suppose that a company has a piece of PI. This piece of information is either collected from its proprietor, from another agent, or created internally by the agent. Starting with any of these sources, that piece of PI flows into the PI information handling system (manual or automatic) and is subjected to different acts such as processing, utilization, mining, and so forth. This path of acts can be traced through *chains*.

Application to P3P

The basic concept in this paper is to incorporate chains in P3P to represent privacy constraints on PI. Such a method is presented as a solution to many problems encountered in the specification of privacy policies. Our objective is not to introduce a modification of P3P, rather, it is to provide justifications for adopting PIFM as a foundation for conceptualizing PI handling. No specific changes in the syntax or technicalities of P3P will be given at this stage; however, our analysis shows that the notions of *personal information* and *purpose* need to be drastically refined to implant chains in the specification of policies.

To compare the chains approach with the proposed W3C Platform for Privacy Preferences Draft (2006), we used the following scenario, used in the W3C Working Draft (2006):

> Claudia has decided to check out a store called CatalogExample, located at http://www.catalog.example.com/. Let us assume that CatalogExample has placed P3P policies on all their pages, and that Claudia is using a Web browser with P3P built in. Claudia types the address for CatalogExample into her Web browser…

Assuming that the credit card has been issued by a different company, the phrase "her data will be used only for completing the current transaction, her order" means on the face the chain $A_1(F_1E_1)B_1O_2I_2G_2D_2O_1$ where the subscript 1 refers to CatalogExample and 2 refers to the crediting company. The parenthesis in the chain denotes the “don't care” sequence. The chain represents a well-understood series of acts, as shown in figure 4. The dotted lines in the figure indicate the flow of PI in this scenario. In English, this chain expresses the following:

Your personal information will be stored (E) and used for delivery (F) by us and disclosed to your credit company (BO) which solely will process (IG)) it to check your credit and returns OK/not OK for us (DO).

The chain method is an explicit specification of this instance of acting on PI (supplement by retention period, etc.), while ""her data will be used only for completing the current transaction," is ambiguous specification." “Completing the current transaction" can mean many things. For the CatalogExample, it covers different chains in the CatalogExample's region of actions; it also covers the credit card company's region of acts. The PIFM represents a "closed system" that excludes what is not specified. Thus, the specified chains are the permitted acts on PI, while the rest of the chains are not permitted.

We propose to define the *intended purpose* of PI as a chain of acts on this type of information. Clearly, the 16 acts in the PIFM affect the purpose element in the P3P statement the most. P3P data categories are more affected by the new definition of PI.

Purpose element in P3P contains one or more pre-defined values, such as current, admin, individual-analysis, and historical, that can be qualified with values such as opt-in, opt-out, and always. Interestingly, P3P does not distinguish acts on PI from *using* PI. Acts on PI have already been discussed in terms of the 16 basic acts in the PIFM. For example, the “historical” is an act on PI denoted by *store* in the PIFM. However, such P3P purposes as “telemarketing” and "contact" are *uses* of PI and not acts on PI where the *object* (patient) of acting is not information.

With respect to user agents, the chains represent an unambiguous interpretation of the intended acts on PI specified in the privacy policy. An objection to the chains method is related to user interface. Expressing the acts on PI requires the user to follow different acts on his/her PI. This problem is worth investigating; however, we notice that the method provides the user with a full general description of what is being performed on his/her PI.

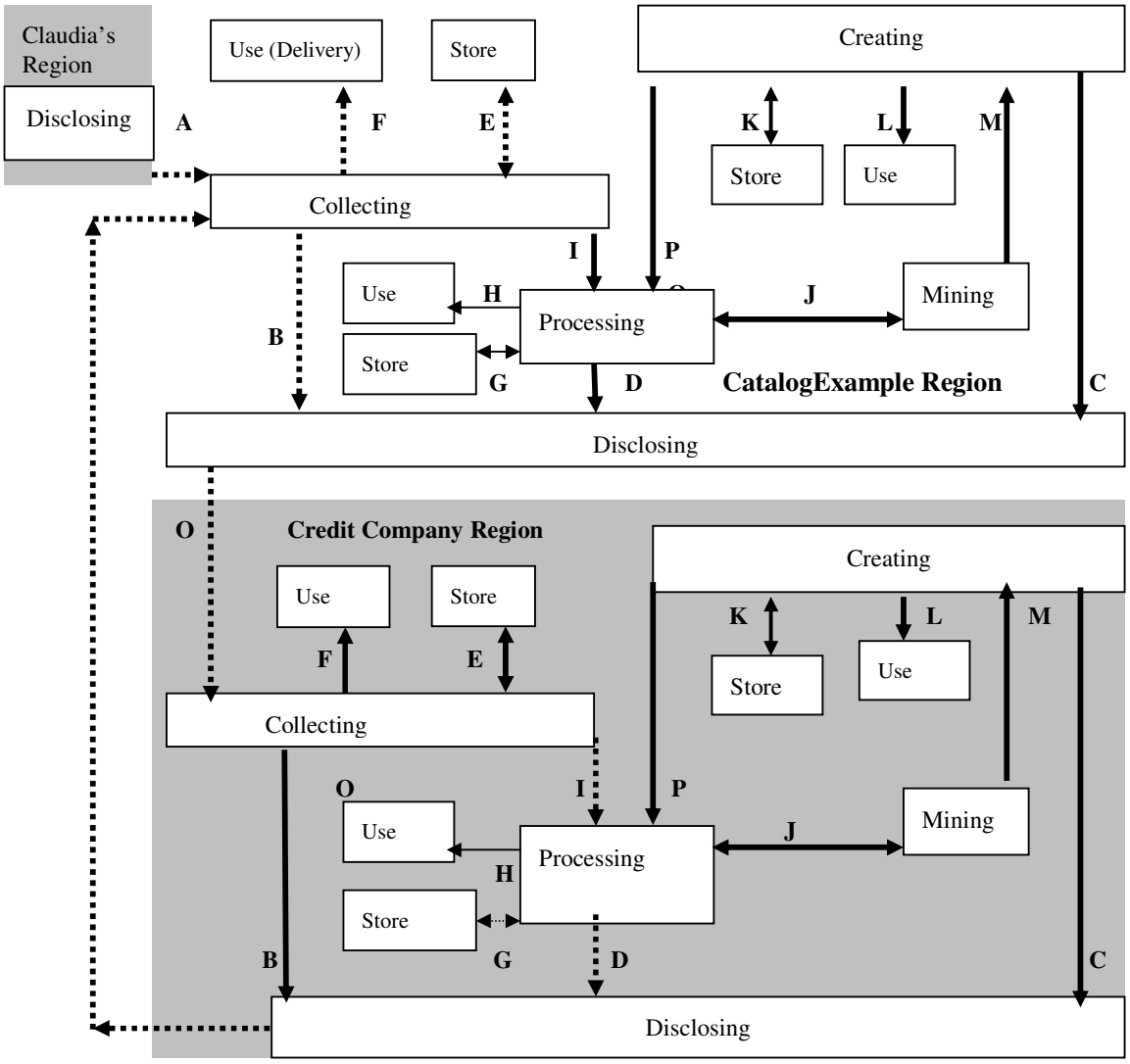

Figure 4. The chain $A_1(F_1E_1)B_1O_2I_2G_2D_2O_1$

References

Al-Fedaghi, S. (2006a). Anatomy of Personal Information Processing: Application to the EU Privacy Directive, International Conference on Business, Law and Technology (IBLT 2006), Copenhagen, December 5-7.

Al-Fedaghi, S. (2006b). Aspects of Personal Information Theory, 7th, The Seventh Annual IEEE Information Assurance Workshop (IEEE-IAW), West Point, NY: United States Military Academy, June 20-23.

Al-Fedaghi, S. (2005). How to Calculate the Information Privacy, The Third Annual Conference on Privacy, Security and Trust, St. Andrews, New Brunswick, Canada.

Hogben, G. (2003, June 18-20). Suggestions for Long term changes to P3P, "Long Term Future of P3P" workshop, Kiel, Germany. http://www.w3.org/2003/p3p-ws/pp/jrc.pdf

Hogben, G. (2002, November 12-13). A technical analysis of problems with P3P v1.0 and possible solutions. "Future of P3P" workshop, Dulles, Virginia, USA. http://www.w3.org/2002/p3p-ws/pp/jrc.html

Schunter, M., Van Herreweghen, E., & Michael Waidner M. (n.d.) Expressive Privacy Promises — How to Improve the Platform for Privacy Preferences (P3P), W3C Workshop on the Future of P3P. http://www.w3.org/2002/p3p-ws/pp/ibm-zuerich.pdf

Thibadeau, R. (2000, August 23). A Critique of P3P: Privacy on the Web (Postscript, April 20, 2004). http://dollar.ecom.cmu.edu/p3pcritique/#postscript

W3C Working Draft 10,. (2006, February). The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. http://www.w3.org/TR/P3P11/