View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

# PRIVDAM: Privacy Violation Detection and Monitoring Using Data Mining

Jaijit Bhattacharya
Oracle HP e-Governance, Center of Excellence, Gurgoan
*jaijit.bhattacharya@oracle.com*

Rajanish Dass
Computer and Information Systems Group, Indian Institute of Management, Ahmedabad
*rajanish@iimahd.ernet.in*

Vishal Kapoor
Oracle HP e-Governance, Center of Excellence, Gurgoan
*vkapoor@cse.iitd.ernet.in*

Debamitro Chakraborti
Computer Science and Engineering, Jadavpur University
*debamitro@yahoo.co.in*

S.K.Gupta
Department of Computer Science and Engineering, Indian Institute of Technology, Delhi
*skg@cse.iitd.ernet.in*

## Abstract

*Privacy, its violations and techniques to bypass privacy violation have grabbed the centre-stage of both academia and industry in recent months. Corporations worldwide have become conscious of the implications of privacy violation and its impact on them and to other stakeholders. Moreover, nations across the world are coming out with privacy protecting legislations to prevent data privacy violations. Such legislations however expose organizations to the issues of intentional or unintentional violation of privacy data. A violation by either malicious external hackers or by internal employees can expose the organizations to costly litigations. In this paper, we propose PRIVDAM; a data mining based intelligent architecture of a Privacy Violation Detection and Monitoring system whose purpose is to detect possible privacy violations and to prevent them in the future. Experimental evaluations show that our approach is scalable and robust and that it can detect privacy violations or chances of violations quite accurately.*

## 1 Introduction

The area of privacy enhancement technologies has seen tremendous growth in the last couple of years. This is mainly due to the enactment of privacy legislations and the wide-spread use of the Internet and its inherent weakness in the protection of the privacy of individuals as well as organizations. Till date, most of these technologies have focused on privacy middleware [1] and on privacy policy expression 0 . Moreover, research on the detection of privacy violation and the proactive determination of privacy violation patterns to prevent future privacy violation is sparse, if

not non-existent. Consequently, there is a gaping requirement for a method to automate the detection of privacy violations 0.

The need of a privacy violation prevention mechanism becomes evident whenever organizations deal with Personal Identifiable Information. With the increase in the amounts of personal data being collected, stored and processed in information systems, the threat of violation of individual privacy and consequent commercial damage to large enterprises is on the increase. Moreover, control over personal information has also decreased as individuals are unaware of which systems store their information and what all has been stored. Sensitive data, such as detailed transaction summaries including social security number, shipping and billing addresses, e-mail id and credit card details are being put to risk on a routine basis [19].

The issue of privacy violation detection is tricky since any violation detection has to be done on the log created by the privacy middleware. The creation of this log itself can be a privacy violation as it captures information about individuals that they may not want to be stored. This information might include the individual's surfing history and his data-access patterns, using which, one can build back personally identifiable information about the individual. Previous work in this area includes the minimization of the identity information of a user and the employment of techniques like anonymisation and pseudonymisation of log files. This paper ignores this issue of privacy violation due to maintenance of a log and focuses primarily on using the log for Privacy Violation Detection (PVD) 0.

Privacy violations are defined as events that breach a privacy policy or an agreement between a customer (data subject) and the data collecting entity. Generally, an individual's privacy can be protected in two ways, either by minimizing the amount of personal data stored, or by enforcing privacy policies.

As defined by Tina Hermandsen Krekke 0, a Privacy Violation Detector (PVD) aims at detecting such privacy violations.

In this paper, we introduce the implementation of a Privacy Violation Detection and Monitoring system that we have termed PRIVDAM. PRIVDAM uses a suite of machine learning techniques for automated identification of malicious violations and may be a part of a system that enforces

privacy policies. We use the system described in [1] as the basis on top of which the work in this paper has been done.

The rest of the paper is organized as follows: In Section 2 we present the motivation behind creating such a system. Section 3 describes related work previously done in intrusion detection systems as well as privacy violation detection. In Section 4, a privacy broker from our earlier work is discussed which is used to enforce the privacy policy in the database. Section 5 presents an anomaly based PRIVDAM architecture using two data mining techniques. Section 6 describes the implementation of PRIVDAM for a hospital situation and section 7 presents the results.

Finally we conclude with experimental results and make a few suggestions for future research work.

## 2 Motivation behind PRIVDAM

We have conceptualized PRIVDAM as not just a reactive protection mechanism but also an intelligent, proactive one: it is designed to detect privacy violations and learn from such violations so as to prevent their future occurrences. However, this system does not propose to replace the role of the human analysts but simply attempts to reduce their burden.

The need for such a system arises from the fact that vast amounts of data being processed and collected have the potential to become a privacy nightmare in the near future. A few organizations and corporations have a privacy policy enforcement mechanism in place 00, but no systems exist for monitoring and detecting the violations that might occur. Users come across hundreds of privacy policies while visiting various websites, but very few of these privacy policies are actually enforced using automated systems. Moreover, assuming the presence of automated privacy violations, there may not be many systems that would have preventive measures for stopping violations. We have termed programs that enable automated privacy violations as *Leeches* since they leech out the data. This is especially true for high valued information like credit card information. Recently a major credit card provider's data was compromised and this put an estimated 40 million customers at risk 0. This risk includes exposure of the individual to fraud, identity theft while it makes the enterprise susceptible to costly legal liabilities 0.

There are numerous issues that are adding to the problem of privacy violation: like personal data driven e-businesses which are highly motivated towards exploiting personal information; negligence to security and continual technological glitches and an ever increasing community of malicious hackers who want to gain from the situation 0. New bills are being tabled by law-makers to set standards for companies handling sensitive consumer data 0 but until specialized and effective privacy protection systems are put into place, it will be difficult to implement and monitor compliance to the privacy legislations.

With increased consumer awareness even a single privacy violation can lead to costly lawsuits. Businesses want customers to have trust in their institutions and approximately $15 billion dollars is lost every year by e-businesses due to the intensifying mistrust 0. A cable giant, an airline carrier and a major toy manufacturer were all involved in lawsuits over alleged privacy violations 000, which cost these enterprises millions of dollars in settlement fees and lost revenues and goodwill.

We believe that just as network intrusions occur despite network access controls being put into place 0000 similarly, privacy violations will happen despite the traditional policy based privacy violation prevention mechanisms being put into place.

Moreover, once the information is leaked, misuse cannot be prevented, detected or even rolled back and this is the driving factor behind stopping the access before it causes a privacy violation. In the following subsections, we will discuss about the existing classification of perpetrators and scenarios of privacy violations.

## 2.1 Classification of perpetrators

The notations used for various categories of attackers are 0:

- **External Users**: Nefarious attackers trolling the net trying to hack their way into the network. These users are ignorant of the privacy policies and the internal infrastructure and form the unauthorized class.
- **Regular Employees**: Malicious users who have intimate knowledge and inside information about the privacy policy as well as the network. They employ subterfuge to overcome their lack of security credentials to access the privacy data. They have to employ illegal hacking to access the information.

- **Misfeasors**: Insidious users who have the required security credentials. They are probably the most detrimental as they abuse their power and position to access and use information that they are not supposed to. Misfeasors are also employees like in the above category but the difference lies in the fact that they do not have to resort to any illegal hacking to access sensitive data.

Obviously, preventing privacy violations by *misfeasors* is a considerably difficult, if not impossible task. This paper focuses on privacy violations by *External Users* and *Regular Employees.*

## 2.2 Possible scenarios of privacy violations

Even if one can construct privacy middleware that ensure that all data requests comply with the privacy policies, it will not prevent masqueraders and misfeasors from accessing the data by various techniques 0.

### *2.2.1 Hacking Attempts*

A person gets to know the login/password and/or gets access to the privacy policy that is defined in the system either by hacking it or by social engineering methods and therefore assumes the identity of a genuine user or changes the privacy policy such that the person gets access to privacy constrained data. For example, a hacker can break into a hospital's network and sell the patient records to a medical insurance company, enabling the insurance company to provide medical insurance cover at discriminatory process based on detailed health records of individuals.

### *2.2.2 Treacherous employees*

A regular authorized user makes changes to the privacy policy and accesses information he was not supposed to. An example of such a behavior would be the case taken up by Agarwal et al. 0 of a malicious internal user, Mallory while discussing Hippocratic databases. Mallory is an employee with questionable ethics who can retrieve customer records in off peak hours and can potentially sell them to any rival company. This will cause loss of revenue as well as potential lawsuits by customers.

### *2.2.3 Technical faults*

The system can be infected with malicious codes like a virus or a worm. They can be made to send confidential information to outside systems through unsupervised ports. We term such codes as *Leeches* since they leech out the data. Backdoors or loopholes can exist which are a result of badly

configured security systems. They can be used surreptitiously to bypass the normal logging and auditing mechanisms.

#### *2.2.4 Denial of Service*

A DoS attack, in terms of privacy, would lead to the violation of the privacy principle of information, notification and access rights of the data subjects 0, 0. Data subjects have the right to information, to notification and the right to correction, erasure or blocking of incorrect or illegally stored data. A Denial of Service attack violates privacy by impinging on the right to access one's own information.

A crucial yet unrelated form of privacy violation is breaking the confidentiality agreement between two parties. An organization might agree to a non-disclosure agreement but might not adhere to it and sell the information to a third party for illegitimate gain. Sending unsolicited spam mails might be the motive behind such a crime.

## 3 Related Work

While discussing Hippocratic databases 0, a *Query Intrusion Detector* has been proposed which runs on the query results to spot queries whose access patterns is different from the usual access patterns. However, this is similar to the misuse detection approach in Intrusion Detection Systems and is hence would be unable to detect new attacks whose nature is unknown.

Strawman architecture of a Privacy Violation Detection system has also been discussed in Tina Hermandsen Krekke's thesis 0 . However, the work does not go beyond the strawman architecture and does not include any implementation. Moreover it is not extended to malicious hackers and is somewhat limited to the employees who attempt to gain additional privileges for which they are not authorized, or employees who misuse the privileges given to them, i.e. due to accidental disclosure, insider curiosity and insubordination.

### 3.1 Privacy Violation Detection V/s Intrusion Detection Systems

Intrusion Detection Systems have been extensively studied 0 and they have been classified according to their granularity of data processing, source of audit data, detection methods, response to detected intrusions, security, degree of interoperability, manageability, adaptability and network infrastructure requirements.

Protecting the infrastructure from external unauthorized access is a security issue whereas the protection of individual's information from intentional or unintentional abuse of authorized access is a privacy issue. This thought can be exemplified by the disparity between privacy and security. Privacy pertains to an individual's information whereas security pertains to the enterprise information access and is focused on the enterprise systems.

However, security and privacy are also weakly co-related. In most cases, privacy requires security, but sometimes security functions may hinder or actually be a cause of privacy violations e.g. intrusion detection systems and logging. This is called the Security-Privacy Paradox 0.

PRIVDAM borrows some of the learning from Intrusion Detection Systems 00000. However as privacy is inherently different from security, our architecture has some novel features that support privacy violation detection. They will be discussed in detail in Section 5.

### 3.2 P3P Policy Specification Language

Considerable work has been done in creating standards for specifying a website's privacy policy. The Platform for Privacy Preferences Project (P3P) provides a standard to websites to communicate their data practices 0. It provides the syntax and semantics of privacy policies and the mechanisms for associating policies with Web resources. It includes machine-readable privacy policy syntax that web browsers and other agent tools can use to fetch P3P privacy policy automatically.

The specification includes

- A standard vocabulary to describe a web site's data practices
- A set of base data elements that web sites can refer to on their P3P policy
- A protocol for requesting and transmitting web site privacy policy.

The P3P protocol is a simple extension to the HTTP protocol – it uses XML. However, the privacy specification language does not support implementation of the stated privacy policies. Moreover, it does not allow personalization of privacy policies and merely helps in specifying the generic privacy policy in a machine understandable form. P3P also does not provide enforcement. Furthermore, it does not include mechanisms for transferring data or for securing personal data in transit or storage. Intermediaries such as telecommunication providers, ISPs, proxies and others may be privy to the exchange of data between a site and a user, but their practices may not be governed by the site's policies.

In some cases, P3P vocabulary may not be precise enough to describe a website's privacy practices.

### 3.3 Enterprise Privacy Authorization Language

In order to address the shortcomings of P3P, an enterprise privacy policy authorization language (EPAL) has been proposed by Schunter et al 0.

An EPAL policy is essentially a list of privacy rules that are ordered with descending precedence (i.e., if a rule applies, subsequent rules are ignored). A rule is a statement that includes a ruling, a user category, an action, a data category, and a purpose. A rule may also contain conditions and obligations. Rules are used to determine if a request is allowed or denied. A request contains a user category, an action, a data category, and a purpose.

Legislation and privacy policies may state that when a certain action is performed, the enterprise is obligated to take some additional steps. An example is that all accesses against a certain type of data for a given purpose must be logged. Or children's data shall be deleted within 30 days unless parent consent is obtained. In EPAL such consequential actions are called obligations. EPAL is not designed to encode the logic of an obligation. The system which evaluates a request against an EPAL policy must be capable of executing all obligations given the unique name of the obligation 0.

## 4 Privacy Middleware

The PRIVDAM system described in this paper uses the log file from the privacy middleware shown in Figure 1.

The privacy middleware ensures that all the data stored in the data repository adheres to the privacy policies that are stored in the Privacy database. Data requests are granted only when the EPAL based privacy policies match the characteristics of the request and the requestor.

The architecture of the privacy middleware uses EPAL to store the privacy policy in XML format and the access requests also conform to EPAL. The entire transaction is then logged in the query log. This query log, along with some network data, is then utilized by PRIVDAM for constructing features that are used in the data mining analysis.

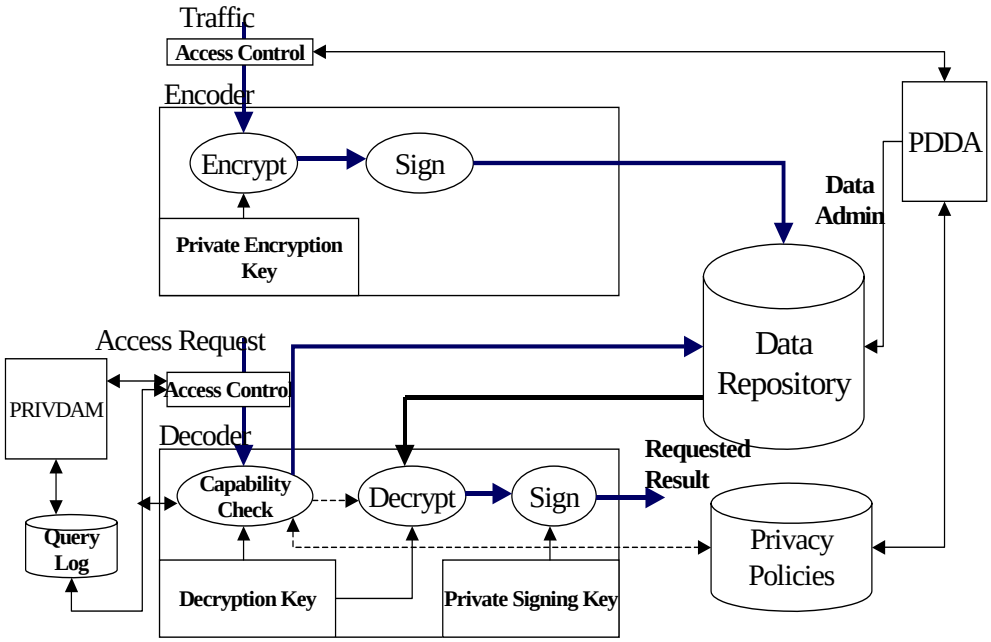


**Figure 1: Architecture of Privacy Middleware**

## 4.1 Logging

The PRIVDAM architecture operates on the privacy middleware log which is created whenever access is granted to some information.

Logging is done via collating data from various sources. The log contains a feature set containing network attributes and privacy attributes.

### *4.1.1 Network Attributes*

Network attributes include Network traffic data. Such data is collected using Netflow tools. Netflow tools capture only packet header information like source IP address and port number, destination IP address and port number, type of packet etc. These traffic flow data is stored in a flat file.

#### *4.1.2 Privacy Attributes*

Privacy attributes are extracted from the data request and the corresponding applicable EPAL privacy policy. Privacy attributes include the action, purpose and data category. Moreover the login ID and number of records accessed are derived from the database access logs.

## 5 Privacy Violation Detection and Monitoring

As previously discussed in section 2.1, privacy violations can occur due to external hackers or from regular employees (we are ignoring misfeasors as PRIVDAM does not address this kind of privacy violation). In addition, section 2.2 describes the mechanisms that may be used to violate privacy. Privacy violations using each of these mechanisms will leave a *signature* or *pattern* in the privacy log file. These *patterns* or *pattern deviations* that arise from such violations are described in the following subsections.

### 5.1 External Hackers (Access through Internet)

This kind of attack involves hackers trying to break into the network through remotely logging on to the enterprise computers. We assume that the hackers are oblivious of the privacy policy and hence cannot change it. They can only masquerade as genuine employees of the organization and the following symptoms arise from such a possibility:

1. The source IP address and possibly the source port deviates from the usual enterprise intranet address. The hacker could certainly spoof her IP address, but that would require her to have prior information regarding the enterprise's network configuration.
2. The access time pattern deviates from the normal access time patterns.
3. The purpose provided by the perpetrator deviates from the normal pattern of purposes provided by the actual authorized personnel.
4. The malicious codes like leeches that infect the system can cause unsupervised flow of data out of the computer. This will cause the number of records requested to go up drastically and possibly thousands of records might be requested at a time. This would deviate from the typical pattern of the number of records accessed in a single request.

### 5.2 Regular Employees (Access through Intranet)

This scenario involves the employees tampering with their privacy policies and hence getting access to previously unauthorized data. This may includes the misfeasors that do not have to change their access rights but simply misuse their authority. The following cases arise:

1. The data category accessed deviates from the data category that the user normally accesses. For example, the log shows the nurse accessing the medical history of a patient whereas her previous access patterns show that she has never accessed medical history in the past.
2. The purpose and action could similarly be modified. This might result in the log showing the testers writing physician's orders for treatment or some such other absurd possibility which does not have a precedence in the log pattern.
3. The time of access might be the same but the number of records requested might go up when an employee with malicious intent wants to read the records of all patients.
4. A combination of the above scenario might also result, e.g. the data category might change along with either action/purpose or time of access. All such combinations that can lead to privacy violation have been considered in our work.

### 5.3 PRIVDAM Methodology

The proposed PRIVDAM system attempts to detect and monitor all the eight symptoms using clustering and data mining algorithms. However DoS attacks and NDA annulments are not considered for simplicity sake. This is largely due to the fact that if the organization itself cannot be trusted, then all auditing mechanisms are rendered irrelevant.

PRIVDAM acts on the privacy log created by a Privacy Violation Prevention system like the one developed by Bhattacharya and Gupta [1]. Privacy Violation Prevention restricts access to privacy records in accordance with the privacy policies. The Privacy Violation Prevention is built into the privacy middleware [1].

PRIVDAM operates on the principle of unsupervised anomaly detection approach using statistical methods. Data accesses are identified that are possibly illegal but were able to pass through the existing privacy violation prevention frameworks, in a manner similar to network intrusion and database hacking, as described in sections 5.1 and 5.2.

Consequently, there is a pressing need to have a PRIVDAM approach that is able to generate access patterns that are *definitely* normal and access patterns that are *definitely* not normal. Such an approach would necessitate the definition of what constitutes *definitely* normal and *definitely* not normal.

PRIVDAM requires the generation of positive and negative access patterns, and since we do not have any existing sources, we need to bootstrap the pattern generation process. If there were one or two existing patterns, then one could have followed an approach of using the existing patterns to filter the transactions that are potentially normal or not normal and then could have performed data mining on them to generate patterns that a human privacy analyst could have investigated. However, unlike in the case of network intrusion detection where intrusion traffic patterns already exist, the PRIVDAM approach has no existing patterns and hence bootstrapping is necessary.

Also, the attempt is not only detecting privacy violations but also to prevent future privacy violation using the *definitely* not normal access patterns. Hence, the system needs to be a Privacy Violation Detection and Monitoring System (PRIVDAM).

The process of detecting privacy violations can be categorized into two parts 0: (a) collection and organization of anomalous data that potentially documents privacy violations and (b) analysis of the anomalous data to look for potential privacy violation patterns.

Anomaly detection requires that a profile of a normal behavior is generated, and that a certain threshold or a statistical deviation from this normal profile is defined such that deviations larger than the threshold are labeled as potentially anomalous and hence possibly a privacy violation.

A policy-based detection approach requires that the log is compared to a defined machine readable privacy policy such as EPAL 0.

The proposed PRIVDAM approach uses clustering of the privacy log data to find outliers that correspond to access to privacy-constrained data. The approach is illustrated in Figure 2.

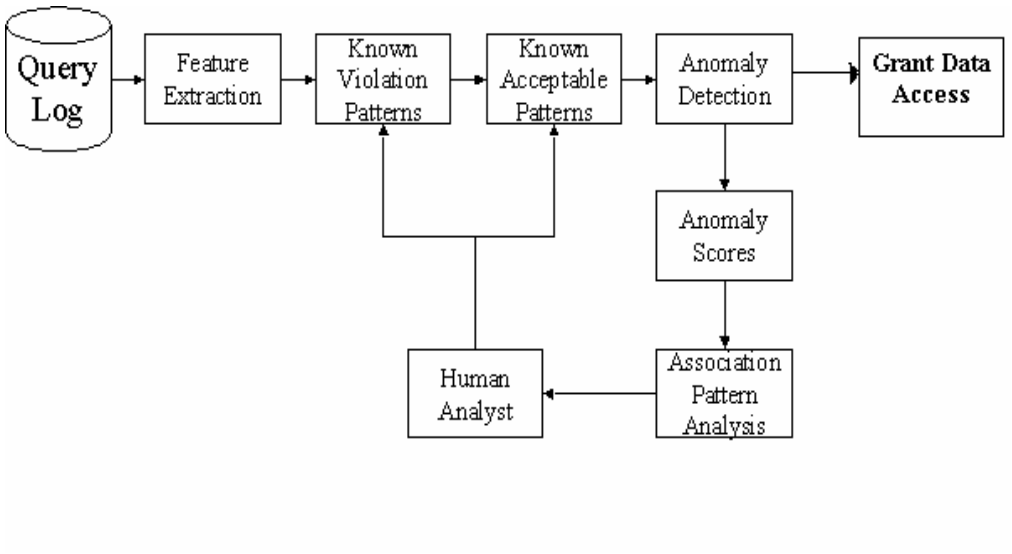


**Figure 2: Privacy Violation Detection and Monitoring (PRIVDAM) approach using Clustering and Data mining**

The input to the system is the Query Log from the Privacy middleware (ref Figure 1). This query log forms the base on which the violations are sought to be detected.

### 5.4 Feature Extraction

The first step in PRIVDAM is extracting the features that are to be used in the data mining analysis. As described in section 4.1, the feature set consists of network features and privacy features. The network features include time of access, source IP address, source port, destination IP address and destination port. The privacy features are extracted from EPAL request to obtain user category, action, purpose, data category, privacy policy id used . These features are especially useful for identifying irregular data access.

Once the feature extraction has been done, the feature set is compared with the known violation patterns to remove the previously known perpetrators. This step reduces the overheads on the system. This feature set is then compared with existing known *definitely* acceptable patterns and the matching sets are removed. Initially, since there are no patterns of either known privacy violations or known acceptable patterns, hence this step has no impact and the system bootstraps itself without any patterns.

### 5.5 Mining Distance based Outliers

An outlier is defined as follows*; provided with user defined parameters p and D, and a distance function F, an object O in a dataset T is said to be an outlier if at least fraction p of the objects in T lie greater than distance D from O. 0*

During the next step the feature set is provided to the PRIVDAM anomaly detection module that uses a clustering based outlier detection algorithm. This module searches for deviation from the normal patterns and assigns an anomaly score to each data access log entry.

PRIVDAM adopted an outlier detection algorithm for anomaly detection since it had provided superlative results in the case of intrusion detection systems 0.

We used the data mining tool Orca 00 for discovering outliers. It uses the distance from a given example to its nearest neighbor to determine its unusualness. Outliers can be viewed as candidates who have a low nearest neighbor density.

### 5.6 Association Pattern Analysis using BDFS(b)

While the improvement of detection rate and reduction of false alarms is an important objective but the task cannot be limited to this if the data collated is very large. It becomes impracticable for analysis of hundreds of violations that might get detected.

For this purpose, we have used BDFS(b) [22], a frequent pattern mining algorithm for finding out the most frequent rules occurring among the parameters of the feature set (described in section 6.1), above the user-defined support and confidence. BDFS(b) is a frequent pattern mining algorithm for association rule mining, based on a novel combination of the staged search and the depth first search [36]. As a result, it has the merits of both best-first search and the depth-first-branch-and-bound (DFBB) search [37], and at the same time, avoids bad features of both. Thus this algorithm introduces a new search strategy, not limited entirely to breadth-first or depth-first search, and explores the given search space in stages. When we can assign specific merit to a pattern depending on the particular context, BDFS(b) has the ability to ensure that patterns of higher merit will be preferred over those of comparatively lesser merits. This ensures that we can come up with interesting meritorious patterns faster, which in turn will help us reacting to them efficiently for better decision making.

Hence, after detection of the outliers, which are possible privacy intrusions, PRIVDAM runs an association rule data mining on the outliers to generate the patterns of potential violations.

*Let T be the set of log entries and A be the set of attributes defined over T. For example A consists of {action, purpose, login id, data category}. Let I be a set of attribute-values pair defined over A. For example I= {action= "read", purpose= "diagnosis", login id=15, data category = "patient history"}. Each attribute-value pair is termed an item. Subsets of I are called itemsets.*
*Association rules are defined between two disjoint itemsets X and Y as:*

$$X \rightarrow Y(c,s)$$

*Where c is the confidence and s is the support value for the rule*

The PRIVDAM association pattern analysis module summarizes data accesses that are ranked highly anomalous in the anomaly detection module. An association rules aims at finding interesting intra-relationship within a single log entry.

The outcome of the association pattern analysis is presented to a human privacy analyst who then decides if the patterns are indeed indicative of privacy violations and hence decides whether these summaries are helpful in creating new rules that may be further used in the known violation detection module. Thus one can start with no known patterns and the process will bootstrap.

Once a set of patterns is known, PRIVDAM uses these patterns for the monitoring of potential privacy violations. When a data access request comes, its features are extracted and compared with known privacy violation patterns. If the patterns match, then the request is blocked else the request is granted. However, as discussed in Sub section 2.2 even if the request is granted, the request may still actually be a privacy violation.

Depending on the privacy sensitivity of the organization and the data, the access may be granted, but in order to detect potential violation pattern in the future, the output of the check against known privacy violation patterns is fed into the known acceptable patterns. Here the organization has to keep in mind the impact of not providing the data e.g. denial of data access to a genuine doctor in a hospital due to a false alarm might cause irremediable harm.

PRIVDAM operates on the data collected on a particular machine rather than the traffic relating to the whole network.

## 6 Implementation

We take the case of a health care information system based on the work of Krekke 0. We extend the case to our PRIVDAM architecture and then present the results in the next section.

The medical information system is especially privacy sensitive due to the fact that patient information collated at health care centers is highly confidential in nature and any leakage of medical information can be highly damaging to an individual. For example, an individual inflicted with AIDS may not want his employer and colleagues to know about his or her medical status as it may compromise the individual's quality of life. Moreover, medical records were subjected to some of the earliest privacy legislations and standards like HIPAA.

Personnel working at the health care center, using the medical information system deployed in the health care center, can access patients' personal identifiable information (PII) for treatment and diagnosis. In our simplified case, we categorize medical workers into the following categories of actors: receptionists, nurses, testers and doctors who all have predefined roles. The privacy policy of the hospital can be implemented using EPAL 0 or similar constructs. The data-categories, actions and purposes in the medical policy are explained using Table 1.

Each column specifies the action, purpose and data category of each user. The purposes are treatment (t), diagnosis (d), localization (l), registration (r) and testing (te), and only two actions, read (r) or write (w) are allowed.
This privacy policy is enforced by the means of a privacy broker [1], but as we have already discussed, it is not sufficient as privacy violations might still crop up. Therefore we need to have a PRIVDAM system to operate on the privacy log.

**Table 1 Authorization matrix**

| Data Category | Doctor | Nurse | Tester | Receptionist |
|---|---|---|---|---|
| Patient id data | | | | r(t),w(t) |
| Contact data | r(d,t), w(d,t) | r(t) | | r(l),w(r) |
| Medical history | r(d,t), w(d,t) | | | |
| Physician's order | r(d,t), w(d,t) | r(t) | | |
| Progress notes | r(d,t), w(d,t) | | | |
| Depart-mental Reports | r(d,t), w(d,t) | | w(te), w(d), r(d) | |
| Nursing data | r(d,t), w(d,t) | w(t), r(t) | | |
| Operative reports | r(d,t), w(d,t) | | | |
| Discharge Summary | r(d,t), w(d,t) | | | |

## 6.1 Feature Set of Privacy Log

The created log, as discussed in section 4.1, contains the following attributes:

- Time: The access time of the user converted in seconds.
- Src_ip: IP address of the source computer that has initiated the request.
- Dest_ip: IP address of the destination computer that has the data stored.
- Src_port: Port number of the source computer.
- Dest_port: Port number of the host where the requests are terminated.
- Login_id: The unique id of the hospital employees.
- Number of records requested from the database.
- Action: This is extracted from the EPAL request and can be either read or write depending on the type of request.
- Purpose: Another feature extracted from the EPAL request and possesses a predefined value as stated above.
- Data category: The category type of data requested by the user.
- Weekday/ Weekend: This aspect reflects whether access is being carried out on weekends.

The first six attributes are continuous in nature and the next five are categorical attributes, which after transformation, are converted into ordinal values. For example, read was assigned a value of 0 and write was given 1 in the action attribute. Similarly, the data category was numbered from 1-9 in the increasing probability of them resulting in a privacy violation.

### 6.2 Simulation of Data

Synthetic network datasets like the Darpa'99 0 and KDDCup'99 data sets are publicly available. Similarly, real life network data is also easily available. However these data sets are known to possess serious limitations 00and moreover real life data containing features required for our PRIVDAM system is not obtainable beforehand as there are no dataset of acceptable quality available for privacy violation. Hence a synthetic dataset was created 0 using the Information Exploration Shootout 0 and privacy characteristics were appended using extensive simulation.

A total of sixteen employees were considered to be working at the hospital. Each user possessed a normal, privacy violation free access pattern and 400, 000 such records were simulated. Statistics for one month of access was simulated, the details of which are listed in Table 2. The action and purpose fields conform to the privacy policy previously defined.

The simulated data is temporally skewed in terms of the number of log entries due to the fact that the day shift employee encounter more patients than the off hour employees. Similarly, the disparity in the number of records requested between the receptionists and the testers can be attributed to the fact that receptionists need to view the information of already registered patients when registering new patients. However, the tester is usually treating only one patient at a given point in time and is not required to access any other data.

### 6.3 Probable Privacy Violations

The various symptoms that our architecture attempts to detect have been discussed thoroughly in Section 5. Anomalous data was simulated for each of the case and the various log entries were appended to the synthetic normal behavior. 10,000 such entries were created and distributed randomly in the normal data set.

Apart from possible violations, some probable false alarms such as doctors accessing patient records from their homes for treatment or diagnosis were simulated. Moreover, the advising doctors might work from off-site and give their opinion from different hospitals.

**Table 2: Distribution of log entries**

| Login-ID | User | Access Time | Records accessed | Data Category | Log entries |
|---|---|---|---|---|---|
| 1 | Receptionist 1 | 0900-1700 | 0-200 | 1,9 | 60,000 |
| 2 | Receptionist 2 | 1700-0100 | 0-200 | 1,9 | 30,000 |
| 3 | Receptionist 2 | 0100-0900 | 0-200 | 1,9 | 30,000 |
| 4 | Tester 1 | 0900-1700 | 0-10 | 7 | 40,000 |
| 5 | Tester 2 | 1700-0100 | 0-10 | 7 | 30,000 |
| 6 | Tester 3 | 0100-0900 | 0-10 | 7 | 30,000 |
| 7 | Nurse 1 | 0900-1700 | 0-100 | 1,2,5 | 40,000 |
| 8 | Nurse 2 | 1700-0100 | 0-100 | 1,2,5 | 30,000 |
| 9 | Nurse 3 | 0100-0900 | 0-100 | 1,2,5 | 30,000 |
| 10 | Treating Doc 1 | 0000-1200 | 0-50 | 1,2,3,4,5,6,7,8 | 30,000 |
| 11 | Treating Doc 2 | 1200-0000 | 0-50 | 1,2,3,4,5,6,7,8 | 20,000 |
| 12 | Advising Doc 1 | 0000-1200 | 0-50 | 1,2,3,4,5,6 | 20,000 |
| 13 | Advising Doc 2 | 1200-0000 | 0-50 | 1,2,3,4,5,6 | 10,000 |
| 14 | Exam. Doc 1 | 0000-1200 | 0-50 | 2,3,5,6,8 | 20,000 |
| 15 | Exam. Doc 1 | 1200-0000 | 0-50 | 2,3,5,6,8 | 10,000 |
| 16 | Surgeon | All Day | 0-10 | 4 | 10,000 |

# 7 Evaluation of Results

This section describes the results obtained by applying the PRIVDAM methodology to the synthetic dataset. The architecture had no data labeled as good or bad data and the system bootstrapped with no existing patterns.

## 7.1 Anomaly Detection Module

The distance based outlier algorithm was run on the given dataset of 410,000 and it resulted in all the log entries being correctly assigned an outlier factor. The dataset had 10,000 records arising out of privacy violations. Each of the columns was normalized using their standard deviation so that the calculation of distances is not skewed in favor of attributes possessing large values. Clustering was done to extract the top n outliers in the dataset where the value of n was varied for different results.

The cases simulated in section 6.2, were found to be assigned the highest scores by the anomaly detection module and consequently detected successfully. Sample attacks are shown in Table 3.
Although all of the attributes are part of the clustering, only those attributes which have significantly variant values between normal and anomalous behavior are shown for clarity.

**Table 3: Outliers coming out of the Outlier Detection module**

| Case | Interesting Attributes | Normal Behavior | Anomalous Behavior | Average Scores |
|---|---|---|---|---|
| 1 | Src_ip , Src_port | 192.168.0.21 8080 | 61.16.215.75 21 | 0.314 |
| 2 | Time , Weekends | 0900 - 1700, 0 / 1 | 0100 – 0200, 1 | 0.493 |
| 3 | Purpose | 4,5 | 1 | 0.218 |
| 4 | Number of records | 0–100 | 6000 - 15000 | 0.526 |
| 5 | Data category | 1/2/5 | 9 | 0.317 |
| 6 | Purpose , Action | 4/5,0/1 | 2,1 | 0.246 |
| 7 | Time , Number of records | 1700–0100, 0 - 10 | 0900 – 1700, 0 - 500 | 0.473 |
| 8 | Data category, Action , Purpose | 1 / 9, 0 / 1, 1 / 2 | 7,1,3 | 0.307 |

Some false alarms were encountered due to larger variation of some normal records as compared to the anomalous ones. Figure 3 shows the comparison of the accuracy of captured attacks v/s the number of normal records, with 10,000 anomalous records and with 8,000 anomalous records.

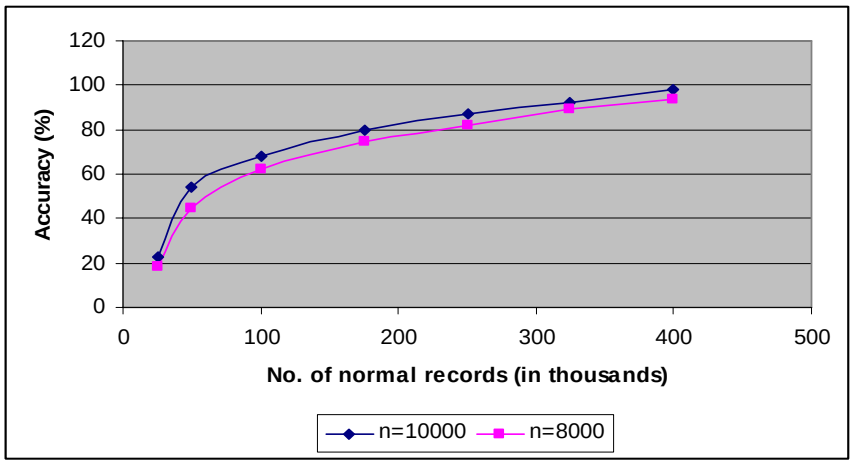


**Figure 3 Comparison Study**

The value of accuracy is calculated as:

*Accuracy=Number of detected Outliers/Total number of Outliers*

The two different plots are drawn by varying the value of ‘n’ during the mining of the top ‘n’ outliers.

We notice that the accuracy improves as a result of increasing the value of ‘n’ but there is tradeoff that exists between accuracy and false alarms, which has to be taken into account.

Figure 4 shows the outliers scatter plot along the login-id and time, with outliers marked.

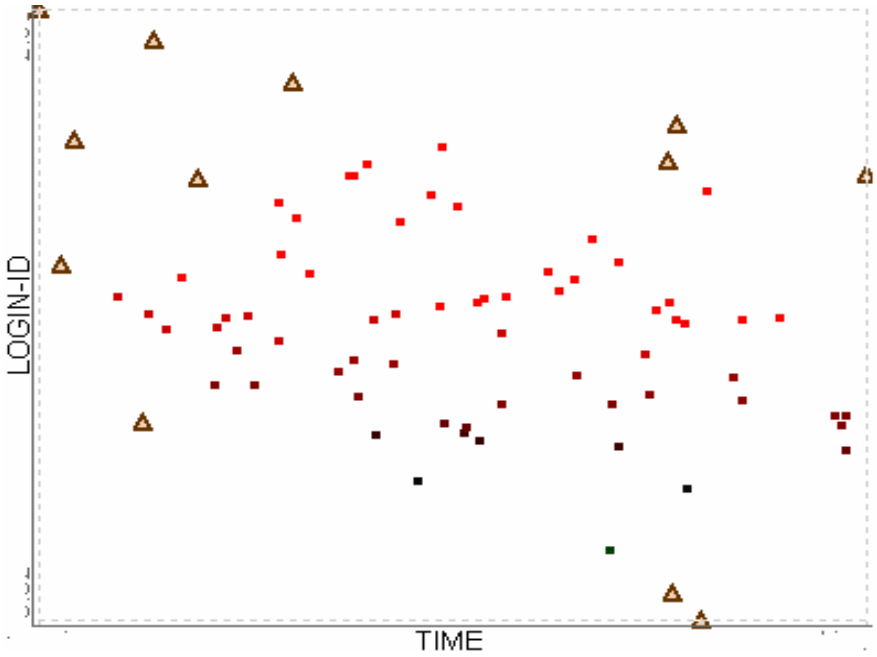


**Figure 4 Visualization of outliers**

## 7.2 Summarization through association rules

In the next step the highest ranked outliers were passed through the associative mining algorithm to find interesting patterns and summarize the results.

The frequent item sets of various lengths were generated along with their support value. In this section we report some of the highest ranked patterns generated by PRIVDAM on the outliers:

- {Src_ip=61.16.215.75, Src_port=21, Number of records = 481, Login_id=6} (s=521)

This pattern indicates an intrusion attempt by a source computer having the specified IP address and port number. It requests a large number of records and has a big support value s as well, which makes the pattern more interesting to the analyst.

- {Data category= “Contact data”, Purpose= “Treatment”, Login_id=13} (s=376)

The pattern above implies an unusual behavior where an advising doctor is accessing the contact data of patients for treatment. The support value is also quite high to warrant a possible violation.

- {Action= “Read”, Number of records = 2304, Weekend, Login_id=3 } (s=457)

The pattern indicates an unusually large data reads on the weekend. The employee is a receptionist and it might be a false alarm with the access being normal.

- {Data category = “Medical history” , Purpose = “Registration” , Login_id = 16} (s=472)

This pattern implies a violation and the access rights of the surgeon must be revoked immediately. The purpose of registration is not included in the privacy policy of the hospital relating to the specified personnel.

Hence PRIVDAM methodology is able to successfully produce meaningful summarizations of possible privacy violations allowing it to be used by a human analyst to partially automate privacy violation detection.

## 8 Conclusion and Future work

In this paper, we have introduced the PRIVDAM methodology that uses intelligent data mining techniques for detecting privacy violations. The approach taken by PRIVDAM allows bootstrapping of the privacy violation detection. Experimental evaluations show that our proposed methodology performed significantly well and at larger ratios of normal to anomalous data, it detected all the anomalous records.

However, the methodology has the limitation of not being able to detect privacy attacks like those of DoS, slow scanning and multiple location attacks. Moreover, unsupervised learning techniques are best suited for unlabeled data having high dimensionality and huge volume. The low percentage of violations among total logged data renders standard data mining techniques to be of lesser use. We propose to employ other algorithms for their efficacy in a future work.

An interesting scenario that PRIVDAM can successfully detect is when the pattern of reads and writes can have anomalous behavior. Say for example, a doctor would normally have a couple of read accesses onto a patient's medical history and then a single write to update the diagnosis and prescription. However, if a masquerader attempts to access the data as a doctor, he could simply make reads and no writes. Since this would lead to a variation in the privacy log pattern, it can potentially be captured by PRIVDAM. Looking at similar cases, a couple of domain-dependant heuristics can be designed as well, that will make the proposed PRIVDAM methodology more accurate and robust for privacy violation detections. The fully-developed PRIVDAM system can go a long way if some good visualizations techniques are also provided as an aid for better comprehension of suspicious behavior captured by our system.

Appreciating the huge amount of automated data traffic being encountered in any organization, a number of steps of efficient querying and aggregation of attributes must be carried out efficiently. We wish to extend in a future work by incorporating these various optimization techniques as well. For better summaries and efficient pruning of insignificant rules, we also propose to try using frequent episode rules [30], by only considering the axis and reference attributes while summarizing the attacks.

A major significance of the work will be felt when we can extend PRIVDAM for real-time (or near-real-time) privacy violation detection. Considering the main proponents of privacy, that information once leaked into the wrong hands it is impossible to be undone, we plan to extend the current PRIVDAM methodology to react in real-time.

We hope that this paper will encourage more work on privacy violation detection and will promote developments of various intelligent machine learning techniques for fast and accurate detection of privacy violations.

## 9 References


[1] Bhattacharya J. ,Gupta, S.K., *Privacy broker for enforcing privacy policies in databases*, KBCS 2004

[2] Barse E. L. , *Logging For Intrusion And Fraud Detection*, Thesis For The Degree Of Doctor Of Philosophy, ISBN 91-7291-484-X, School of Computer Science and Engineering, Chalmers University of Technology, Technical Report no. 28D ISSN 1651-4971

[3] Ertöz L., Eilertson E. ,Lazarevic A. , Dokas P.T. , Kumar V. , and Srivastava J. , *Detection and Summarization of Novel Network Attacks Using Data Mining*, Technical Report, 2003

[4] Westin, A.F., *Privacy and Freedom, Atheneum*, NY, 1967.

[5] Krekke T.H. , *Privacy Violation Detection*, Master's thesis has been carried out at the Norwegian University of Science and Technology (NTNU), 22nd June 2004,

[6] Lazarevic A., Ertoz L., Ozgur A., Kumar V., Srivastava J. , *A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection*, Proceedings of Third SIAM International Conference on Data mining, May, San Francisco, 2003.

[7] Lazarevic A.*, Dokas P. , Ertoz L. , Kumar V. , Srivastava J., Tan P. N., Cyber Threat Analysis – *A Key Enabling Technology For The Objective Force (A Case Study In Network Intrusion Detection)* Army High Performance Computing Research Center, Computer Science Department, University of Minnesota

[8] D. Barbara, N. Wu, S. Jajodia, *Detecting Novel Network Intrusions Using Bayes Estimators*, First SIAM Conference on Data Mining, Chicago, IL, 2001.

[9] W. Lee, S. J. Stolfo, *Data Mining Approaches for Intrusion Detection*, Proceedings of the 1998 USENIX Security Symposium, 1998.

[10] D. Denning, *An Intrusion Detection Model*, IEEE Transactions on Software Engineering, SE-13:222-232, 1987.

[11] Emilie Lundin, Erland Jonsson, *Survey of Intrusion Detection Research*, Technical Report 2004

[12] The Information and Privacy Commissioner / Ontario, Deloitte & Touche The Security-Privacy Paradox: Issues, Misconceptions, and Strategies, August 2003, available at *http://www.ipc.on.ca/user_les/page_attachments/secpriv.pdf*

[13] Emilie Lundin, Hakan Kvarnstorm and Erland Jonsson, *A synthetic fraud data generation methodology*

[14] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, Yirong Xu, *Hippocratic Databases*, Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002

[15] B. Teasley, Does Your Privacy Policy Mean Anything? *http://www.clickz.com/experts/crm/analyze_data/article.php* ,January 11, 2005

[16] Bowman, L. Comcast hit with privacy violation lawsuits, *http://news.zdnet.com/2100-1009_22-923285.html*

[17] Toys "R" Us faces suit for alleged privacy violation Associated Press, August 3, 2000

[18] Fliers File Suit Against Jetblue, Associated Press, Sep 23,2003

[19] Bruno, J. B. Security Breach Could Expose 40M to Fraud, Associated Press, June 18 2005

[20] Information Exploration Shootout , *http://www.cs. uml.edu/shootout/*

[21] Wen Jin. Anthony KH Tung. Jiawei Han. *Mining Top-n Local Outliers in Large Databases,* Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining

[22] R. Dass and A. Mahanti, *"Frequent Pattern Mining in Real-Time – First Results,"* presented at TDM2004/ACM SIGKDD 2004, Seattle, Washington USA, 2004.

[23] J.P.Anderson (1980). *Computer Securiy Threat Monitoring and Surveillance. Technical report*, James P Anderson Co., Fort Washington, Pennsylvania

[24] Federal Trade Commision , *http://www.ftc.gov/ privacy/index.html*

[25] Regan, K. Privacy Times: *Internet privacy, an oxymoron?*.Feb 3,2000

[26] R. P. Lippmann, R. K. Cunningham, D. J. Fried, I. Graph K. R. Kendall, S. W. Webster, M. Zissmal, *Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation*, Proceedings of the Second International Workshop on Recent Advances in Intrusion Detection (RAID99)

[27] J. McHugh, The1998 Lincoln Laboratory *IDS Evaluation (A Critique)*, Proceedings of the Recent Advances in Intrusion Detection, 145-161, Toulouse, France, 2000

[28] IBM Tivoli Privacy Manager for e-business , *http://www-306.ibm.com/software/info/ecatalog/ en_TH/products/K106003J38182X80.html*

[29] AT&T privacy bird, *http://www.privacybird.com/*

[30] Jianxiong Luo, Susan M. Bridges, Rayford B.Vaughn, Jr., *Fuzzy Frequent Episodes for Real-Time Intrusion Detection,*

[31] The World Wide Web Consortium. The Platform for Privacy Preference (P3P}. Available from *http: //www. w3.org/P3P/P3FAQ. html.*

[32] M. Schunter et al, Enterprise Privacy Authorization Language (EPAL 1.1), IBM Research Report, http://www.zurich.ibm.com/security/enterprise-privacy/epal

[33] D. Baumer, J. Brande Earp, F. Cobb Payton, *Privacy of Medical Record*s: IT implications of HIPAA, Computers and Society, December 2000

[34] Orca: A Program for Mining Distance Based Outliers, *http://www.isle.org/~sbay/software/orca/*

[35] Bay, S. D. and Schwabacher, M. (2003). *Mining Distance-Based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule*. Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[36] N. J. Nilson, Artificial Intelligence: A New Synthesis. Los Altos, CA: Morgan Kaufmann, 1998.

[37] V. N. Rao and V. Kumar, "Analysis of Heuristic Search Algorithms," University of Minnessota, Technical Report Csci TR 90-40, 1990.