PIDS: A Privacy Intrusion Detection System

Hein S Venter[1], Martin S Olivier[2], Jan HP Eloff[3]

Information and Computer Security Architectures Group, University of Pretoria, South Africa
[1]hventer@cs.up.ac.za, [2]molivier@cs.up.ac.za, [3]eloff@cs.up.ac.za

Abstract

It is well-known that the primary threat against misuse of private data about individuals is present within the organisation. This paper proposes a system that uses Intrusion Detection System (IDS) technologies to help safeguard such private information. It is assumed that the private information is stored in a central networked repository (using, for example, network-attached storage). The proposed Privacy IDS (PIDS) is used on the border between this repository and the rest of the organisation to identify attempts to misuse such information. The system works by identifying anomalous behaviour and reacts by throttling access to the data and/or issuing reports.

Keywords

Privacy, Intrusion detection

1. Introduction

Personal privacy is often defined in terms of control: one has privacy to the extent that one exerts control over one's personal information. While such a definition of privacy is far from perfect, it highlights a fundamental issue of privacy. This issue is that, if control over use of private information is not actively enforced and monitored, a loss of privacy will occur. Over many years privacy-enhancing technologies (PETs) have been developed that help individuals to retain such control. Examples include Crowds (Reiter et al, 1999), LPWA (Gabber et al, 1999), P3P (Reagle et al, 1999) and PrivGuard (Lategan et al, 2002). Even encryption, when used in a privacy context, is about ensuring that only the intended recipient has access to the information — another example of controlling, who gets access to that shared information.

In more recent years, a substantial amount of work has been done to protect personal information after it has been collected by an organisation. This includes work done on the legal, policy and technical areas. Examples of laws that restrict use of personal information and require protection of such information include the (well-known) US Privacy Act of 1974 and the EU Data Protection Directive (Pfleeger, 2003). It is also obvious that privacy policies on websites and in other contexts have become common. A few years ago such policies existed in a limited number of places. This paper is specifically interested in work done in the technical area. Examples of technologies proposed or developed in this area include Hippocratic databases (Agrawal et al, 2002), E-P3P (Karjoth et al, 2003; Ashley et al, 2003) and work on decision-making in this context (Olivier, 2003a). Clearly, after one's information has been collected by an organisation, personal control becomes much harder to enforce. Some of the proposed solutions in this context do store individual preferences that are taken into account before such information is accessed. In a more general sense they limit access to personal information in the organisational context. This is indeed necessary given the fact that

some of the major known breaches of privacy in the past occurred when individuals who had access to personal data, misused their privileges to obtain access to that information (GAO, 1993; GAO, 1997). Access rules might however not be sufficient. In the case of more traditional access control, it is common knowledge that intrusion detection systems (IDS) augment such systems in a natural manner. These systems are not only able to thwart attacks that might otherwise have breached the normal access control system, but also lead to an improved understanding of the strategies used by attackers, as well as improved knowledge about the frequency and severity of actual attacks launched against an organisation. This raises the question: Is it possible to use an intrusion detection system (IDS) that is specifically tailored to detect attacks against a collection of personal information stored by an organisation, and then react to such attacks? This paper addresses this question by proposing a Privacy IDS (PIDS) that does exactly that.

The application of IDS techniques to enhance privacy offers interesting challenges and opportunities. A challenge is that it is extremely difficult to distinguish between legitimate access to private information and access by someone who, under slightly different circumstances, should have been allowed access but is actually 'just browsing' when access is made. It is therefore necessary to cater for a very high number of false positives and false negatives. We contend that it is possible to react in a manner that makes the impact of a false positive or negative tolerable, but still improves the privacy of stored data. An opportunity that arises is that, given the specific domain of application, it becomes possible to take more direct steps to deal with (possible) attacks that are in progress.

This paper is structured as follows. Section 2 contains further information about information privacy and privacy-enhancing technologies. Section 3 examines the role that a current IDS could play in a privacy system. Section 4 proposes a framework for a privacy IDS, and section 5 concludes the paper.

2. Privacy

Perhaps the most widely accepted principles for parties processing personal information are the OECD guidelines (www.oecd.org). The guidelines consist of the following principles: Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation and Accountability. A lack of space precludes a detailed discussion of all these principles here. We will therefore only consider those that are directly related to the specific area this paper considers; the names of the remaining guidelines have been listed for the sake of completeness and to give the context for those that are discussed.

The three principles that apply directly to the problem at hand are the Use Limitation, Security Safeguards and the Accountability principles. The first of these principles is central to the problem at hand: Only when a valid reason exists and the intended use is compatible with the reasons for collecting the data in the first place, should the action to use the information be allowed to proceed. The only exceptions are when the law allows such use as intended, or where the subject has given permission for such use. IDS technology promises to detect the use of information outside these boundaries.

The second principle singled out above, requires that Security Safeguards be in place to protect private data. Given the knowledge that a specific domain implies about the data to be protected, an IDS can use such knowledge to be better able to identify possible intrusions. It also has a greater range of available actions to take in the case of possible intrusions that it detects. The work in Section 4 below will explore this possibility in more depth.

The final principle was that of accountability. In an organisation this will typically imply that the organisation is accountable for misuse of private information by any of its employees. This clearly places an obligation on the organisation to identify possible misuses of such information by employees. While sophisticated logging of access requests can solve some aspects of this problem, logging does not normally provide all the functionality that an IDS does. Moreover, it has been reported that logging leaves loopholes, for example when information is retrieved via a system other than that providing the logging information (GAO, 1997). In addition, current approaches to logging are implemented on a too high level and do not give details on the specific access to information contents, but rather to the containers, such as databases, in which the information is held. Logging is also done in a static manner, whereas IDSs are able to detect intrusions dynamically and in real time.

Since the PIDS to be proposed in this paper is a form of privacy-enhancing technology, it will be useful to consider other forms of privacy-enhancing technologies that the IDS can interact with or complement. The Layered Privacy Architecture (LaPA) (Olivier, 2003b) provides a useful framework for the discussion of privacy-enhancing technologies. LaPA classifies PETs in four layers, viz the personal control layer (PCL), the organisational safeguards layer (OSL), the private (confidential) communication layer (CCL) and the identity management layer (IML).

The PCL includes technologies such as P3P (Reagle et al, 1999) that allows individuals to express their personal preferences about how data should be processed. The CCL uses encryption, such as PGP (Garfinkel, 1995), to ensure that communicated information remains confidential. The IML allows individuals, where applicable, to remain anonymous or to use a pseudonym. Examples of technologies in this category include LPWA (Gabber et al, 1999) and Onion-routing (Goldschlag, 1999). For the purposes of this paper, the OSL is the most important layer. Examples of technologies that can be used on this layer have been given above. As a technology used by the organisation to protect information collected by the organisation, the proposed PIDS will itself form part of this layer.

All existing state-of-the-art IDS implementations attempt to identify intrusions in network and host domains originating from the outside world. Although privacy intrusion can originate from the outside world, the main threat of compromising privacy of a system originates from the inside of organisations. In other words, existing approaches to IDS mainly view this technology as a perimeter defence, similar to firewalls. PIDS operating on the OSL layer in contrast aims to propose a technology that can be employed to detect privacy-compromising behaviour, linked to networked attached storage from internal and external sources. We assume that private data is best stored in a central repository within the organisation connected to the rest of the organisation via a network. This enables one to carefully monitor access requests to the repository. Often such central repositories are implemented as network-attached storage (NAS) units.

3. IDS functionality applied to privacy

Sundaram (1996) classifies intrusions into six main categories, viz attempted break-ins, masquerade attacks, penetration of the security control system, leakage, denial of service and malicious use. Often atypical behaviour is used to detect a specific type of intrusion. In the normal security context, attempted break-ins, for example, are detected when the behaviour of a subject differs from the typical behaviour profile, or if a subject violates specified security constraints. These aspects of an IDS can clearly apply to an IDS in the privacy context with only minor modifications: atypical access of private information may indicate misuse of such information. Similarly, constraints that apply specifically to the use of private data can be specified and violations of such constraints could indicate misuse of private data.

Masquerading, a common vulnerability identified by currently available IDS technology, is also a concern in the privacy context. It is likely to occur when (a) a user attempts to assume a role that the user is not authorised to assume, or (b) when a user attempts to act in a work context that is not expected. While Role Based Access Control (RBAC) and workflow security mechanisms should ensure that this does not happen, IDS technology can identify attempts to bypass these mechanisms and react if these systems are indeed compromised.

Leakage of private information in the privacy context can be associated with identity theft. Trojan Horses have, for example, been used to determine the identifying attributes (such as user identifiers and passwords) of individuals and so gained access to their personal information. Once this information has been lost, the normal mechanisms used to protect access become ineffective. IDS technology could help to identify attempts to acquire such personal information.

4. Model for PIDS

From the previous sections, it is clear that there is potential in combining IDS functionality with privacy to form a PIDS. In order to demonstrate how this is feasible, a model for a PIDS is introduced in this section. Before this is shown, it is however necessary to consider the architecture of an IDS. It should be noted that there are two main types of IDS: anomaly detection and misuse detection (Bace, 2000). The IDS architecture for these two types of IDS is essentially the same, except that where an anomaly-based IDS architecture has an anomaly detector and an anomaly profile, a misuse-based IDS has a pattern matcher and policy rules. A misuse-based IDS, thus differs from an anomaly-based IDS in that, instead of looking for anomalies, it attempts to match a specific pattern from the audit data using the policy rules. These patterns are known in advance and hence specified by the policy rules. For the purpose of the PIDS the architecture of the anomaly-based IDS is adopted.

The main components for an anomaly-detection IDS include a **data source**, a **profile engine**, an **anomaly detector**, a **profile database** and a **report generator**. The way in which an anomaly-detection IDS performs intrusion detection is as follows. Each piece of source data is carefully grouped by the profile engine to form sets of related user or system behaviour. Such a set of behaviour is referred to as a profile. A profile database contains profiles of normal user or system behaviour. The profile database can be set up manually by a human expert to define profiles. Another option is to use a computer in compiling profiles by using statistical techniques, which can be updated automatically. The anomaly detector then compares each

profile compiled from the source data by the profile engine to the normal user and system behaviour profiles from the profile database. When the anomaly detector finds a profile that appears to be abnormal or unusual compared to a specific user or system profile in the profile database, the behaviour is labelled as intrusive and a report or alarm is generated. A model for a PIDS is shown in figure 1. This model is based on a model defined by Denning (1986) and the anomaly-detection IDS architecture as discussed above. The scope of the PIDS in figure 1 is depicted by the dark black line. The data request first travels trough a number of perimeter security and privacy technologies before it arrives at the PIDS for further evaluation. This is only possible if it passed successfully through all the perimeter security and privacy technologies. The components of the PIDS map directly to the original anomaly-based IDS architecture except for a new component called the **privacy enforcer**. This component ensures that the request is either routed successfully to retrieve data from the database or not depending on whether a privacy intrusion was detected. The privacy enforcer then sends a reply whether the request was granted or rejected.

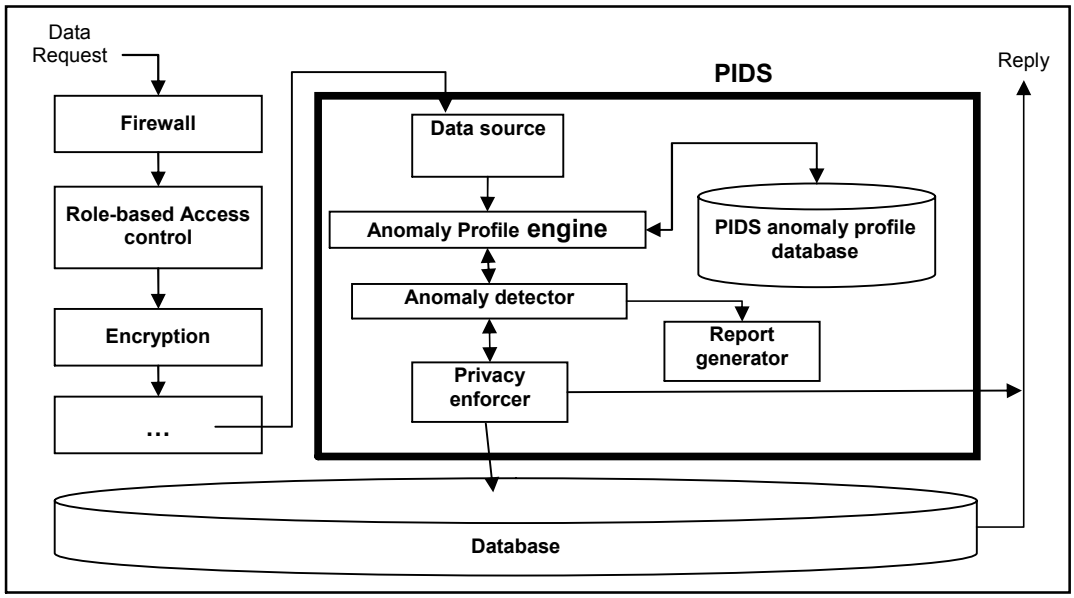

Figure 1: Model for a privacy IDS

Since it is clear how the PIDS components map to the original IDS components, the components of the PIDS can now be discussed in more detail. In order to understand these components, consider the following case scenario. Suppose a representative of a government's revenue service is querying the revenue database. As he has legitimate access to the database, he can retrieve only information he has been granted access to. He might though be able to use this information for unethical purposes. Suppose he retrieves the information of persons over the age of 65 years who are millionaires. A safe assumption is that the revenue service has a privacy policy that does not allow an employee of the revenue service to disclose personal information of any taxpayer to third parties. Suppose, though that the representative retrieved the personal and contact details of the selected group of people and supplied the list to his wife who happens to be a property marketing agent for an exclusive retirement village. The request that the representative made may have been allowed

based on his access rights and role but, an invasion of people's privacy has occurred! It is this type of privacy intrusion that a PIDS will attempt to detect.

The case scenario described above will be examined by the PIDS as follows. First, consider the data request "Get contact details of people with age larger than 65 and financial income of more than ZAR1 million" as input to the PIDS. This request will have to pass through the normal perimeter security and privacy technologies. These technologies can range from low-level perimeter security such as a firewall and simple access control to higher-level security such as a rule-based access control system, a secure workflow system, or any other PET. In the PIDS component the data requests of the user is carefully examined by the anomaly profile engine and compared to the PIDS anomaly profile database by the anomaly detector in a bid to find a profile that appears to be abnormal or unusual. In this particular case, an interest in people who are over 65 might be normal, if they, for example, qualify for a specific tax rebate. Similarly, an interest in the contact details of tax payers and, in some cases, even an interest in those who are relatively wealthy are likely to be normal queries. However, the combined interest is what should be noted as anomalous. If one was to classify all queries that have not previously and regularly been executed by those associated with the current profile as possible intrusions, one would end up with an extremely high rate of false positives. This problem can be solved in two ways though. Firstly, the normal range of parameters for features of a profile is often easy to determine. Normal working hours are, for example, well known and activities outside these hours could indicate suspicious behaviour. Such a determined range on its own does not solve the problem. If an employee decides to work a few minutes late it should not trigger alarms. If employees are not allowed to work outside normal working hours, this would indeed be simple to address using normal access control measures. It is therefore necessary for this to be combined with the second part of the solution — throttling. We use the term *throttling* to refer to the dynamic adaptation of the parameters used for the profile and/or the level of service that the system provides to the user. If the number of `unusual' queries, perhaps involving certain sensitive fields such as contact details, forms one of the monitored features, the threshold for reporting or halting such activity could be lowered with each such query entered during a given period. Alternatively or simultaneously, the speed at which records are returned could be lowered with each record retrieved. This would not prevent our example of the tax inspector from obtaining any contact details at all, but could limit the number of such records he can get and could also lead to his activities being reported to management if he persists.

Examples of facets that could be monitored and throttles are shown in table 1. The **association** refers to a specific PIDS anomaly profile "feature", meaning that "association" is one of a few thresholds that are checked when detecting privacy anomalies. Examples of other such features in a privacy anomaly profile are shown in table 1. To illustrate the concept, fictitious threshold values have been allocated to each feature in the table. In addition, the scope of what is monitored as well as what action to take when a specific value is found to be outside of the valid threshold range is shown in table 1. All such values will differ extensively from one organisational environment to another.

It is important to realise that each request made to the database is linked to a specific person or system referred to as a subject and so a PIDS anomaly profile will have to exist for each subject, but would be derived from the subject's role. The specific features identified for this example include the time of day, duration, number of records accessed, number of records

edited, association of records and frequency of usage. There are 10 entries shown in table 1 for subject Bob; more entries could, however, exist for other subjects. The "time of day" feature for Bob specifies that Bob is normally supposed to access the database between 08:00 and 17:00; Alice, however, might be a night worker and would normally access the database between 18:00 and 02:00. In order to prevent false positives from occurring, Bob may request special permission to work late when he anticipates the need to do so due to. Likewise, the features are set up with specific valid threshold values for each feature for each different subject. It is also possible to take some appropriate action in order to throttle the normal behaviour of the subject accordingly when a specific feature is breached.

Entry	**Subject**	**Feature**	**Valid threshold range**	**Storage component**	**Action(s) taken on threshold violation**
1	Bob	Time of day	08:00-17:00 (override: 08:00-20:00)	Database	1) Report 2) Close database connection
2	Bob	Duration	0-10 minutes	Database	1) Report 2) Close database connection 3) Throttle threshold range
3	Bob	Duration	0-3 minutes	Records	1) Report 2) Throttle threshold range
4	Bob	# records accessed	1-10 records	Records	1) Report 2) Reduce threshold range with 2
5	Bob	# records accessed	11-100 records	Records	1) Report 2) Reduce threshold range with 20
6	Bob	# records accessed	>100 records	Records	1) Report 2) Close database connection
7	Bob	# records edited	0	Specific Record	1) Report 2) Close database connection for the remainder of the day
8	Bob	Association of records	0-2 records associated	Records	1) Report 2) Close database connection
9	Bob	Usage frequency	0-10 times per day	Database	1) Report 2) Close database connection for the remainder of the day
10	Bob	Usage frequency	0-3 times per day	Specific Record	1) Report 2) Disallow access to specific record for remainder of the day
11	Alice	Time of day	18:00-02:00	Database	1) Report 2) Throttle threshold range
…	…	…	…	…	…

Table 1: Features in a PIDS anomaly profile database

5. Comparison with other work

The idea of using an IDS approached to protect privacy is not new. In a Hippocratic Database (Agrawal et al, 2002) a Query Intrusion Detector (QID) is proposed, but few details are given. PIDS differs from the QID in three significant respects: PIDS considers queries while QID considers the results of queries (before data is released). Secondly, PIDS uses an intrusion detection model based on the expected activities of a user. This is derived from the role of the user, as well as individual traits. In contrast QID builds a profile from past queries. Thirdly QID apparently only flags suspect queries, whiles PIDS attempts to limit damage by using throttling.

6. Conclusion

This paper described the concept of a PIDS. It was shown how it could be used to augment normal security and privacy-enhancing technologies to better safeguard private data. It should be emphasised that no such system could ever be perfect. It only takes one person to disclose one piece of information that that individual was perfectly authorised to see to violate privacy. Clearly this would not be noticed by any automated means. We believe that any system that helps to eliminate some violations of privacy that might otherwise have gone unnoticed is a worthwhile effort.

Future research on this system will focus on the construction of a prototype that will allow experimentation, particularly to determine the effectiveness of our application of throttling.

7. References

Agrawal, R., Kiernan, J., Srikant, S., and Xu, Y. (2002), "Hippocratic databases", In *the Int'l Conf. on Very Large Databases (VLDB)*, Hong Kong.

Ashley, P., Hada, S., Karjoth, G., and Schunter, M. (2003), "E-P3P privacy policies and privacy authorization", In *Proceedings of the ACM workshop on Privacy in the Electronic Society*, pp103–109, ACM Press.

Bace, R. G. (2000), *Intrusion Detection*, Macmillan Technical Publishing, USA.

Denning, D. E. (1986), "An intrusion Detection Model", *Proceedings of the 1986 IEEE Symposium on Security and Privacy*, Oakland, California.

Gabber, E., Gibbons, P. B., Kristol, D. M., Matias, Y., and Mayer, A. (1999), "Consistent, yet anonymous, web access with LPWA". *Communications of the ACM*, Vol. 42, No. 2, pp42–47.

GAO (1993), "National crime information center: Legislation needed to deter misuse of criminal justice information", *Document GAO/T-GGD-93-41*, United Stated General Accounting Office, Washington, USA.

GAO (1997), "IRS systems security: Tax processing operations and data still at risk due to serious weaknesses.", *Document GAO/AIMD-97-49*, United States General Accounting Office, Washington, DC, USA.

Garfinkel, S. (1995), *PGP: Pretty Good Privacy*. O'Reilly.

Goldschlag, D. M. (1999), Reed, M. G., and Syverson, P. F. (1999), "Onion routing", *Communications of the ACM*, Vol. 42, No. 2, pp39–41.

Karjoth, G., Schunter, M., and Waidner, M. (2003), "Platform for Enterprise Privacy Practices: Privacy-enabled management of customer data", In Dingledine, R. and Syverson, P. (Ed.) *Privacy Enhancing Technologies: Second International Workshop, PET 2002, San Francisco, CA, USA, April 14-15, 2002, Revised Papers*, Springer.

Lategan, F. A., and Olivier, M. S. (2002), "PrivGuard: A model to protect private information based on its usage", *South African Computer Journal*, Vol. 29, pp58–68.

Olivier, M. S. (2003a), "Using organisational safeguards to make justifiable decisions when processing personal data", In Eloff, J. H. P., Kotzé, P., Engelbrecht, A. P., and Eloff, M. M. (Ed.), *IT Research in Developing Countries (SAICSIT 2003)*, pp275–284, Sandton, South Africa.

Olivier, M. S. (2003b), "A layered architecture for privacy-enhancing technologies", In Eloff, J. H. P., Venter, H. S., Labuschagne, L., and Eloff, M. M., (Ed.), *Proceedings of the Third Annual Information Security South Africa Conference (ISSA2003)*, pp113–126, Sandton, South Africa.

Pfleeger, C. P., and Pfleeger, S. L., (2003), *Security in Computing*, Third edition, Prentice Hall.

Reagle, J., and Cranor, L. F. (1999), "The platform for privacy preferences", *Communications of the ACM*, Vol. 42, No. 2, pp48–55.

Reiter, M. K., and Rubin, A. D., (1999), "Anonymous web transactions with Crowds", *Communications of the ACM*, Vol. 42, No. 2, pp32–48.

Sundaram, A. (1996), "An Introduction to Intrusion Detection", ACM Crossroads, http://www.acm.org/crossroads/xrds2-4/intrus.html.

HS Venter, MS Olivier and JHP Eloff, "PIDS: A Privacy Intrusion Detection System," in Proceedings of the Fourth International Network Conference, SM Furnell and PS Dowland (eds), Plymouth, UK, 255-262, July 2004

©The authors 2004