The art of information destruction

Johan J. Lukkien
Department of Mathematics and Computer Science
Security and Embedded Networked Systems
Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

***Abstract*—Increasing numbers of facts and figures of a private nature become accessable to authorities, businesses and sometimes to any interested party on the internet. Embedded sensing, long-term information storage, data mining, profiling and advanced information access techologies are behind this trend. To date, most information processing and gathering systems are indifferent with regard to what happens with the information they process. The main focus is on protecting this information against intrusion, implicitly assuming that access control mechanisms are sufficient for information protection. I think that systems should be built that are much more aware of the ownership of information and that implement information protection and destruction methods that mimic physical limitations of information processing that existed before the digital era.**

I. Introduction

A few weeks ago I obtained a new passport. In the Netherlands this means I had to give fingerprints of four fingers along with a photograph and a signature. I asked what would happen to the fingerprints. Nobody knew. Some thought they would be deleted after they had been stored in some form on the passport; some thought they would be stored in a central database. It is the latter, of course. The next question is what is done, can be done or will be done with those data. This can be found out by digging in the regulations: it is only admitted to search in this database for persons that are already known to the department of justice. A reverse lookup to identify fingerprints is not allowed.

In principle, such an explicit policy is good. However, I have no way of finding out if the policy is enforced. Also, since the law can change afterwards I cannot be sure about what may happen to my fingerprints in the future. For example, at some stage the United States might require a copy. The same doubts hold, of course, for my photograph and signature as these can be (mis)used much more easily in their digital form than before, when they were stored in some physical file in the town hall.

Most people do not really bother about what happens to data obtained from security camera's, personal health records, government agencies and the like, mainly because they see little apparent downsides and do not comprehend the potential. However, there are more and more reports of negative effects of careless handling of private data, for example, data on social websites. It also turns out to be difficult to delete such data, which means that it is difficult for a person to cover up or forget mistakes he made in the past. In addition, the internet has some painful records about the effect when databases get polluted with wrong information, e.g. credit information. A striking example is the story of man who received a mobile phone number that formerly belonged to a registered criminal. On several occasions he was searched and treated as a criminal, and he had difficulty travelling.

II. Privacy

The privacy of a person is the level of control he has about his personal information. With respect to this, the negative elements in the examples above can be summarized in the following three points.

- The lack of control a person has about his information;
- the lack of insight a person has in what happens with his information;
- the poor system support in putting the information owner as the first class citizen.

I summarize these three points in a positive way as the need for having *transparent control* on ones information.

The typical approach to support privacy is not to limit gathering and distribution of data but to define access policies and to (try to) protect the data against unauthorized access. For obtaining transparent control, however, we need to design our systems differently. I suggest a number of ideas in this direction.

Access accounting

One step in the right direction is to build systems such that access to private data is recorded. This does not increase control but it does increase transparency and traceability. An improvement on this is to let the information owner know this access and a further improvement is

to let the information owner grant this access explicitly. In my opinion these are essential features of the new electronic patient file (EPD). In the same way of thinking, I want to be able to examine where the shopping data from my AH customer card has been used for or who accessed it.

Limit information production

The predominant way of dealing with information processing is to gather data at a central location and then process it. A good example is the way the public transportation chipcard system works in Holland. All transactions of a day are put together and processed overnight. Rather than this centralized data collection we must design the system such that the *purpose of data gathering becomes part of the system operation*. We see this idea coming back in policies where a queury is brought to the data to answer it rather than sending data to answer the query. The challenge in the chipcard example is to perform the needed checks and balances without actually having the complete record of precise personalized transactions. The system must further be built such that obtaining such information is impossible.

Privacy interfaces

With the above definition of privacy, privacy protection is much alike a form of DRM. In DRM, the goal is to have the information owner define handling policies, and to have certified procedures and systems to enforce those policies. If we want to have transparent control there must be support for this at all *interfaces* across which information is exchanged. At these interfaces, privacy properties must be specified to which the information receiver will adhere to. A privacy property is a mapping from an information receiver R and a data item d to a data handling property P and can be regarded as a contract of the form: "R will only do P with d". Examples of such properties are

- R will only display d on a screen;
- R stores d for at most a period p;
- R will (not) forward d;
- R will forward d only to a set X of receivers;
- the operations R does with d (and that cascade from R) are traceable.

It requires careful consideration where to place system boundaries. For example, do these properties go all the way into the physical architecture to the level of memories and processors or do they play at the level of communicating stations? Should we maintain a notion of privacy domains? Also the way feedback is given to users must be carefully thought through. Many information-generating devices do not have a display. Finally, enforcing that such properties are maintained even upon malicious behavior of device owners is certainly a challenge.

Privacy levels

Thinking further in this direction we may define a *privacy index* which is function of the above privacy properties. Devices can then be certified to behave according to a certain privacy level. For example, security camera's may be certified to not give out any pictures, only event triggers. Or information from the camera cannot be stored for more than a day. Similar properties can be defined for wearable sensors.

III. Conclusion

Summarizing, current means to deal with personal information are not sustainable. The pressure to use information that exists will increase while time goes by. Consider, for example, the scenario in which at the end of a car trip the embedded printer of the car prints tickets for traffic and speed violations. The challenge therefore is to build systems that cannot generate unneeded information and explicitly destroy information. In addition, ownership over and control of personal data should be much better supported, and actually be part of the system design. The challenge for the coming years is to conceptualize the notion of privacy and privacy properties and find means to put these concepts systematically into the system design. It means that systems must be much more aware of the data they process. Auditing and certification procedures must be developed to verify and certify that a device or system behaves according to the privacy properties. But most importantly, stakeholders must start to think differently about information.