SHEILA:
Securing Healthcare data by Efficient and Intelligent Log file Analysis

Uppsala University, Uppsala, Sweden
Rose-Hulman Institute of Technology,
Terre Haute, Indiana, USA

itinsociety@gmail.com

January 31, 2011

Abstract

The revolution of information technology has made information easily accessible within the hospitals. Electronic Healthcare Records (EHR) are now the standard documentation material for hospital staff, who are connected to the system on a daily basis. As the systems grow larger, both by registered users and patients, it is now only a question of time before all the hospitals in Sweden are connected to one national EHR database.

This student collaboration report is founded on article investigations, interviews, field investigations and surveys. We answer what ethical and integrity issues that arise within EHR systems, both for users and patients. Then we investigate solutions used today and the technical methods that can be used to reduce the ethical and integrity problems.

As the systems are merging and the patient database grows larger everyday, these are unanswered questions that need investigation before the problems grow out of hand.

Contents

1 Introduction

Electronic Healthcare Records (EHR) have been used by the the Uppsala County Council and other parts of Sweden for several years. Even though this provides benefits such as increased healthcare treatment due to better informed physicians, easier communication between healthcare professionals and larger amounts and more detailed data available for researchers, there are also downsides to their use. In every situation where large amounts of personal data are collected by companies or state agencies, the risk of illegal access, unintended use or loss of data occurs. When it comes to EHRs, a lack of data confidentiality seems to be especially harmful as medical data is considered as absolutely private. Although different measures have been taken to secure EHRs, there have been several cases in Sweden during the last years where patient data has been illegally accessed.

One of the best known intrusions occurred when the Swedish Minister for Foreign Affairs, Anna Lindh, was assassinated. During the time she was treated at hospital and the time afterwards, many intrusions were recorded. One of the nurses who read Anna Lindh's healthcare record was required by the court to pay 30 day-fines [39].

While it seems obvious that celebrities and politicians are especially vulnerable, other cases show that every citizen can become a victim of EHR intrusion. In 2009 for example, a woman in Kiruna unlawfully accessed a relative's healthcare record. She claimed that she felt threatened by the man and wanted to know if he was really dangerous. She accessed his healthcare records up to three times and later on was sentenced to pay 30 day-fines by the district court [29].

Furthermore, there are cases where the delinquents are not even aware that it is wrong to access other peoples' records. In the year 2010 for instance, a male nurse viewed his ex-girlfriend's healthcare records in Gävle. The woman got worried when she saw the nurse in the hallway of the hospital. After contacting the head of the department, a lawsuit against the male nurse was filed. The nurse explained his behavior by saying he was unaware that the action was illegal [25]. In 2009 the Data Inspection Board raised concerns about the issue that some county councils around Sweden did not control the access to the healthcare records. They felt that the problem was widespread throughout the country.

This report investigates the problem of illegal access to EHRs and provides recommendations for the Uppsala County Council to avoid misuse. There will be a special focus on the issues of confidentiality and integrity of patient data that arise with the use of EHRs. Furthermore a description of EHR usage is provided to achieve a better understanding of it and the necessity of logs in the context of medical care in particular and in other contexts. Possible solutions and methods to prevent and detect inappropriate use will be discussed. There will be special attention to methods that use log analysis as a means of prevention and detection. The report will also provide recommendations for the Uppsala University Hospital and the Uppsala County Council. These will be based on the findings of this report concerning ethics, law and technical ways

of prevention and detection.

The report has been created by 1 Colombian, 1 German and 14 Swedish students that took part in the course "IT in Society" at Uppsala University and 4 American students participating in the course "Computing in a Global Society" at Rose-Hulman Institute of Technology, Terre Haute, Indiana, USA in autumn 2010. There have been collaborations between these courses for the last six years. In order to improve the collaboration, the American students and their teacher, Cary Laxer, stayed at Uppsala for one week in September and another week in December. The names of all authors can be found in Appendix D.

The report begins with an introduction to the scenario (Section 3.1) that is used to illustrate the problems and techniques and a short description of the methods used(Section 4). Afterwards, the ethical problems from the patient's and medical staff's perspective are described to get a deeper understanding of the whole set of problems (Section 5). Sections 6 and 7 describe ways of securing Electronic Healthcare Records. Whereas section 6 focuses on techniques to avoid illegal access, section 7 includes measures of spotting illegal access that already took place. B describes investigated solutions of log handling in other contexts than healthcare records. Section 8 serves as a wrap up to summarise the previous sections. Finally, the findings of this report and potential future work are discussed (Section 9) and turned into recommendations for the Uppsala County Council (Section 10). As problem solving always takes place within a certain legal context, the most important laws in this area are described in Appendix A

2 Definitions and Clarifications

2.1 Definition of Electronic Healthcare Records

Many different definitions of Electronic Healthcare Records (EHR) are used in existing literature. According to Wynia and Dunn, an EHR can be "any electronic means of storing and transferring health-related information"[46, p.64]. Iakovidis provides a more concrete description by defining an EHR as "digitally stored health care information about an individual's lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times"[15, p.106]. According to Gartner Consulting, "EHRs are not the legal medical record of any healthcare organization" but "a natural repository of summary information"[31, p.3]. Apart from the term EHRs, several other similar expressions can be found: Personal Health Record [46, p.64], electronic Personal Health Record [46, p.65], Electronic Medical Record [31, p.2] and computer-based patient records [31, p.2].

2.2 Difference Between Paper-Based and Electronic Healthcare Records

When talking about the confidentiality problems arising with EHRs, one should take into consideration that paper-based patient records have existed for a long time. Even though they could not be accessed from many computers at the same time, there has always been the risk of illegal access to patient data. In contrast to digitally stored records, there is almost no way of recognizing or proving that somebody has accessed a paper-based record if the person is not caught in action[21, p.124]. However, the number of people that had access to a paper record is significantly lower compared to EHRs: "unless copied, they were in only one place at one time"[7, p.46]. To assure that EHRs are at least as safe as paper-based ones, the Patient Data Act was enacted in 2008. Its content is summarized in Appendix A.1.

2.3 Stakeholders of Electronic Healthcare Records

Since many people are affected by the use of EHRs, different interests arise. These interests lead to information demands that the stakeholders want to be fulfilled by the EHRs. Unfortunately, these demands are not complementary in many cases. On the one hand, the success of physicians' and nurses' daily work depends on the quality and completeness of their patients' healthcare records which might lead them to wish that all available information be accessible. The staff responsible for data integrity on the other side wants to limit the accessibility of private data to avoid misuse. Furthermore, patients rely on the confidentiality of their data but also have high expectations on the quality of their healthcare, which of course requires shared information about the patient with many involved professionals [12, p.151]. Additional stakeholders are identified by Gaunt: "Health insurers, administrators, government departments and

professional regulatory bodies are demanding greater detail about care delivery and clinical performance, putting in jeopardy the privacy of the patient and clinician alike. Human rights groups are campaigning for personal privacy and the right of the individual to have access to the data recorded about them" [12, p.151 f.]

The remainder of the report is limited to the interests of the following stakeholders: patients, medical staff and researchers.

3 Problem Description

3.1 Scenario

In order to aid understanding and to concretize the research questions for this report, a fictional but realistic scenario is presented as follows:

Maria Eriksson is a 53-year-old woman living in Uppsala, Sweden. She is a popular local politician, making her a high profile person whose medical details could be of interest to a large number of people. Maria also has several friends within the medical care sector in Uppsala. This further increases the risk for illegal access of her Electronic Healthcare Record (EHR).

One day Maria is hit by a car while she is out riding her bicycle. A person on the street calls the emergency number. An ambulance arrives to help her and the ambulance crew immediately makes a quick review of her condition. After performing a quick immobilisation they transport her to the Uppsala University Hospital. On their way to the hospital the ambulance crew alerts the staff at the emergency department by phone, telling them Maria's personal number and their assessment of her injuries. They also report that they sensed a smell of alcohol from her breath and that a blood sample should be taken. Arriving at the emergency room, the ambulance crew gives a quick verbal report to the staff present in the room. Then the trauma team takes over and the ambulance crew leaves to finish their reports.

A surgeon examines Maria and checks the injuries reported by the ambulance crew, and concludes that there are no signs of internal bleeding but a possible brain trauma and a suspected broken leg. An emergency CT-scan shows no signs of cerebral bleeding or other serious injuries and an X-ray confirms an uncomplicated fracture in her leg. Maria is then transferred to the observational ward where her condition is determined to be a light concussion. Later in the cafeteria the nurse from the emergency room tells some of her colleagues that the politician came in earlier that day. One of the colleagues finds this very interesting. She opens Maria's record even though she has no legal right to do so. She has no problem at all getting access to Marias healthcare record since all medical employees at the hospital have access to the EHRs.

An orthopaedic surgeon is consulted regarding Maria's fracture. Upon reading her record and looking at the X-ray images the surgeon repositions the bone and places a cast on the leg for it to heal. Maria is then asked to stay overnight for observation. The orthopaedic surgeon adds his conclusions and plans for follow-up treatment to Maria's EHR. In the morning she is discharged, and the physician on the ward creates a summary of her hospitalization in her EHR. After three weeks she visits the hospital again to renew the cast. The doctor in charge reads her record. She says she would like new medication for the pain. The doctor prescribes some new painkillers which she can buy from any drugstore in town. Later she meets a physiotherapist who shows her new gymnastic exercises which she shall follow during the recovery.

Before the accident Maria had just shared a bottle of wine with her friends. The media get hold of the story and very private details from her record such

as a heightened level of alcohol in her blood during the accident. It becomes a big scandal because she is a popular politician and her reputation is seriously damaged. She gets very upset with this and files a charge. The police investigate the logs and find out that the only person who has accessed her record without a legal reason is the nurse in the cafeteria. Later the case is closed and no charges are filed since the police could not prove that she sold the story to the media.

After three months Maria makes a new visit, this time to a primary care unit to remove the cast. The medical staff opens her EHR and makes a note in it. Maria is now regarded as fully recovered and the medical case is closed.

The only reason that the illegal access was discovered was that the nurse sold the story to the newspapers. As the system is constructed now there is no way of detecting illegal accesses unless the information leaks out. In case of an illegal access to a health care record, the risk of getting caught is very small since the system administrators only examine a hundred randomly selected log files each quarter of a year.

3.2 Other Research Questions

Besides the questions that arise from the scenario, this report will also try answering the following questions:

1. Does the medical staff know what ethical rules apply when they're using EHRs?
2. How should ethical training of the medical staff be conducted?
3. What detection techniques are best suited to detect illegal access of an EHR?
4. What detection methods are used in similar domains and how do they work?

4 Methods

In September 2010, the international team gathered in Sweden to establish the foundations of communication and organization for the remainder of the collaboration. During the week, the team divided into the following groups to consider various sub-elements of the project:

Prevention and Detection: This group was responsible for finding different ways to prevent and detect attacks on the patient database. The group did not formally divide, but some group members focused primarily on methods of preventing illicit access, while others focused on technical algorithms to detect illegal access after it occurred.

Laws and Ethical Issues: This group was responsible for researching laws and investigating other ethical issues associated with access of medical records. The group primarily was concerned with understanding the objectives of research over the course of the project. Furthermore, it investigated the moral ramifications in Sweden's political environment as well as the global society in which the proposed system would operate.

Other Systems and Solutions: This group was responsible for researching how other systems handle security issues pertaining to personal records. These systems encompassed implementations found not only in the medical industry, but in law enforcement agencies as well. This research was meant to provide supplementary solutions to the ones provided as well as introduce topics of interest to the other groups.

Information was primarily gathered through web-based research, discovering scholarly articles on related subjects. In addition, a series of interviews were conducted with healthcare professionals in Sweden in order to gain a better understanding of the current state of affairs and the major concerns of medical staff. The team also visited Stockholm's läns sjukvårdsområde (SLSO), the Healthcare provision of Stockholm County and the Swedish National Police Board to learn about their methods of log analysis. Finally, the team attended several lectures from experts in the areas of ethics and patient concerns.

To streamline the information gathered over the course of the collaboration, the Swedes and the Americans communicated via Skype to discuss several topics. Overall, there were three types of collaborative meetings. First, class time was allocated to reinforce and update the overall structure of the report. Additionally, the in-class meetings served to update everybody on the progress of the project as well. Secondly, each team was responsible for setting up its own meetings to plan out how work was going to be distributed among the members of the group. These smaller meetings served to guide individual research. Lastly, the team leaders also met weekly on Skype to discuss current progress and objectives for the coming week.

5 Ethical Problems with Electronic Healthcare Records

All employees having access to personal data as Electronic Healthcare Records (EHR) should get a basic training in ethics and the laws in their countries. But is that training enough when they use EHRs and manipulate information in an IT system? Do they have ethical and moral dilemmas with the EHR? What kind of experiences do they have with EHRs?

In order to answer the questions above, seven interviews were conducted with two nurses, two doctors, a Ph.D. student and two researchers. None of their names will be disclosed here, since some of the people who were interviewed requested this. All of them work at Swedish hospitals such as Karolinska, Uppsala Akademiska, Sahlgrenska Universitetssjukhuset and Östersunds Sjukhus and have access to patient information. Even though some of these hospitals use different systems than COSMIC (the IT system used in the Uppsala County Council), it can be assumed that the general procedures of handling patient data are similar. Furthermore, the opportunity arose to get insights in different ethical trainings, guidelines and handling of IT systems by interviewing medical staff from different hospitals. The results of these interviews are mainly presented in this chapter and the conclusions section. This section provides a description of the most important ethical issues in the context of EHRs from the patients' perspective (Subsection 5.1) and the medical staff perspective (Subsection 5.2).

5.1 Ethical Problems from the Patients' Perspective

Privacy and confidentiality are important topics for human beings in all areas of daily life. However, people tend to be even more concerned when it comes to medical information. Gostin states that maybe "the most intimate and sensitive form of personal information is an individual's health care records"[13, p.683]. The reason why health care records are highly sensitive is that they are directly connected to single persons, their health status and the treatment they received [17, p.129]. The research on this topic has shown, that "sexual problems and mental health" [32, p.2378] are the topics where patients hesitate the most to disclose information. Interestingly, the reasons for non-disclosure are quite different: "a belief that the problem was not worthy of discussing with a physician, [..] certain characteristics of the physician, such as age or sex, or [..] the patient's sense that the physician is not interested" [32, p.2378 f.].

Once information is disclosed, it is important for patients to know who has accessed their records. A survey conducted in the United States in 2005 found that "81 percent want to be able to personally review who has had access to their personal health information"[2, p.80]. In the Uppsala County Council, patients are given the right to view the log of the accesses of their EHR. Nevertheless it has to be considered that patients might not always completely understand who needs to access their record. Therefore patients might suspect employees of illegal access if their names are unknown to them, even though the employee

had the right to access the record. This might be the case if a lab technician or another physician give some advice to the patient's doctor. Consequently, it is necessary to document or explain all accesses to the patient to provide a clear understanding of legal and illegal access. Another important topic is the ownership of Electronic Healthcare Records which is discussed in the following subsection.

5.1.1 Responsibility and Ownership of Electronic Healthcare Records

As Electronic Healthcare Records contain personal data, the question arises who owns the data. Is it the hospital, as they store the data? Is it the physician, as he creates the data? Or is it the patient, as he is the subject that the data is about? Figure 1 shows the distribution of answers in a survey conducted in 2005 with 2100 participants that tried to find out the patient's point of view on this question.

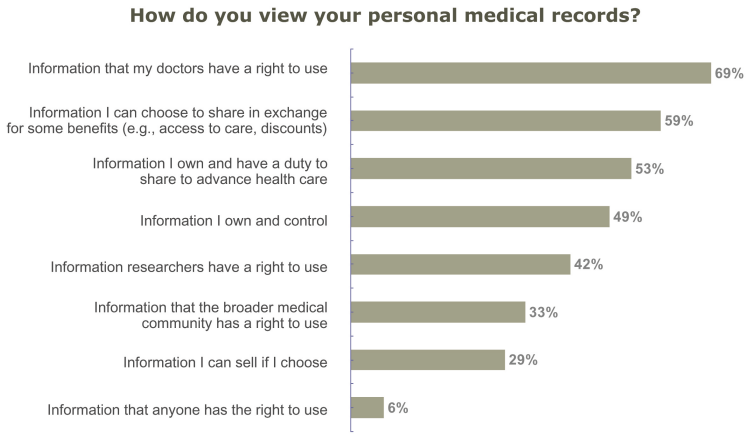

Figure 1: Patients' view on EHR ownership [6, p.23]

Interestingly, the distribution of people who think they own their EHR and those who do not is quite balanced. This might be caused by patients' lack of familiarity with the concept of stored personal data concerning them.

5.1.2 How is the Relationship between Patients and Medical Staff Affected by Electronic Healthcare Records?

The relationship between physicians or nurses and a patient provides an essential basis for successful healthcare. Even though it is obvious that medical staff has to rely on the information they are given, patients tend to keep secrets about their health or even lie. Besides the reason that they might not trust

the physician, lacking trust in the security could also stop people from disclosing information. Therefore it should be guaranteed that the principles of the Hippocratic Oath should be applied in the area of IT systems as well. In this Oath, physicians state to keep all information confidential that is disclosed by patients. It states: "Whatever, in connection with my professional practice or not, in connection with it, I see or hear, in the life of men, which ought not to be spoken of abroad, I will not divulge, as reckoning that all such should be kept secret"[20]. Obviously, when information is stored digitally, everything necessary should be performed to guarantee the same confidentiality.

5.1.3 In Which Cases Should Electronic Healthcare Records be Disclosed Without the Patient's Consent?

Personal information has to be protected and the patient's choice of non-disclosure has to be accepted, nevertheless there are situations where information should be accessed without permission. The most important reasons are: "(1) when treating in an emergency; (2) when accessing a record is necessary for commitment for mental health care; (3) when protecting third parties from dangerous patients; (4) when responding to statutory reporting requirements; and (5) when discussing a case with treatment collaborators for a patient"[40, p.114]. Even though it seems to be reasonable in these contexts to access the necessary information, there have to be clear definitions when they apply, e.g. what is an emergency or what does it mean to be a dangerous patient. According to Magnus Bergström from the Data Inspection Board, there are only two acceptable situations: emergency and the clear consent of accessing the data given by the patient. In the latter case it seems to be problematic how to make sure that a patient does not change his mind after some time and blame the physician for illegally accessing his or her EHR.

5.2 Ethical Problems from the Medical Staff Perspective

Ethical problems arise in the daily work of the medical staff. The privacy of the patients must be protected, e.g. by protecting their healthcare records. There are some problems however for the medical staff as well. Due to the new possibilities that Electronic Healthcare Records provide, the actions of the medical staff need to be logged. The medical staff can feel concerned about who is monitoring them, who is viewing the logs and if their actions are correct. Furthermore they might not know what they are allowed to do as they receive too little training in the systems they have to use every day. The interviews with physicians, nurses, researchers and a Ph.D. student revealed that most of them seem to know that they are not allowed to check their own EHR, unless they ask for a booking time with a doctor. Universities in Sweden and some hospitals train their personnel in the professional codes of ethics such as the Helsinki [44] and Hawaii declarations[45]. Nevertheless, some of them do not even seem to know that they are not allowed to view their own record which might be due to a lack of communication. The following subsections describe the

current situation concerning technical and ethical training and the awareness of logging in the COSMIC system at the Uppsala County Council.

5.2.1 Technical and Ethical Training

In 2008, the Swedish Data Inspection Board stated: "The entity responsible for storing personal data should create routines and tasks in such a way that it enables staff to work and think sensibly in regards to security" and "The entity responsible for storing personal data should make sure that everyone with access to personal data receives relevant training" [38]. According to the IT department at Uppsala Akademiska Hospital, all medical staff take an online course in COSMIC in which they learn the technical aspects of the system. All staff also receive a lecture in both technical and secrecy aspects of the Electronic Healthcare Record system. The online course is the same for all staff, but the lecture is arranged by each separate unit at the hospital. Therefore there is no guarantee that all staff receive enough ethical training before using the system. The only control of secrecy knowledge is made by having all employees sign a paper containing the following statements:
I am aware that:

- All registration I enter into a health care system is traceable and can be connected to my user account.
- I should always log out of the health care system when leaving my computer.
- I am responsible for protecting my password and not disclose it to others.
- Follow ups on events in the county council's health care system occur regularly.
- Breaking the rules stated above may lead to a police report.

For detailed information see Appendix C. This paper however is out of date since it states the laws affecting the use of the system. Two of these laws (Patientjournallagen - 1985:562 and Vårdregisterlagen - 1998:544) are not in effect any more. The laws replacing them (Patientdatalagen - 2008:355 and Patientdataförordningen - 2008:360) are not listed. The outdated paper is probably very hard to use to enforce proper use since the laws stated in the document are no longer in effect.

Even though training is an essential part in preventing illegal access to EHRs, it can not guarantee total security. This is similar to the fact that everyone has been taught that stealing is wrong, but there are still reports of thieves in the news. So on the one hand, claiming that information and schooling solve the problem is quite naïve. On the other hand, would there be more thieves in a society where some never get taught that stealing is wrong? A probable answer is yes. If we accept this assumption we should be able to lessen the amount of misuse of EHRs by informing the staff that it's wrong. In one case of misuse

reported to the police, the nurse stated that he did not know that his actions were illegal[25]. The lack of knowledge of how to use the system is backed up by Ph.D. student Rebecka Janols. According to her nurses and hospital staff feel unsure about the system; they have an ethical dilemma. They do not know what is ethical, non-ethical, legal or illegal. It appears that the signing of the document listed above is not enough. The ethical schooling is not enough either. The staff must be given more information about the ethical guidelines. It is also recommended to have regular reminders to increase the awareness of logging in the COSMIC system which will be discussed in the following subsection.

5.2.2 Awareness of logging in the COSMIC system

When looking for a patient in the COSMIC system, the employee can search by social security number or name. Therefore, it is possible to see whether or not a person has an EHR without being logged, as the logging starts only when opening the record and its content is viewed. There are no signals given to the employee that his actions are being logged. If the ethical and technical training did not emphasise this enough, the employee may not have any idea about what is logged. Some of the hospital staff do not know that the log system exists such as the Ph.D. student, some of them have a clue about it but not how it works or its consequences because they come from other countries outside Sweden and they have not been taught in the local laws as the doctor that works in Uppsala. The Ph.D. student said that she is very interested to know more about the log system and the Swedish rules during a training session at the hospital.

6 Prevention

With a system that can prevent users from illicitly or accidentally accessing EHRs, there will be less need for advanced detection algorithms that find these accesses afterwards. In addition, it is better to prevent misuse, as once misuse occurs the patient's privacy has been violated, and it is impossible to restore it. Preventing users from illegal or unethical use is mainly a question of education, but also a question of technical implementations into the existing system.

6.1 Prevention by Training

During the interview with the hospital staff, most of them said that they have not been educated in ethics. They suffer from legal and ethical dilemmas compared to those that have been trained or have a hotline to call and ask legal or ethical questions. A researcher from Stockholm and a nurse from Östersund got trained and are well informed about the ethics and integrity. The nurse is thankful for the hotline that answers their questions, the other people interviewed did not know about a legal committee or hotline at their working places.

There is ethical information online, but most of those interviewed said that they do not read it because they do not know where it is or because they do not have time to read it. In contrast the hospital staff in Östersund are trained and think differently about ethics, laws and logging system and are thankful for such training. A nurse who works in a hospital in the north of Sweden confirmed that they have an ethical and legal committee that answers their questions, thus decreasing their ethical concerns. Hospital staff feel better with hotlines that answer their questions immediately.

Hospital staff interviewed agreed that they would be happier if they recieved basic training in the Swedish laws that regulate the hospital system. The ethical training they get is mostly based on the professional ethical code (as learned at university) but not in the IT systems.

All users interviewed answered that they are afraid of writing, misusing, reading, or unintentionally changing an EHR, so they are very careful not to type in wrong names or medication.

As a preventive measure a doctor in Uppsala expressed that the system should help to remind the users about the laws and ethics by providing warning messages and more passwords when writing reports such as death certificates, instructions, or when reading and changing the information. An interviewed Ph.D. student said that it is easy to write a wrong id-number in the system and therefore get into the wrong EHR. They mentioned that it would be good if the system warns the user before opening and reading the wrong EHR.

In Östersund and Karolinska there seem to be fewer problems about misuse of information because the personnel are trained. In Östersund for example, they have a hotline for ethical and legal questions, the human resources department propagates information about the integrity and they check the log system every month.

Most of the interviewed users responded that the hospital staff is too busy to look into other information files such as legal laws and ethical advices. It would be good to train the personnel about the log system, integrity and ethics at the same time with the COSMIC system training, along with providing information about their personal account and responsibility.

The interviewed nurses from Östersund and from Stockholm and the researcher from Stockholm are trained in ethics and the log system. An interviewed user said that "trained staff feels thankful for getting a better understanding about the system, their rights and better integrity relation with the patients".

Most of them admit that they forget to close their accounts and normally they have also been using others' accounts. This happens because they are busy and need to attend to a patient urgently. Sometimes some of the personnel have been found making repeated misuse under another user name. Interviewed users admitted that they forget to close their accounts (they forget, they have to run to see a patient in emergency, they feel lazy since the system takes a long time to log in, etc.), and therefore others use their access to the system. The user should be trained in the importance of closing accounts. On the other hand the system should also understand that hospital staff are busy and they might forget about this, therefore there should also be an automatic system that saves and closes their accounts.

6.2 Prevention by Guidelines and Disciplinary Actions

When designing guidelines that people need to memorize it is important not to drown the users in too much information. A bad example can be found in the *Guidelines for information security*[28,] used in the county of Östergötland. These guidelines contains a five page listing what users must think of when using the system. For example it is stated that users are not allowed to install other applications on the hospital's computer and users should check for virus alerts and report them to IT maintenance. These are guidelines that should be dealt with by system administrators and should not concern the users.

When new policies and new guidelines are introduced that tighten the ethical standard in one area, users need to be educated but there also need to be clear disciplinary actions against people disobeying the new policy. If the system stakeholders do not set an example then users easily fall back into old habits again. A concrete way to do this would be to also inform all users when disciplinary actions are taken. For example when users log in to COSMIC a pop-up informs that a user has committed a certain action and thereby was subject to disciplinary action.

During the interview with the hospital staff, most of them said that they know disciplinary action may be taken if someone violates EHR access policy and that the punishment depends on the action, but they did not know exactly what kind of punishments were possible. The doctors and Ph.D. student agreed that there should be a punishment if someone is doing something in violation of the ethical standards, such as looking at an EHR without having any medical

relationship with the patient.

6.3 Whiteflagging

When selecting logs to be checked for illegal access the logs with a positive flag will be removed from the random selection process. The current system for finding illegal access in logs is both time consuming for the people involved, and has up to this point not produced any results that has lead to the identification of any illegal access of the healthcare records.

This method would work with the existing scheduling system and provide doctors who are scheduled to meet a patient with "a white flag" to access that patient's healthcare record. If the patient then is scheduled for a check-up later on, the white flag will be valid up to this meeting. In addition the staff of a given unit will be granted access to an Electronic Healthcare Record, if the patient is admitted to that unit, and will retain that access until the patient is released or signed out.

6.4 Account Management

When using this approach the system takes a defensive stand by only allowing users that have been granted access in advance to access the system. To manually handle what user should have access to a which patient's EHR would be too time consuming and might threaten critical access that needs to be done immediately. This method would consist of a detailed list of requirements that have to be met in order to grant a user access automatically. The list would include:

- Computer ID
- Medical unit that is treating the patient
- Users that the patient has been referred to
- Emergency unit has full access

When a user requests access to an EHR, the list of requirements are checked. If one ore more of the requirements are met the healthcare record will be shown.

An option to prevent all access that does not pass the list of requirements, would be to implement a feature with forced access. In this way parts of a healthcare record or the whole record can be made classified and not accessible unless the user makes a forced access which is logged separately and checked more often and more thoroughly. Additionally the system should make the user aware of the fact that he or she is making a forced access, for example with a pop-up screen.

Interviews conducted with current users EHR systems show that users acknowledge the problem with being logged in even though they leave the computer. One way to solve this problem would be to start using personal Smart Cards together with a PIN-code. If the Smart Card is included in the staff

pass card it will force the users to bring the card with them when leaving the computer and thereby be logged off. This would be necessary for the account management technique to be effective. However, even if the account management technique is not used, using Smart Cards would help to ensure the accuracy of the logs.

6.5 Search Term Limitation

Today the COSMIC system lets the users search by several key words such as name or part of a name. In this way it is easy for a user to access, for example, a neighbor's or a friend's Electronic Healthcare Records. The system that the county council in Stockholm is using, TakeCare, has another approach. In that system users can only search by patients' personal numbers. By doing this, users will be prevented from finding the people's records unless they know their full personal number. A feature that gives a user the right to search by name should be given to a restricted group of users, for example the emergency unit.

6.6 Prevention by Log Analysis

Log analysis might be a useful tool not just for detecting illegal accesses, but also for preventing them. When users of the system are aware that everything that they do is being logged, and that the logs are being analyzed, they may be less likely to abuse the system because they know that there is a higher risk of being caught.

The nurse in the scenario who accessed the politician's record illegally might not have leaked the information to the press if she knew that she was being logged and probably would get caught. She might not even have accessed the record in the first place.

During our interviews at the hospital, the employees also expressed concerns about the log analysis. They believe that if there are any uncertainties about what they are allowed to do, people will be afraid to use the system because of the fear of being caught with something illegal. The employees need to be able to trust the system, otherwise the quality of their work might suffer.

7 Detection

Since log analysis is required by law, it would be good to make it as efficient and effective as possible in order to save both time and money. This section describes and analyzes methods for selecting logs more intelligently, by finding patterns and record accesses that are different from the normal usage in some way. This is called novelty detection. Another idea is to improve the way the selected logs are analyzed, by supplying tools for the staff who today go through the logs manually. Some of the content of this section is quite technical in nature. The information is summarized in a less technical fashion in Section 8.

7.1 String Matching

The idea of string matching comes from the way some animals', including humans', immune systems work. The body protects itself against foreign elements without knowing what these elements are. The immune system categorizes things as either part of the self (S) or the non-self (N) (see Figure 2). As long as it does not look like something that the body knows, the body will assume it is dangerous and attack it [9, p. 1]. In string matching, this is done by looking at data known to be legitimate and in some way generating the opposite of these. If new data then matches something in the set of opposites, that data is deemed abnormal. This technique has been used within the areas of detecting breakage in tools and network intrusion. There are different methods of generating these opposites and comparing the new accesses, all with different advantages and disadvantages. The following sections deal with the technical aspects and problems of some of these techniques and their variations.

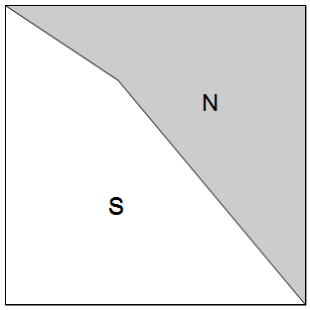

Figure 2: Actual partition between self and non-self space

7.1.1 Prerequisites/preparation

In order to use string matching the data needs to be encoded into some language Σ^l (normally a binary language). The strings will look different depending on

if the system is used to look at individual accesses or one employee's usage. Each position in the string will represent a feature that is of some importance to figuring out if the access should be regarded as normal or not. If Σ is the binary language of {0,1} and one feature can take on more than two values then more than one position can be used for this feature. If the system is looking at each employee then it would probably be best to aggregate an employees record accesses over some time and perform the analysis once a day/week/month instead of in real time. Regardless of what the system is looking at, the algorithm creates two data sets, S (self) and D (detectors), both subsets of Σ^l. The self set is usually derived from observing the system and taking points or in this case, record accesses or staff patterns, that are known to be proper (see Figure 3). Given the self set, there are then different ways of generating the non-self (see Figure 4). Three of those methods are described below.

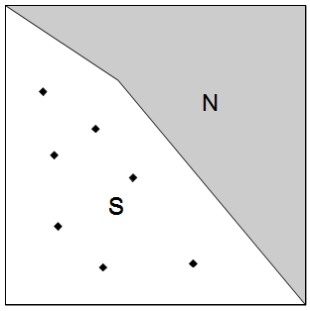

Figure 3: Observed samples in the self space

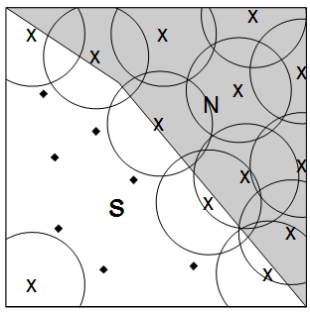

Figure 4: Generated detectors not covering the known samples in self space

7.1.2 r-chunk and r-contiguous

Elberfeld and Textor present two different ways of generating the detection set, D, and measuring the closeness of a sample to D; the r-contiguous and r-chunk

methods [9, p. 4]. In both cases D is created by randomly generating strings and comparing them to the data known to be part of S. If the generated string matches any string in S, it is removed from D. Unknown data is classified by comparing it to strings in D. If it is considered to be close enough to a string in D, it is said to be part of the non-self and therefore an anomaly. An r-chunk detector consists of two parts, a string of length r (also known as the r-gram) and an index i. For a sample s to match the r-chunk detector the substring of length r that starts at index i in s must be equal to the r-gram. An r-contiguous detector on the other hand is of the same length as the samples, and must match at r adjacent positions, regardless of what index the match is found at. One could illustrate the r-contiguous detector as lining up two strings next to each other and seeing if there is some place where they correspond. For example, consider the samples $s_1 = abaaaa$ and $s_2 = bbaaba$. The 3-contiguous detector $bbaaaa$ would match s_2 because of the first 3 positions and s_1 because of the last 3, while the 3-chunk detector $(aba, 1)$ would only match the first sample, as the r-gram aba is found at index 1 in s_1 and at index 4 in s_2.

7.1.3 Schemata

Hang and Dai present a way of converting vectors of numerical and categorical values to a binary schema representation [14, p. 278]. They also present a way of generating detection rules by using co-evolving genetic algorithms to generate different schemata to represent the self and non-self data space. In essence, all parameters in the system are not specified up front, but rather worked out by the computer generating different kinds of detectors and testing how well they perform and using the ones that perform the best as a base to generate new kinds of detectors. Note that what is evolved here isn't the actual detectors but rather the way they look, i.e. what different attributes to include as well as length and form of the detectors. These forms, or schemata, are then used as a template to generate the detectors, which are then pruned in a similar way as in the r-chunk/r-contiguous methods.

7.1.4 Complexity

The r-chunk and r-contiguous techniques can be proven to have a training complexity of $O(|S|lr|\Sigma|)$ and a run time complexity of $O(l)$ [9, p. 8]. This could mean that the actual checking of record accesses could be done in real time, i.e. if the system is trained to look at individual accesses to see whether or not they are novel then this check could be done every time a record is opened.

Since generating the rules with genetic algorithms, as proposed by Hand and Dai, is more advanced than the r-chunk and r-contiguous approach it will probably be computationally more expensive to generate rules this way. It will, however, generate rules that promote the more important features in the problem space and thus more efficiently generate an appropriate number of diverse detectors [14, p. 282]. Once the detectors are created, these can be used just as quickly as the r-chunk and r-contiguous detectors.

7.1.5 Problems/requirements on data

Neither of these approaches learn from new results. The only way to incorporate new data into the algorithm is to retrain them and include the new data in the training. The problem with using string matching in detecting suspicious health record accesses is that there is no good representation of what "normal" usage is available. Since no clean training data is available, the algorithm will include the suspicious usage when training and generalize from this and then assume that accesses that are similar to these are normal. There is also the problem of overfitting in the training. This means that the system will generate too many detectors, and thus come too close to the training samples given and not be able to generalize much beyond those (see Figure 5). On the other hand, if too few detectors are generated, then the system will generalize too much and illegal activity will be classed as part of the self set.

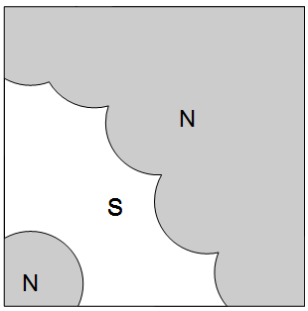

Figure 5: Inferred partition between self and non-self space

As mentioned above, the data needs to be converted to some language Σ. This probably means it would be it easier to export the logs to some separate software instead of trying to incorprate it into COSMIC.

7.1.6 Conclusions

String matching has been proven to work in the field of novelty detection, but it would probably be difficult to implement at the Uppsala County Council because of the problems mentioned above - mainly the problem with the technique requiring a lot of training data in order to work. If one could find a good way of representing the record accesses and manually supply a lot of training data, then string matching could be a viable option for making a more effective selection of logs to analyze.

7.2 Clustering Algorithms

Cluster analysis is used to divide data into various groups. These groups are selected such that members of a group are highly similar to each other, and dissimilar from members of other groups. These clusters, or meaningful groupings,

are most commonly used to understand and interpret data [36]. Humans are very good at developing groupings and classifying new stimuli into these groups. However, with increasing data sizes, it is necessary to develop algorithms to find such clusters automatically. These algorithms attempt unsupervised classification, deriving clusters only from the data without any user input. Various algorithms exist, each possessing various strengths and weaknesses. Clustering algorithms appear in three main types: partitional, hierarchical, and density based. Clustering algorithms are all fundamentally based on a distance measure between points. While the technical details of this process are beyond the scope of this report, a difference metric for log records will need to be developed. This may be accomplished by translating log records into a datatype with a defined metric, or developing one directly. Such metrics exist for comparing text.

7.2.1 Partitional Clustering

Partitional clustering algorithms are those algorithms which partition the data, placing each data point in exactly one cluster [36]. The most common partitional clustering algorithm is K-Means. K-Means [36] divides the data into exactly K clusters, where K is a user specified variable. In practice, the algorithm is performed for many values of K, and the model that yields the best results is chosen. The algorithm proceeds as follows:

1. Select K random points as centroids
2. REPEAT
3. Form clusters by assigning each point to the closest centroid
4. Recalculate centroids
5. UNTIL Centroids do not change.

The selection of random starting points can obviously be problematic, as poorly chosen starting points can produce highly inaccurate clusters. As such, the algorithm is generally run multiple times for a specific value of K, especially if the first results do not seem accurate.

K-Means is comparatively fast, even with multiple calculations to control for the random starting points. K-means' principal disadvantage is that it cannot handle irregularly shaped clusters, or clusters of differing sizes or densities well.

7.2.2 Hierarchical Clustering

In contrast, hierarchical Clustering [36] allows each data point to be a member of multiple clusters. Clusters form a hierarchy, ultimately resulting in a cluster of all points. The algorithm begins with each data point as its own cluster. At each step, the closest two clusters are combined until only one cluster remains. This method can be problematic as the combination is irreversible. As such,

larger clusters may form incorrectly, based on the actions of smaller clusters. In addition, the hierarchy created is ill-suited to log analysis. For log analysis, a specific log must be grouped as either illicit or legitimate, not both.

7.2.3 Density Based Clustering

Density based clustering is used to partition data by finding areas of high density that are separated by areas of low density. One very common algorithm is DBSCAN. DBSCAN requires two user parameters, E and M. To begin, points are classified as either core, border, or noise, according to the following definitions:

- Core: A point which has at least M other points within a distance of E. These are interior to a density cluster.
- Border: A point which is not a core point, but falls within E of at least one core point.
- Noise: A point which is neither Core nor Border.

The algorithm then proceeds as follows:

1. Eliminate all noise points.
2. Connect all core points within E of each other as a cluster.
3. Assign border points to the cluster containing the largest number of core points within E of the border point.

This algorithm is also reasonably fast. One potential issue is the removal of noise points. Illicit log accesses may be noise points, as they are (presumably) much less frequent than legitimate accesses. However, it would be possible to consider all noise points for further analysis, with proper selection of the parameters.

7.2.4 Efficacy

With proper parameters, it is likely that a clustering algorithm could separate illicit log accesses from legitimate accesses. One issue may be the comparatively small amount of known illicit data. This may make it difficult to determine which clusters indicate improper access, as this data may be overwhelmed by the quantity of legitimate access. On the other hand, clustering techniques will improve in performance as more illicit access patterns are found. Since the clustering is unsupervised, additional illicit patterns will not directly change the clusters produced. However, selection of likely illicit records will be based on those records which cluster with known illicit records. As more records are available, this will be easier and clearer.

7.2.5 Cost

Clustering algorithms are highly scalable, as they were designed to work on large quantities of data. Once clusters are created it may be possible to quick check records against the existing clusters. In this process, a record would be grouped into a cluster, but the clusters would not be recalculated. This would be much faster than recalculating the clusters. Recalculation is unnecessary after each record addition, as the clusters are unlikely to change significantly based on one additional record. Clusters could be recalculated on a periodic basis. Clustering algorithms are fairly easy to implement, and tools implementing such algorithms exist. These tools would need to be modified to work with log records, however. Further, it is unlikely that the clustering analysis could be directly integrated into COSMIC. However, data could easily be exported from the COSMIC database to an external clustering application for analytic processing.

7.3 Association Analysis

7.3.1 Background and application areas today

When a computer technician thinks of data mining, the method of association analysis (AA) is what usually comes first to their mind. Derived from mathematical statistics, database technology and retail business, it was one of the very first widely used data mining techniques. The first major application was to efficiently classify customers and their shopping behaviors. The classification of customer had such a deep impact that there was even a modern myth created about the phenomena. The myth tells that Wal-Mart (a large American retail business) software engineers, after testing the new analysis method found a strange and unsuspected customer behavior. The system found that customers who bought diapers, were very likely to also buy beer. This led the Wal-mart owners to put the beer right next to the shelf of diapers, to see if it would raise profits, which it did.

Besides the usage within the retail industry, AA is now commonly used within the areas of bioinformatics, medical diagnosis, web mining and scientific data analysis. Since it was based upon statistical mathematics, many earlier underlying mathematical methods could soon be adapted to this new application. This resulted in a quick development which gave birth to many new AA -variants and similar techniques. The methods soon came to define a new stand-alone scientific subject, namely data mining.

7.3.2 Method

The process of data mining usually contains a number of different steps which are common to many different techniques. One frequent first step is “data preparation”. This is because when a system stores data, it is very likely that it is stored in database system. To be able to do efficient AA on interesting data, it is usually necessary to clean, filter, transform and select subsets of the

data, in order to be able to perform AA. This data preparation is often time consuming since massive amounts of data can be available in a bad format.

When the data is ready to be analyzed, the first step of AA is Frequent Itemset Generation. An itemset is a set of data items which are in the same transaction. A transaction is a collection of items which are in the same time-period.

E.g. when a customer buys both milk and butter and then pays at the cashier, it generates a transaction with the itemset (Milk,Butter)
The occurrences of the itemsets are then counted so that mathematical filtering rules can be applied. By tuning the parameters of these rules, we are then able to find itemsets which more and less related to each other. For example, perhaps buying diapers at Wal-Mart implies other purchases are likely on the same trip.

7.3.3 Healthcare and loganalytics

To prepare an electronic health care system for AA, with the intention to detect misuse of Electronic Healthcare Records (EHR), it is necessary to reconsider the definition of a transaction. It can no longer be what the patient (earlier customer) has bought, but who has accessed his or her EHR.

Given that one transaction is either a patient's EHR accesses for one day, or a patient's EHR accesses for that patient's "health care visit", and given information about illegal access patterns, it is possible to classify and find both typical and unusual behavior. For example, consider the registries below.

Example Staff Registry:
Doctor1: id=19641210-0123, name=Adam, dep=Surgery, adr=[Studentroad 12a,UPPSALA,75230]
Doctor2: id=19560420-3421, name=Peter,dep=Urgent_care, adr=[Eventroad 1,UPPSALA,75223]
Nurse34: id=19740103-6423, name=Anna, dep=Surgery, adr=[petersroad 3,UPPSALA,75540]
...
Example Patient Registry:
Patient1: id=19761002-5342, name=Niklas, Adr=[Studentroad 14b,UPPSALA,75230]
Patient2: id=19870405-0523, name=Jimmy, Adr=[Anderssonsroad 11,UPPSALA,75342]
Patient3: id=19900203-0425, name=Cammeron, Adr=[Palmdrive 56,UPPSALA,75675]

A scenario where Doctor1, Doctor2 and Nurse34 accesses Patient1 healthcare records during 2009-09-02 and Doctor2 and Nurse34 accesses Patient1's healthcare records ones again the day after would appear as:

Patient1-2009-09-02:[Doctor1, Doctor2, Nurse34]
Patient1-2009-09-03:[Doctor2, Nurse34]
In this case, AA would indicate a strong correlation that if patientX has meet Doctor2, patientX has also meet with Nurse34.

As these entries will build up logs for many EHR accesses, the system administrators will soon have the possibility to determine what is common and uncommon behavior. Even if there is too much information for a human inspec-

tor to be able to observe suspicious patterns, the system will quickly find them. For example, combining the right data in a suitable algorithm will swiftly find that Doctor1 and Patient1 actually live on the same street and that they are almost neighbors. This is not a reason for a judicial sentence, but it might be suspicious enough to recommend further investigation.

7.3.4 Analyzing a moving target

Through interviews with the medical staff at Uppsala Akademiska Hospital we have found that there are a huge number of different specialties and positions a staff member can posses. Furthermore, these specialties and positions are likely to change over time. Even if the AA-method is effective in classifying typical and atypical behaviors, this interview result indicates there might be a huge number of different kinds of users/patterns to be found. This volume of patterns could make it very difficult for the algorithm to locate patterns, as there may be little data for each pattern. The largest negative of AA for these purposes would be that it is a non-learning algorithm. Since an AA-run processes the whole dataset at once, new behavior by the users in the system will be valued exactly the same as already known behavior.

7.3.5 Cost and complexity

Running AA over data is very expensive due to the sorting and counting of all included elements in all the involved transactions. When the data mass to be analyzed increases linearly the amount of operations to be performed scales $\Theta(2^n)$. This soon becomes an obstacle for anyone who wishes to analyze more than just transactions. There are several solutions to this issue, the best known is the Apriori algorithm. The Apriori variant tackles the problem by exploiting a few mathematical properties and effectively reducing the itemsets involved in the calculation, thereby lowering the complexity. The result is all the itemsets which appear at least n times, where n is a user defined variable.

Since AA originally runs over an set of transactions which has to be chosen before runtime, it can not, by definition, run in real time. If the technique was to be implemented for log analysis, the algorithm would have to run over a separate set of logs, possibly overnight or at other times when the computing load is low. To remake the AA algorithm for real-time-analysis, there would have to be changes made to make the algorithm learn and adjust to new circumstances. This technique is then called association rule learning.

AA is a well-known and easily implemented algorithm. The problem at hand at the Uppsala County Council is a bit more complex than the algorithm can handle, and the algorithm should thus be considered a small building block in a larger log analysis tool.

7.4 Neural Networks

Artificial neural networks are a set of simplified methods to mimic how the human brain processes information. The human brain contains billions of nerve cells called neurons that are interconnected with other nerve cells to construct a huge complex parallel processing network of so called synapses where each synapse has an input, an output and a body. The artificial neural networks consist of artificial counterparts of the neurons and the synapses where each neuron, or "node", takes other nodes' outputs as input and thereby mimics the brain's way of processing data.

Neural networks are widely used in applications designed to detect and recognize patterns or behaviors. The areas range from face recognition or detecting ships in noisy images[1] to create automated segmentations of magnetic resonance images using pattern recognition[26].

7.4.1 The perceptron

One of the more simple types of neural networks that can be used as a linear classifier is the perceptron. It consists of a single node which takes an arbitrary number of inputs, each assigned a dynamic value known as a weight. The weight is dynamic in the sense that it changes while the network is learning. The final output of the standard perceptron is a binary value depending on whether the cross product of the input vector and the weight vector is greater than a certain threshold [16, p. 2-4]. Expressed mathematically, where w is the weight vector, x is the input vector, and b is the threshold, the output of the perceptron is given by:

$$f(x) = \begin{cases} 1 & if \text{w} \cdot x + b > 0 \\ 0 & \text{else} \end{cases}$$

To be able to learn and to recognize, the network needs to be trained using various methods such as supervised or unsupervised learning. In supervised learning, sets of positive and negative training data are available in form of certain input data that yield certain output data. Positive examples tell the network how to act on that kind of input and negative examples tell the network how not to act on certain data. Based on this training, the network changes its weights to produce the correct results. A network can also be trained by letting the training set be generated by some deterministic function. The possibilities for the single perceptron are limited since it only separates linear problems i.e separate two sets of patterns with a linear function. The multilayer perceptron on the other hand can distinguish data that is not linear separable by using a more complex function to separate it. An illustration of the difference between linear and non-linear data can be seen in Figure 6.

The multilayer perceptron is well represented as a novelty detection method and there are plenty of variations of the technique that can be applied to different problems. Despite the fact that multilayered perceptrons are the most used and best known type of neural networks, it is hard to bypass the generalization

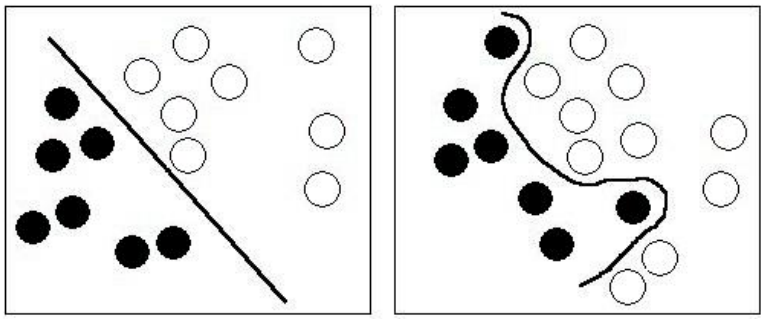

Figure 6: Linear separable data versus non-linear separable data

problems neural networks have when dealing with unseen data[18]. The problem occurs when a network becomes trained to fit a specific set of data resulting in poor generalization since the network is not familiar with data outside its training set and becomes overfitted. The design of the network also affects the generalization performance since an overuse of hidden nodes and layers, i.e nodes in some middle layer, also yield poor generalization [16, p.9].

If the environment is very complex and very different in the way a pattern is considered as normal behavior, as it could be when dealing with log analysis in a hospital since a lot of combinations of different actions can be classified as normal behavior, the network must be designed to fit the environment without being too generalized or overfitted.

7.4.2 Hopfield Networks

The standard Hopfield Network is a type of neural network where all nodes serve as both input and output nodes. This type of neural network is a recurrent network, in which the connections between nodes form a directed cycle and the output from this is the current state of the network itself, often called the "energy of the network" [19]. Feeding an input vector to the network will generate an energy value that is either high or low. If the input is recognized, as in seen in training, the energy will be low. On the other hand when the input is novel the energy will be higher and therefore easy to detect[4]. In the equation below the energy function is defined with N being the number of neurons, x_i is the activation of neuron $i = 1, ..., N$ and w_{ij} being the weight of the connection between neuron i and j.

$$E(x) = -12 \sum_{i=1}^{N} x_i \sum_{j=1}^{N} x_j w_{ij}$$

In [47, p.147] the authors summarize "The Hopfield network provides one of the strongest links between information processing and dynamics. However, spurious memories limit its capacity to store patterns." The Hopfield net might

give good results as long as the number of patterns that describe when someone is accessing EHRs are limited so as to not affect the performance of the network's energy function.

7.4.3 Validation

Bishop[3] writes about network validation being one problem when implementing neural networks in the field. It is difficult to prove that the output generated by the network is accurate. Bishop separates network validation into different levels. At the first level, some additional associated measure of confidence is combined with the network output to make it easier to distinguish the more reliable output data from non-reliable output data. The non-reliable data can then be analyzed further by other methods or by human experts. For example when a multilayer neural network processes EHR log data, it could be possible to flag logs as suspicious, non-suspicious and uncertain.

7.4.4 Computational Complexity

The computational complexity of a neural network depends on its design. There are many elements which factor into the complexity. For example, network topology, number of nodes and layers, the network activation function that is used to map input to output, training phase with update of weights and various back-propagation methods used to feed the result back to the network, as well as other factors affect the complexity of a neural network.

7.4.5 Conclusions

In order to get good results from a neural network it needs to be trained with a lot of training data. Examples of training data in this context include combinations of access patterns to some EHRs that are classified as either acceptable or suspicious. A problem here is the lack of known suspicious patterns and the complexity of the patterns to make the network generalized enough.

A novelty detection method based on the example above can not distinguish between these two classes if there is not enough training data since the network would probably label data incorrectly. If a good enough amount of training data can be gathered this is a great technique to use, especially in combination with other similar techniques.

7.5 Log Anomaly Visualization Application - LAVA

The Log Anomaly Visualization Application (LAVA) is a monitoring framework using access logs to detect anomalies in an EHR system proposed in [27].

Many system designers have tried to improve security in EHR systems by incorporating greater security protocols. However, standard work flow and information needs of clinicians differ so much from the practices existing security models attempt to protect. As mentioned above, the protection of patient information from misuse of employees in the organization is not sufficient. Trying

to remedy this issue, many organizations use auditing as the primary defense against insider threats. However, EHR systems with thousands of clinicians and tens of thousands of patients require enormous auditing systems, and it quickly becomes impractical to use systematic auditing.

LAVA is a proposed monitoring framework that uses access logs to model current trends in the system, and later detect anomalies compared to those trends.

Typical solutions for potential insider attacks are threats of disciplinary actions if misuse is discovered. Unfortunately, using auditing this way is more a threat than a tool, since the volume of accesses precludes auditing from ever finding the "needles in the haystack", claims Paulett [27, p.1]. However, he does recognize auditing to be productive in the rare situations when evidence indicates improper access. In contrast, LAVA can analyze access patterns of users, visually display access trends and outliers of those trends, and present suspect accesses in a ranked list. The application transforms the random audit to a guided audit.

LAVA is compatible with systems from any EHR-vendor, and can handle multiple EHR systems using a database table as an abstraction layer. Using LAVA, logs are transferred from the EHR systems into a database table. The application will run with any backend database. Depending on the number of metrics that are available for extraction from the logs, i.e. the number of different categories of parameters in the log files, LAVA uses these metrics for scoring and aggregating log entries. Scoring log entries of users is the method used to detect anomalies in the system. It is possible to find aggregated scores for all users of the system, but also the score of one single user. Examples of different entries in the log that are valid for scoring are *Number of Patients Accessed, Number of Total Patient Accesses, Accesses per Patient, Time between Accesses* and *Time between Patients*. Using an anomaly scoring algorithm, LAVA is capable of pinpointing users with anomalous metrics calculated from the users respective log entries with regards to the different entries exemplified.

These functions make LAVA a helpful tool by turning the existing random audit to a guided audit.

7.6 Log Analysis

The sections above describe several techniques for extracting log entries concerning illicit access of EHRs. The techniques are more or less intelligent methods of extracting suspicious candidates. In this circumstance a log analysis is to be regarded as merely testing a set of rules or conditions against a log sample. For instance, one could test a log excerpt, depicting an EHR access against a rule to verify that the patient actually was admitted to the hospital at the time. In the subsequent sections two different means of log analysis will be introduced and explained in detail.

7.6.1 Using Log Analysis as an Extension

Once a set of suspicious log records has been obtained using the methods introduced one could evaluate the set of rules against them. Thus, the analysis will give a further indication of whether the log record depicts an illicit access or a legitimate one.

7.6.2 Using Log Analysis as an Alternative

Regarding the analysis as an alternative would mean that the set of rules shall be evaluated against a sample from the log. These samples could involve perhaps every given log during a certain period of time or a random selection of the total collection. A good example of using the analysis as the whole solution is the Systematisk Automatiserad LoggAnalys (SALA) system, see B.2.2 for reference, where analysis is performed on random samples. This section also contains the criteria used for analysis.

7.6.3 Extending the Attributes

In order for the analysis to be effective the set of rules must cover as much as possible. Hence, it is crucial that the rules are constantly evaluated, extended and improved. The evaluation ought to be performed against some established data. Furthermore, a system which is designed to evaluate these rules should possess the capability to dynamically alter, add or remove rules. A great example of such a system is the Swedish National Police Board's log management system (see B.3).

The list below consists of possible attributes to consider during analysis. It is mainly an extension of both the LAVA and SALA systems. However, some of the examples below are the exact same ones from the LAVA system as described in Section 7.5.

- **Number of accesses to a specific EHR**
  The number of times a specific employee accesses a specific EHR, evaluated relative to the employee's work task and work place. It may be that certain employees have a legitimate reason to view many patients records once on several occasions, perhaps a receptionist, and others may view a specific record several times during a period of time, perhaps a psychologist.
- **Geographical Location**
  The geographical distance of home addresses between the patient and the employee, as many cases have consisted of employees snooping on their neighbors.
- **Name**
  The surname of the patient and the employee to avoid checking up on relatives. This can perhaps be somewhat dubious since many people share the same last name without being relatives. Those names may therefore be excluded from the comparison.

- **Number of patients viewed by an employee**
  The number of patients a certain employee has viewed during a certain period of time, again evaluated relative to work task and work place.
- **Number of accesses to a medical record**
  The number of accesses per medical record. If a certain medical record has been viewed an unusual number of times, it might be a record belonging to someone famous or important.
- **Time between accesses to a medical record**
  The time between accesses to a medical record. The time that has passed between specific accesses to a given medical record.
- **Time between accesses to different medical records**
  The time between accesses to different medical records with regards to a certain user. The time between accesses made by an employee to different medical records. If the passed time is shorter than the normal time it takes for an employee to treat patients it could be an indication of illicit use.

7.6.4 Using Log Analysis in Practice

The most beneficial scenario would be to use the analysis as an extension to an intelligent selection of suspicious logs. However, this would imply producing two separate systems which would work in cooperation and therefore necessarily increase the cost of developing the combined system.

Otherwise it would be possible to use only the log analysis to determine illicit use, much like the SALA system. Furthermore, the log analysis could be disregarded all together and in favor of a solution which only involves the intelligent selection.

In summary, it depends on the requirements of the organization, the funds available and the timeframe for the development of this system.

8 Summary

This part of the document contains the principal conclusions given by the results, interviews and investigations during the project.

8.1 Prevention

8.1.1 Education

All hospital staff have received significant amounts of medical training during the years leading up to their employment at various care units. In order to keep the patients' trust, most personnel also receive ethical training. After becoming employed by a healthcare provider they are further trained in the use of technical equipment and computer systems.

All training is standardized by either the Swedish National Agency for Higher Education or the county council offering the care.

There is however no standardization of how to educate medical staff in the ethical issues of data handling. Without this standardization there is no way to guarantee that all medical staff receive the same training or any training at all.

8.1.2 Technical Solutions

We have examined three types of technical preventive solutions:

- Whiteflagging (section 6.3): Pre-approve a staff member's access to a patient's record (e.g. due to a scheduled appointment). This allows for those EHR accesses to be exempt from subsequent analyses.
- Account Management (section 6.4): Disallow access to a patient's EHR unless you have the proper authorization. A forced access feature could be implemented for use in case of emergency. Automated log-out using combined smart cards and pass cards.
- Search Term Limitation (section 6.5): Find a patient's EHR by personal number only and not by the patient's name or a part thereof.

All three could quite easily be implemented at the Uppsala County Council as long as you keep in mind that authentication and access control are two of the nine services covered by BIF (Appendix B.5). The BIF time frame would thus have to be taken into account when the decision is made on what solutions to implement.

8.2 Detection

Based on the detection methods presented in section 7, the following conclusions can be made with regards to implementation at the Uppsala County Council.

- String matching (section 7.1) seem to be difficult to implement at the Uppsala County Council. The main problem with the technique is that it requires a lot of training data in order to work. If a good way of representing the record accesses can be found, and manually supply a lot of training data, then string matching could be a viable option for making a more effective selection of logs to analyse.

- Clustering algorithms (section 7.2) are fairly easy to implement, and tools implementing such algorithms exist. These tools would however need to be modified in order to work with the log records in COSMIC. Further, it is unlikely that the clustering analysis could be directly integrated into COSMIC. However, data could easily be exported from the COSMIC database to an external clustering application for analytic processing.

- Association Analysis (section 7.3) is a well known algorithm and seems to be easy to implement. However, the log analysis requirements at the Uppsala County Council is of greater complexity than what the algorithm can handle. Association Analysis should therefore be one part of a log analysis system that use additional methods.

- Neural Networks (section 7.4) are complex and resource-demanding. It would most likely be difficult to implement in a log analysis system at the Uppsala County Council as they need extensive training data in order to function adequately.

- LAVA (section 7.5) is a detection tool that seems promising and should be investigated further. It is an EHR-vendor neutral system witch means that it could be used with COSMIC or any other EHR system. However the EHR logs needs to be translated into a format that LAVA can understand. There is one big problem, it seems that LAVA has not come into production yet.

LAVA has some interesting similarities with Systematisk Automatiserad LoggAnalys (SALA) (Appendix B.2.2), for instance; neither LAVA nor SALA act as a replacement for the privacy officer's manual job. Usually, the role of privacy officer is taken by the director of a specific unit who has knowledge about the COSMIC users at his/her unit. LAVA and SALA rather act as tools that can assist the privacy officer in finding the most interesting access patterns and manually reviewing them. Another aspect is that both have similar rules, or metrics that they base their analysis on.

9 Discussion and Future Work

9.1 Introduction

This section of the report presents a discussion where the results of the research are explained further and put into context. In the discussion we will interpret the results as well as deduce them with research papers and reports.

As argued in the report, log management in Electronic Health Care Records is an intricate subject. It involves both ethical and technical issues which should be addressed properly. The new possibilities that electronic health records provide means that current legislation and guidelines have to be adapted to the use of this new technology. Ethical codes are an important aspect of health record handling and must also be adapted to these new conditions. These and other consequences of the use of logging of computer systems will be elaborated further below.

The studies at the academic hospital shows that the Electronic Health Care systems available today enables hospital staff easy access to a vast database of personal information. To maintain security in the new electronic systems, activities have to be logged in order to detect and prevent misuse and breaches. But how extensive should the log system be? The medical staff who are the end user of the system need to feel that they can work with the system without feeling insecure. This can be resolved with different kinds of measures.

9.2 Prevention

When technical systems evolve, it seems that they get more and more advanced, and if they are not designed with the user in focus,[5] the interaction between the system and user can present difficulties. Hence it is important that the end users receive proper education before using the system. However, the people who were interviewed for this report raised concern that they had not received a satisfactory education. This consequently leads to users feeling insecure and uncomfortable when using the system. The development of technical systems typically cost large amounts of money, and it is of great importance to include education costs when systems are budgeted.

One important issue to address in the education of the users, is what consequences individual user actions will lead to. Users need to be informed of what actions and information is logged, and what happens in the systems when they perform these actions. One solution could be more feedback from the system to the user. Warnings could for instance be issued for certain actions, such as "When working in COSMIC, remember that your actions will be logged". The education and information for newly hired medical staff should be planned with given knowledge in mind and preferably also be followed up within a reasonable time to ensure the knowledge is steady among the users.

One of the main questions that arose in the study was if the staff were allowed to access their own records. The people who were interviewed were not sure whether they were allowed or not. We had to contact the National Board of

Health and Welfare in order to get a definitive answer. The only staff members who are allowed to see their own medical records are the doctors, although they are only allowed to see their own records if they are ill and plan on treating themselves. The question of whether the rest of the medical staff are allowed to look at their own records was hard to get an answer to, because of this it is safe to assume that several people in the medical staffs around the country have little knowledge about this. The information regarding this case was very difficult to get.

The interviews revealed that the users are uncomfortable using the system. They tend to see the logging of the system as a way for the hospital to monitor its staff in order to use the information against them. As a result, many users are afraid to use the system, which in turn counteracts the purpose of the system. This issue might be resolved if emphasis is put into explaining to the staff that the log system is intended to protect the users, for instance if they are accused of misuse. Obviously, it is a minority of the users who have bad intentions with their actions. Log systems are intended to catch those particular users. Proper education and information could never prevent planned misuse, but it will help the user with good intentions. People who use the system illicitly will always pose a threat, therefor actions other than education and information have to be taken to counteract these people.

A solution to prevent illicit use could be to handle each user account for every medical staff member manually. Doctors, nurses and other medical personnell should have different level of access to medical records. This solution has drawbacks though, to manually handle each account is time consuming. It could also threaten critical access to time-critical information, for example at an emergency unit at a hospital. This could possibly be resolved if the emergency unit is granted access to medical records without any special account management.

In COSMIC, it is possible to search patient records by using both name and social security number as keywords. This poses some problems. This makes it easier for unauthorized people accessing medical records. If the medical staff was only allowed to search for the patient by the social security number, unintentional and illicit use could be avoided. This solution is implemented at SLSO in Stockholm.

When a forbidden access is detected and proven faulty, the director of the department at the hospital should make a statement and inform the medical staff that an illegal access has occurred, and point out the circumstances of why that the access was faulty or illegal. The staff member who was responsible for the illegal entry should receive appropriate disciplinary action. People who get caught today are often subject to some form of disciplinary action, but far from all users get caught because of misuse. This has to be improved for the future by doing improvements to the log management system. Examples of possible improvements and solutions for logging effectively are described in following sections.

Another task to look into is to investigate how education can be improved to fit the needs of different users. Specifically, research should study both the

necessary content and the best way to integrate the education into everyday work.

9.3 Detection

As been stated in the detection section of the report, it is difficult implementing automated log analysis when there is a lack of good training data. It is difficult because in a workplace such as that of a hospital, there are many ways for medical staff to do tasks, both ways which classify as legitimate and suspicious. If a set of actions corresponds to a task, there are a lot of actions in that task that might be classified as suspicious because the log analysis system has not been trained to handle them. Those actions might in reality be completely legitimate since the environment in a hospital is dynamically changing and the medical staff have individual ways of working. There are simply too many correct ways of doing something and too few ways of saying and proving that something is incorrect. One could link this to the theories of situated action[35] where most people do not tend to work as predicted and not in a orderly fashion using multitasking on their tasks.

9.4 Other systems

That fact that log management and patient integrity is an issue with electronic health care records has been addressed in most sections of this report. At a national level, the BIF initiative was taken with this issue in mind. The implementation of BIF is however progressing slowly and the concept with a centralized system is not very well accepted among caregivers.

At SLSO in Stockholm, the SALA system has been developed and is currently being used. SALA does log analysis on accesses of ten random employees each month, flagging anything that it finds inappropriate and manual viewing of the logs in detail. The system has proven helpful with filtering a vast amount of logs into a more manageable collection, minimizing the work for the employees in charge of the actual log analysis.

The log management system currently in use at Swedish national police board is a good example of a highly adaptable system which performs well for its organisation. One of the reasons for why it works so well is that requirements for the system was researched thoroughly and stated at an early stage before implementation. The requirements are partly derived from the business requirements. Business requirements are important to take into consideration because they indicate what kind of information that could be valuable for the log system. If a clear view of the organisation and its processes is lacking, it will be problematic to know what information is of value.

10 Recommendations

These are the recommendations for the Uppsala County Council. Most of them are explained in greater detail in the summary (section 8).

10.1 Education and Information

- Subtle reminders that a user's actions in COSMIC are being logged.
- Coordinate the education to make sure that all COSMIC users get the same training, including different examples on actions being logged, and applicable laws regulating the use of the system.
- Run campaigns about the privacy, ethical, and legal issues surrounding EHR. This will increase the knowledge and awareness among the staff. The campaigns could consist of simple e-mails, buttons, or pins.
- Permanent posters raising awareness of privacy issues.
- A "hotline" (preferably at the same phone number where technical COSMIC questions are being answered) that a member of the staff can call with ethical and legal questions.
- Inform the staff when someone has been caught misusing the system.
- Encourage informal discussions (e.g. at the coffee machine) about ethical questions.

10.2 Technical Solutions

- Use Smart Cards to log into COSMIC and integrate them with pass cards. It is important that the user is logged out when the card is removed.
- Allow searching for EHRs only by the patients' personal numbers and not by their names or parts thereof, unless it is an emergency.
- Healthcare Provision, Stockholm County (SLSO) has developed a log analyzing tool, SALA, which is:
  - User friendly
  - Developed close to hospital personnel
  - Stand alone from the system handling electronic healthcare records
  - Took less than one year of development time

  Based on the arguments above, the Uppsala County Council is recommended to investigate the possibilities of a cooperation with Stockholm County Council. Regardless if a cooperation takes place we suggest the Uppsala County Council to create a system similar to SALA.

- Among the different techniques presented in this report, association analysis and clustering algorithms has proved to be beneficial with regards to easiness to implement and efficiency. When commencing the development of the log analysis system, a panel of researchers ought to be assigned to determine which technique that would be best suited for COSMIC.

11 Acknowledgements

We would like to give special thanks to the following people that helped us throughout the research for this report. We value your useful thoughts, opinions and ideas. Thanks for spending your valuable time with us.

Magnus Bergström
Hans Blomberg
Åsa Cajander
Annette Cederberg
Mats Daniels
Benny Eklund
Olle Gällmo
Rebecka Janols
Iordanis Kavathatzopoulos
Anita Lakström
Cary Laxer
Johan Lindqvist
Cecilia Lundberg
Håkan Nordgren
Johanna Penell
Kjell Osborn
Roger Lindblom
Bengt Sandblad
Bo Wikström
Erik Zeitler

12 Glossary

In the end a novice should read the report through and highlight hard words which then should be explained.

AHIMA	American Health Information Management Association
AHIP	America's Health Insurance Plans
AHRQ	Agency for Healthcare Research and Quality
AMIA	American Medial Informatics Association
BIF	Bastjänster för informationsförsörjning (Basic Services for Information Management)
CPR	Computer-based patient record
EHR	Electronic Healthcare Record
ePHR	Electronic Personal Health Record
EU	European Union
HIE	Health Information Exchange
HIMSS	Healthcare Information and Management Systems Society
HIPAA	Health Insurance Portability and Accountability Act of 1996
LAVA	Log Anomaly Visualization Application
NHIN	The Nationwide Health Information Network
OCR	Office for Civil Rights
PSQIA	The Patient Safety and Quality Improvement Act of 2005
PSO	Patient Safety Organization
SALA	Systematisk Automatiserad LoggAnalys (Systematic Automated Log Analysis)
SLSO	Stockholms läns sjukvårdsområde (Healthcare Provision, Stockholm County)

13 References

[1] Araghi, L. F., Khaloozade, H., and Arvan, M. R. Ship Identification Using Probabilistic Neural Networks (PNN). In *Proceedings of IMECS'09* (Mar. 2009), International MultiConference of Engineers and Computer Scientists, pp. 1291–1294.

[2] Ball, M. J., Smith, C., and Bakalar, R. S. Personal health records: Empowering consumers. *Journal of Healthcare Information Management 21*, 1 (2007), 76–86.

[3] Bishop, C. Novelty detection and neural network validation. *IEE Proceedings: Vision, Image and Signal Processing 141*, 4 (Aug. 1994), 217–222.

[4] Bogacz, R., Brown, M. W., and Giraud-Carrier, C. High capacity neural networks for familiarity discrimination. In *Proceedings of ICANN99* (Sept. 1999), pp. 773–778.

[5] Cajander, Å. *Usability - Who Cares? : The Introduction of User-Centred Systems Design in Organisations*. PhD thesis, Uppsala University, Division of Human-Computer Interaction, 2010.

[6] California HealthCare Foundation. National consumer health privacy survey 2005. `http://www.chcf.org/~/media/Files/PDF/C/PDF%20ConsumerPrivacy2005Slides.pdf` visited on October 25th 2010, Nov. 2005.

[7] Cantrill, S. V. Computers in patient care: The promise and the challenge. *Communications of the ACM 53*, 9 (2010), 42–47.

[8] CompuGroup Medical. Takecare developer. `http://www.ne.compugroupmedical.com`, visited on December, 2010.

[9] Elberfeld, M., and Textor, J. Negative selection algorithms on strings with efficient training and linear-time classification. *Theoretical Computer Science In Press, Corrected Proof* (2010). doi:10.1016/j.tcs.2010.09.022.

[10] European Union. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal L 281* (Nov. 1995), 31–50.

[11] Export.gov. U.S. – European Union Safe Harbor. `http://www.export.gov/safeharbor/eu/eg_main_018365.asp` visited on November 6th 2010.

[12] Gaunt, N. Practical approaches to creating a security culture. *International Journal of Medical Informatics 60*, 2 (2000), 151–157.

[13] Gostin, L. Health care information and the protection of personal privacy: Ethical and legal considerations. *Annals of Internal Medicine 127*, 8 Pt 2 (Oct. 1997), 683–690.

[14] Hang, X., and Dai, H. Constructing detectors in schema complementary space for anomaly detection. In *Proceedings of GECCO (1)'04* (2004), GECCO 2004, LNCS 3102, pp. 275–286.

[15] Iakovidis, I. Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in europe. *International Journal of Medical Informatics 52*, 1-3 (1998), 105–115.

[16] Jaksa, M. B., Shahin, M. A., and Maier, H. R. State of the art of artificial neural networks in geotechnical engineering. *Electronic Journal of Geotechnical Engineering* (2008), 9.

[17] Katsikas, S. K. Health care management and information systems security: awareness, training or education? *International Journal of Medical Informatics 60* (2000), 129–135.

[18] Markou, M., and Singh, S. Novelty detection: a review part 2: neural network based approaches. *Signal Processing 83*, 12 (2003), 2499–2521.

[19] Marsland, S. Novelty detection in learning systems. *Neural Computing Surveys 3* (2003), 157–195.

[20] Massachusetts Institute of Technology. The oath by hippocrates. `http://classics.mit.edu/Hippocrates/hippooath.html` visited on November 10th 2010, 2009.

[21] McGraw, D. Privacy and health information technology. *Journal of Law, Medicine & Ethics 37*, Sep (2009), 123–149.

[22] Miller, C., Cromwell, T., and Steffensen, S. CONNECTing to the NHIN. *Federal Health Architecture* (2008).

[23] Nationwide Health Information Network (NHIN). Exchange architecture overview. `http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_11113_911643_0_0_18/NHIN_Architecture_Overview_Draft_20100421.pdf` visited on December 2010, 2010.

[24] North Carolina Healthcare Information and Communications Alliance. NHIN Specification Preview. `http://www.nchica.org/HIT_HIE/NHIN2/NHIN1209.htm#S2`, visited on December, 2010.

[25] Öhlander, S. Sjuksköterska läste ex-flickväns journal. *Gefle Dagblad* (Apr. 2010).

[26] Parra, C. A., Iftekharuddin, K., and Kozma, R. Automated brain data segmentation and pattern recognition using ANN. In *Proceedings of CIRAS'03* (Dec. 2003), Second International Conference on Computational Intelligence, Robotics and Autonomous Systems.

[27] Paulett, J. Lava: An ehr log analysis visualization application. Master's thesis, Vanderbilt University Department of Biomedical Informatics, 2008.

[28] Pettersson, L. Å. Riktlinjer för informationssäkerhet, sammanfattning för medarbetare. http://www.lio.se/upload/75406/Riktlinjer_kortversion.pdf visited on November 8th 2010, 2009.

[29] Poromaa, Å. Böter för intrång i patientjournal. *Norrbottens-Kuriren* (Sept. 2009).

[30] Riley, D. L. Authentication in the NHIN. http://www.docstoc.com/docs/52246497/Authentication-in-the-NHIN, visited on December, 2010, January 2010.

[31] Rishel, W., Handler, T. J., and Edwards, J. A clear definition of the electronic health record. Tech. rep., Gartner, 2005.

[32] Sankar, P., and Jones, N. L. To tell or not to tell: Primary care patients' disclosure deliberations. *Archives of Internal Medicine 165*, 20 (2005), 2378–2383.

[33] Shah, S. N. An overview of nhin and nhin direct for software developers. http://www.ibm.com/developerworks/web/library/wa-nhindirect/index.html, viewed on December, 2010, 2010.

[34] Shimanek, A. E. Do you want milk with those cookies? complying with the safe harbor privacy principles. *Journal of Corporation Law 26*, 2 (Jan. 2001).

[35] Suchman, L. A. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, New York, NY, USA, 1987.

[36] Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to Data Mining (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[37] the Direct Project. NHIN Exhange Architecture. http://wiki.directproject.org/NHIN+Exchange+Architecture, visited on December, 2010.

[38] The Swedish Data Inspection Board. Säkerhet för personuppgifter. http://www.datainspektionen.se/Documents/faktabroschyr-allmannarad-sakerhet.pdf visited on November 11th 2010, 2008.

[39] Tidningarnas Telegrambyrå. Böter för att ha läst Lindhs journal. *Dagens Nyheter* (July 2004).

[40] Turkington, R. Medical record confidentiality law, scientific research, and data collection in the information age. *Journal of Law Medicine & Ethics 25*, 2-3 (1997), 113–129.

[41] U.S. Department of Health & Human Services. The health insurance portability and accountability act. http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.htmlvisited on August 30th 2010.

[42] U.S. Department of Health & Human Services. Nationwide Health Information Network (NHIN): Resources. http://healthit.hhs.gov/portal/server.pt?open=512&objID=1194&parentname=CommunityPage&parentid=4&mode=2&in_hi_userid=10741&cached=true visited on December, 2010.

[43] U.S. Department of Health & Human Services. The patient safety and quality improvement act. http://www.hhs.gov/ocr/privacy/psa/regulation/index.htmlvisited on August 30th 2010.

[44] World Medical Association. Wma declaration of helsinki - ethical principles for medical research involving human subjects. http://www.wma.net/en/30publications/10policies/b3/index.htmlvisited on December 06th 2010.

[45] World Psychiatric Association. Declaration of hawaii. *Journal of medical ethics 4*, 2 (June 1978), 71–73.

[46] Wynia, M., and Dunn, K. Dreams and nightmares: Practical and ethical issues for patients and physicians using personal health records. *The Journal of Law, Medicine & Ethics 38*, 1 (2010), 64–73.

[47] Yue, Y., Ding, L., Ahmet, K., Painter, J., and Walters, M. Study of neural network techniques for computer integrated manufacturing. 147.

Appendices

A Law

A.1 Legislation in Sweden

There is a formal law controlling the use of Electronic Healthcare Records (EHR) in Sweden: the "Patient Data Act (SFS 2008:355)" which is described in this subsection.

A.1.1 The Patient Data Act

The main purpose of the Patient Data Act is to allow electronic storage and unified access of healthcare records for medical staff and units. The main topics are:

- Enable electronic storage and unified access of healthcare records.
- The healthcare provider is responsible for controlling access and maintaining access history.
- A patient can request blocking of access for certain staff or units to their healthcare record.
- The healthcare provider can give the patient electronic access to their own healthcare records.

Secrecy
The rule of thumb used here is that medical staff should only be able to access the data they need to finish the task at hand. This should be enforced by controlling access and setting up proper guidelines for all employees. Patients can restrict access for medical staff and units. The healthcare provider is responsible for defining these areas. Medical staff should be able to see if there is restricted or unrestricted data available for a patient without accessing any personal data.

Accessing Data
To access data the following three criteria have to be met: (1) the user has a current medical relationship to the patient, (2) the data can be assumed to be used to prevent or treat illnesses or injuries and (3) the patient authorizes access of data registered by other medical personnel. All accesses of data should be preceded by an active choice by the medical staff. This means that the user has to evaluate the situation and make a decision whether the information is needed to finish the task. If the user chooses to access the data, the access will be logged. In order to access blocked data the patient has to give his or her permission. If the patient is unable to give permission (i.e. the patient is unconscious and in dire need of care) the user should be able to access the blocked data by making another active choice. This access should also be logged.

Controlling Access
The healthcare provider is responsible for restricting access to what personnel need to finish the tasks at hand. Access levels should be followed up and corrected whenever a staff member's position changes. The healthcare provider should also document electronic access to patient records and systematically check the access history of patient records.

Patient Access to Personal Data
The healthcare provider is able, but not forced, to allow access of patients' own records electronically (i.e. over the Internet). The healthcare provider has to evaluate which data is appropriate to distribute in this manner. The healthcare provider also has to take appropriate security measures, such as identification through electronic ID and encryption protocols for transfer of data. The healthcare provider should, when requested by a patient, inform the patient of all access to his or her EHR. The provider has to make this information available on paper and is allowed, but not forced, to allow direct access to the raw data in the database concerning these accesses. The direct access has to follow the same security measures as the access to the patient's healthcare records. Patients are given the right to view the unit and the time of the access, the right to see which person was responsible is not granted by the law. The information has to be formatted in such a way that the patient can evaluate whether or not the access was appropriate.

Supervision
The Data Inspection Board is responsible for supervising the handling of patient data records. This includes ensuring that appropriate security measures are taken. The National Board of Health and Welfare is responsible for the content of the healthcare records.

A.2 Legislation in the European Union

The main law in this field is the data protection directive (95/46/EC). The directive came into effect in October 1995 to protect the individual in the handling of personal data. Every European Union member state had the directive implemented by end of 1998 in their own data protection legislation. The Swedish implementation is the Personal Data Act. The principle of the directive is that personal data shall not be processed at all, except when certain conditions are met, in short based on following seven principles:

1. Notice - data subjects shall be given notice when their data is gathered.
2. Purpose - data shall only be used for the stated purpose and not for any other purposes, for example selling data to third party companies.
3. Consent - data shall not be distributed without the data subject's permission.
4. Security - gathered data shall be kept secure from any potential abuses.
5. Disclosure - data subjects should be informed of who is using their data.
6. Access - data subjects shall be allowed to access their data to correct any inaccurate data.
7. Accountability - data subjects shall have a method available to hold data collectors accountable for following the above principles.[10]

There has been tension between the European Union and the United States about handling of personal data since the United States has a lower level of data protection. The European Union has negotiated with the United States representatives about the protection of personal data, the Safe Harbour Principles. It is designed to protect personal data from being transferred to nations that do not meet the European standard for privacy protection.[11][34]

A.3 Legislation in the USA

In the United States there are two laws concerning the handling of Electronic Healthcare Records: the "Health Insurance Portability and Accountability Act (P.L.104-191)" and the "Patient Safety and Quality Improvement Act (P.L.109-41)" that will be described in this subsection.

A.3.1 Health Insurance Portability and Accountability Act

The main purpose of the Health Insurance Portability and Accountability Act (HIPAA) is federal protection that covers health care entities and patients. The health care entities must be aware about the rights, respect and limitations of the patients records.[41]
It's main topics are:

- Health care entities receive training in ethics, confidentiality, integrity, and they are aware about how to handle patients' information.
- Patients must agree before showing their own information to health care centers.
- The law restricts who is allowed to handle patients' information. Marketing and statistical expertise should not be allowed access without the patients' consent.
- Patients have the right to get information such as: Who accessed the information and when was the information accessed.
- Health care entities are responsible for protecting the patients' information and to limit the amount of information shown to the nurses and doctors.

A.3.2 Patient Safety and Quality Improvement Act

The main purpose of the Patient Safety and Quality Improvement Act (PSQIA) is to provide a voluntary reporting system that assures the care of the privacy and patient information data.[43] It's main topics are:

- Data reporting system designed to show the patients' data available and protects information called as "patient safety work product".
- The violations are punished with monetary penalties.
- This reporting system is checked by two organizations:
  - Office for Civil Rights (OCR) : finds patterns, interprets and implements privacy protections.
  - Agency for Healthcare Research and Quality (AHRQ): update the list of patients at safety organizations from the information collected by the Patient Safety Organizations (PSOs).

B Other Systems

B.1 The Nationwide Health Information Network (NHIN)

The Nationwide Health Information Network (NHIN) is a "network of networks" that allows connection and information exchange between different health organizations. It is not an actual system or program, nor does it have a national database storing any of the patients' files. Rather, it is a set of policies, services, and standards which provide a framework upon which health information can be shared securely over the internet.[42] The NHIN is like a road connecting different buildings; the road provides a way to connect and pass along items, but does not allow the storage of these items.

The NHIN was developed as a solution to connecting electronic healthcare records in the United States. Because universal health care does not exist in the United States, each hospital organization and medical provider could use a different system or format for their records. They may even have specialized programs to view an electronic healthcare record that other health care providers cannot have access to. This variety of practices makes regulating and unifying these records a daunting challenge. One cannot simply ask every hospital to change to a new system; they may already be comfortable with their current one, and changing would take too much precious time. The NHIN, however, does not touch the specific structure, but provides the infrastructure of data sharing.

B.1.1 Basic Architecture

The NHIN is fundamentally a "networks of networks", composed of nodes and links connecting them. The nodes are health information organizations which exchange health information with other NHIN nodes using the NHIN Gateway.[23, pg 6]

Architecture Principles

Highly distributed - there is no central data repository, but patient health information is retained at the local level.

Local Autonomy - the decision of whether information should be distributed to another node in the NHIN is decided on the local level. NHIN transactions must include enough information for the potentially receiving organization to make a decision about participating in the information exchange.[37]

Use of Web Services - As stated earlier, NHIN is not a separate network, but is a set of protocols that run on the existing internet infrastructure. It relies on the internet for transporting, discovering, and exchanging information.[37]

B.1.2 Security of NHIN

There are many items built into the NHIN to ensure a safe, private, and secure network. These include

- Membership services that allow only valid, trusted entities to be involved in the network,
- Digital certificates to validate a user's identity,
- Data Use Reciprocal Support Agreement which informs users about the extent of their privacy and security within the NHIN,
- A secure, private Internet connection for information exchange

[42]

Ensuring Authorization What personnel are authorized to view what level of information on the NHIN is determined on the local level, as was explained in one of the principles of the system. The level of authentication may vary depending on the type of information involved in the exchange.[30] The authorization framework is based on SAML (Security Assertion Markup Language), which provides authentication and authorization between security domains.[24]

Digital Certificates

- Purpose – Certificates are necessary to verify a user's identity and encrypts communication between users.
- Types
  - Development certificate: The development certificate is used for unit and integration testing. The individual organizations must validate these certificates using either a self-signed certificate - the certificate's creator signs and approves it - or a free certificate - acquired through other organizations and provide a limited level of security.
  - Validation certificate: The validation certificate is used from NHIN validation testing.
  - Production certificate: The production certificate is needed for access onto the NHIN exchange.

  Validation and production certificates can be provided by the Office of the National Coordinator for Health IT. These certificates are only valid for a one-year period; after that time, they have to be reissued. This ensures the security of the certificate.

B.1.3 Connect – a System that Uses the NHIN

CONNECT is an open-source system currently available for download that allows a hospital to create their own Health Information Exchange. With this, they can then connect to the NHIN and share their information with other health care providers.

Component Make-up CONNECT is composed of three parts: the Core Service Gateway, the Enterprise Service Component, and the Universal Client Framework.

- The Core Services Gateway implements the core services defined by the NHIN. It allows the organization to locate patients within other organizations, and, if allowed by patient settings, can receive patient documents from other organizations. Other features include authenticating the network participants.[22]
- The Enterprise Service Components provides the tools necessary for electronic health information exchange, including indexing patient identities, maintaining patient health documents, and specifying the rules for the release of medical information, managing consumer preferences, and other functions. Initially, there are default settings for these service components, but individual hospitals are free to replace them with their own implementations.[33]
- The Universal Client Framework is a set of applications so developers can implement enterprise service components.

[22]

B.2 Systematisk Automatiserad LoggAnalys (SALA)

B.2.1 Introduction

Systematisk Automatiserad LoggAnalys (SALA) or in English, Systematic Automated Log Analysis, is a software system developed at Healthcare Provision, Stockholm County (SLSO). SLSO is part of Stockholm County Council and has approximately 11,000 employees at 800 units. The information presented in this section about SALA is mainly based on a presentation given at SLSO by the project group responsible for its development.

The SALA system started off as an idea from an employee at SLSO and since no off-the-shelf software existed that corresponded to that idea, the decision was taken to develop the system at SLSO. The initial development took place September 2009 until April 2010, when it went into production, by SLSO staff with the aid of two external consultants.

The system has a web-based interface and is used as a tool for analyzing logs from TakeCare, an Electronic Healthcare Record (EHR) system developed by CompuGroup Medical [8] (formerly known as Profdoc Care AB). TakeCare is used at SLSO and many other health care providers in Stockholm County Council. SALA was developed in order to facilitate the process of manually reviewing randomly selected logs of TakeCare users. The reviewing needs to be done in order to comply with the Swedish patient data law. The law states that logs should be randomly sampled and reviewed systematically in an entirely or partly automated process.

The success factor for SALA can be summarized in to the following: Development was made very close to the units and the directors that was going to use SALA. The initiative and knowledge was in-house; SALA was developed at SLSO based on their own ideas and mainly by their own staff. One concern is that the Data Inspection Board has not yet reviewed SALA, and so may find it inadequate.

B.2.2 How does SALA work?

Each month 10% of all employees at each SLSO-unit are randomly sampled and their EHR accesses are analyzed by SALA. The log entries associated with these employees are analyzed using the following set of criteria:

- Patient age (if the employee works with adult patients, but opens an EHR belonging to a child)
- Opening of his/her own EHR
- Opening an EHR belonging to a colleague
- Opening an EHR belonging to a patient that is not treated at his/her unit or clinic within a certain period of time
- Opening an EHR belonging to a patient that the employee has not written anything in within a certain period of time

The logs are then put under scrutiny by the director at each unit. This is important because specific knowledge about employees has to be taken into consideration when reviewing the logs. SALA suggests actions to the director based on the patterns derived from the analysis criteria. With support from the suggestions generated by SALA the director can swiftly sign the logs with normal patterns as okay. However, if SALA detects an atypical pattern, no suggestion is made by SALA and the director is forced to make an active decision on how to deal with that specific case.

The standard way for the director to deal with these situations is to ask the employee to give an explanation for his/her suspected unauthorized access of a patient's EHR. If a satisfactory explanation is given, the director marks the suspected log entry as approved. If an unsatisfactory explanation is given, the director marks the log entry to be reported for further investigation.

B.3 Swedish National Police Board - Log Management System

In 2003, the Swedish Foreign Minister Anna Lindh was murdered in Stockholm. The investigation of this murder led to about 250 internal investigations of the Swedish Police for various reasons. Only five of these investigations proved useful, and the rest were cancelled. One of the reasons for this high number of "false alarms" was that over half a million logs had to be analysed manually by the investigators. An automatic log management system could have prevented this high number of internal investigations. Since 2003, a new log management system has been implemented with great success at the Swedish National Police Board.

B.3.1 Introduction

A presentation of the new log management system used at the Swedish National Police Board was held by Roger Lindblom at Uppsala University. He was invited by the project group to hold a presentation due to his long experience and deep knowledge in the area of Log Management.

Lindblom is currently working at the Security Operating Center (SOC) at the Swedish National Police Board in Stockholm. For the past six years he has been working with log management. He has a technical background with studies in Computer System Science at Stockholm University. Since then he has been working as a policeman before he got involved with log management. Being a former policeman, Lindblom has acquired deep knowledge about his organization. Lindblom argues that this knowledge is particularly important when building an effective log management system. He calls this knowledge business requirements and this is explained in more detail below.

The following sections contain summaries of Lindblom's key arguments, and recommendations for a successful log management system.

B.3.2 Business Requirements

To implement a log management system successfully, deep knowledge about the organization in which the system will function is of the essence. Roger calls this deep knowledge "business requirements". Stating the business requirements is an important first step to determine what kind of information is of interest; *what* do we really want to know? The business requirements should dictate why and what the log system should log.

In a large and complex business environment, where a large number of logs must be collected from many different sources, it is important to remember that even weak logs can be used, if supported by other possibly weak logs, there are still possibilities to get out *enough* information. Everything does not have to be perfect to get good information. It depends on what the business requirements say.

B.3.3 Log Management

The log management at the Swedish National Police Board can be described in a number of different steps.

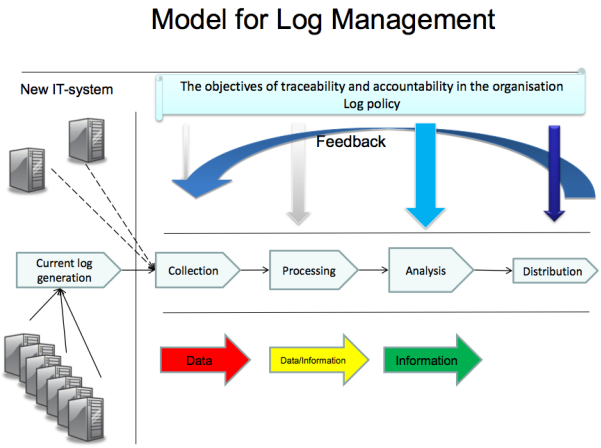

Figure 7: Log Management Model

As shown in figure 7 - which also shows the infrastructure - the information sources are in-house. Gathering the information is thus not a problem. The important part is how to make use of this information. If one doesn't analyze the information in the appropriate context, it will be more or less useless. One key importance is that information is never to be tampered with during any of the different steps in the model. Traceability and accountability is of paramount importance to be able to bind the right person to a specific operation. The information is of little use if it can't be bound to a person.

Another important step of the model is the feedback process, which needs to be done in order to improve the system over time. The log management system at the Swedish National Police Board is rule based, which means that it will react if a rule is violated. It is important to learn from the analysis in order to create new rules.

There are a number of sources for requirements for log management.

- Internal regulations: For instance, at the Swedish Police, logging should be done for *each access* to classified information. This will affect the rules and requirements of the system.
- Laws: Most organizations should strive to have business requirements that are in accordance with the law.
- Business requirements: This is highly depending on the knowledge of the organization and the working methods. The more knowledge one has, the more requirements will be translated into rules for the system to use.

The system running at the Swedish National Police Board today is rule-based and performs real-time monitoring. This means that the system will react if a rule is violated. New rules can be defined by administrators if new requirements arise. This solution is based on the requirements that were found after extensive mapping of the organization. Today, the system is using log data to protect the organization's information, and to prevent or minimize further information leakage.

B.4 Hippocratic Database

B.4.1 Main Objective and structure

The issue of breached privacy that is being addressed by Uppsala Hospital is an iteration of a problem that has become very prevalent in database systems universally. A solution proposed by IBM's Rakesh Agrawal is called "Hippocratic Databases". Like its name suggests, Hippocratic Databases are structured to preserve the desired privacy of the client who provides their information. These clients or data donors all have variable levels of privacy that they want to preserve. As a result, the database system must consider these differences and consequently review its information to ensure its own compliance.

B.4.2 Other Systems

The Hippocratic Database is not the first system to implement structural security for its data donors. For example, Statistical Databases attempt to provide statistical information of its information while maintaining the anonymity of its original donors. Another example are Secure Databases which use security levels to determine who reads and writes files.

B.4.3 Ten Principles that Guide Architecture

Because of its original intention, Hippocratic Databases are defined by ten principles that encapsulate how the system should maintain its privacy. The first principle is *Purpose Specification* which states that all information gathered must be linked with the purpose that it is used for. Second, the information gathered must have been given with the *Consent* of the donor. Thirdly, the system should collect the minimal collection *(Limited Collection)* of information to accomplish its purpose. In conjunction with Limited Collection, the database must ensure *Limited Use* of information by ensuring that queries are consistent with the purposes associated with the information and *Limited Disclosure* by ensuring that the information is only available to parties permitted by the donor. To minimize the possibility of unwanted access even further the database has *Limited Retention* of information. Like any database system, the *Accuracy* of Hippocratic Databases must ensure correct and up-to-date information. Regardless of structural security within the database, the information within the database must have *Safety* from misappropriations. Lastly, the *Openess* of the system should allow the original donor "[to] be able to access all information about [themselves] stored in the database" and ensure system *Compliance* with the previous principles listed.

B.4.4 Basic Architecture and Implications

The architecture that enforces the privacy principles previously outlined can be categorized into four components. The first component is a policy definition

which captures the basic rules set out to enforce limited disclosure. These definitions are stored in the database by meta-data. As a result, when queries are made the database checks the application context that the queries are used in and will return only pertinent information through the query modifier. Lastly, the queries structure of the database is implemented through a disclosure model that holds the relationships between the data being accessed, user, and allowances.

B.4.5 Implementation and Results

The implementation of a Hippocratic database was measured to examine several major factors that could pose potential problems in the elapsed time of queries. Of the tests performed, the most interesting results were the scalability and the impact of filtering. In the experiments conducted to evaluate the effects of scalability the elapsed time to access 15 million records across multiple modified external databases was 200 seconds which is extremely impressive considering the size of information it has to process. Furthermore, the Hippocratic database seemed to reduce the amount of elapsed time for a query if the selection was smaller than 60% of all records.

B.5 Bastjänster för informationsförsörjning (BIF)

Bastjänster för informationsförsörjning (BIF) is a work in progress by Center för eHälsa i samverkan (CeHis) and is meant to be a national EHR security infrastructure for Swedish health care givers. BIF contains a set of nine distinct services used to secure the integrity of patients:

1. **Authentication** - Identification through use of Smart Card and pin code.
2. **Access control** - Checks whether a given person is allowed access to a given patient record.
3. **Logging** - Logs security related events from different systems.
4. **Log analysis** - Tool for analyzing the logs to discover improper record access.
5. **Notification** - Delivering messages to users or systems.
6. **Patient approval** - Checks whether the patient has approved or blocked access to the health record.
7. **Patient relationship** - Checks whether a given person has a patient relationship with a given patient.
8. **Dispensation** - Electronic transfer of health record.
9. **Secure patient context** - Makes sure that the user always has the right patient's information available.

The ultimate goal of the BIF services is that they will replace functions of the EHR systems used around the country, ensuring that all EHR systems follow the Swedish Patient Data Act. This means that the systems have to be modified in order to make use of the services. Whether or not this is a good solution is outside the scope of this report and is left for others to decide.

The rest of this section will focus on the services for logging (service 3) and log analysis (4), since those are of main interest for this report. The information about these services is mainly gathered from the BIF System Requirements Specification (2007) and the BIF Web Interface user manual (2010).

B.5.1 Log service

The log service handles and stores security related logs from the EHR system and other BIF services. Logs are electronically signed and securely transferred to ensure the data integrity.

To increase the performance and the responsiveness of the EHR systems, a local log agent receives the logs, puts them together in batches and puts a timestamp on them before sending them to the log service. The log agent has a configurable amount of local storage, meaning that if the log service for some

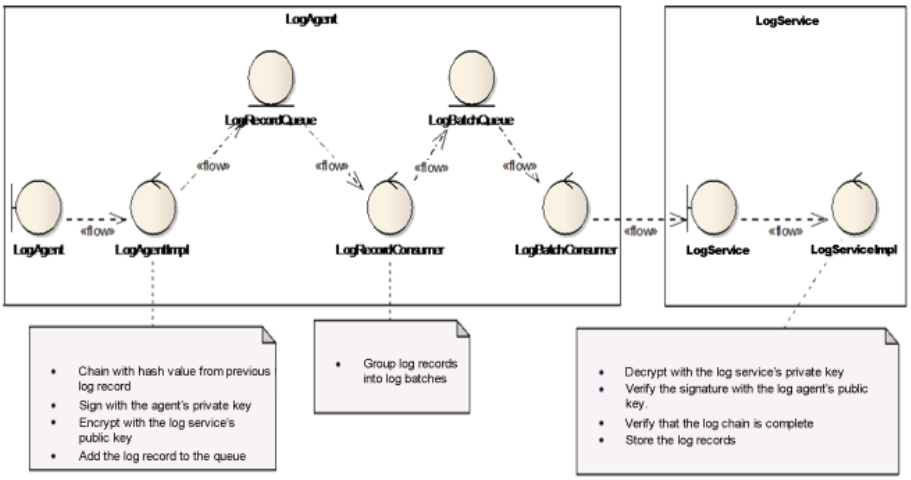

Figure 8: Flow scheme of the logging.

reason is unreachable, logs are put in a local queue and can be transferred when the log service reconnects.

Logs are stored in the XML format and the schema for a certain log is based on the configuration of the category it belongs to. The category configuration specifies a set of required log fields and a set of optional fields. The log service checks each log against the XML schema to make sure that the data is correct.

Access to the logs is only possible through the log analysis service.

B.5.2 Log analysis service

The log analysis service provides a tool for searching and following up on the logs that have been gathered by the log service. Access to the logs is controlled by the authentication service (1). The log analysis service provides a set of predefined analyses that can be run manually or on regular basis, as well as the possibility to create custom analyses. An analysis consists of multiple steps, each running a log query or another analysis. The last step can be used to format the result into something readable.

A log query takes in one or more parameters and returns an output value. When chaining queries in an analysis, the output value from a previous query is used as input parameter to the next. There are predefined queries that perform common operations, one of which transforms raw log data into a readable PDF file.

C Legal Contract for Medical Staff

Landstinget i Uppsala län

Förbindelse om användande av landstingets vårdregister

Förnamn Efternamn

Användar-ID Personnummer

Behörighetsperiod: från och med (åå-mm-dd) till och med (åå-mm-dd)

(om inget slutdatum anges gäller behörigheten tills vidare)

Tillgång till vårdinformation: Som användare i landstingets vårdregister, t ex Cambio COSMIC, kommer du att ha tillgång till uppgifter som rör patienters vård och behandling. Detta är förenat med ett personligt ansvar som regleras bland annat av Sekretesslagen (1980:100), Personuppgiftslagen (1998:204), Patientjournallagen (1985:562), Vårdregisterlagen (1998:544) samt Brottsbalken (1962:700).
Direktåtkomst: Direktåtkomst till uppgifter i ett vårdregister får endast den ha som behöver tillgång till uppgifterna för att kunna utföra sitt arbete[1]. Vidare skall uppgifterna behövas för något av de ändamål för vilka ett vårdregister får användas, 3 och 4 §§ Vårdregisterlagen. Åtkomsten får inte avse andra uppgifter än vad som behövs för att arbetsuppgifterna skall kunna utföras. Direktåtkomst får du genom att du tilldelats behörighet till landstingets vårdregister.

Jag är medveten om:

- att alla registreringar jag gör i ett vårdregister är spårbara och kopplas till min användaridentitet.
- att jag alltid ska logga ut mig från vårdregistret när jag lämnar datorn.
- att jag ansvarar för att skydda mitt lösenord väl och inte avslöja det för andra.
- att uppföljningar av händelser i landstingets vårdregister sker regelbundet.
- att en överträdelse av ovan angivna bestämmelser kan leda till polisanmälan.

Jag har full insikt i ovanstående ansvarsförhållanden och förbinder mig härmed att följa de regler som gäller för hantering av patientinformation i landstingets vårdregister.

Jag bekräftar att jag tagit del av ovanstående information

Uppsala den /

.....

LAGRUM

Brottsbalken 1962:700
Dataintrång
4 kap. §9c Den som i annat fall än som sägs i 8 och 9 §§ olovligen bereder sig tillgång till en uppgift som är avsedd för automatiserad behandling eller olovligen ändrar, utplånar, blockerar eller i register för in en sådan uppgift döms för dataintrång till böter eller fängelse i högst två år. Detsamma gäller den som olovligen genom någon annan liknande åtgärd allvarligt stör eller hindrar användningen av en sådan uppgift.

20 kap. 3§ Röjer någon uppgift, som han är pliktig att hemlighålla enligt lag eller annan författning eller enligt förordnande eller förbehåll som har meddelats med stöd av lag eller annan författning, eller utnyttjar han olovligen sådan hemlighet, dömes, om ej gärningen eljest är särskilt belagd med straff, för brott mot tystnadsplikt till böter eller fängelse i högst ett år.
Den som av oaktsamhet begår gärning som avses i första stycket, dömes till böter. I ringa fall skall dock ej dömas till ansvar.

Lag (1998:544) om vårdregister
Ändamål
3 § Personuppgifter i ett vårdregister för behandlas för dokumentation av vården av patienter eller för sådan administration som rör patienter och som syftar till att bereda vård i enskilda fall.
Behandling av personuppgifter får även utföras för den ekonomiadministration som föranleds av vård i enskilda fall.

4 § Personuppgifter i ett vårdregister får, utöver vad som anges i 3 §, behandlas för följande ändamål:
1. framställning av statistik,
2. Uppföljning, utvärdering, kvalitetssäkring och administration på verksamhetsområdet, och
3. Uppgiftsutlämnande som föreskrivs i lag eller förordning.

Direktåtkomst
8 § Endast den som för de ändamål som anges i 3 och 4 §§ behöver tillgång till uppgifterna för att kunna utföra sitt arbete får ha direktåtkomst till uppgifter i ett vårdregister. Åtkomsten får endast avse de uppgifter som behövs för arbetets utförande.
Med arbete menas i detta fall även deltagande i sjukvårdens arbete för olika studenter.

D Authors

Tim Eriksson
Mikael Gerhardsson
Andreas Göz
Jens Hammarlund
Alexander Ingvar
Henrik Jakobsson
Magnus Jonsson
Peter Jönsson
Lindsay Klaetsch
Wilson Kurian
Anders Lisspers
Erik Löthman
Niklas Malmgren
Luisa Fernanda Angarita Moreno
Michael Pridal-LoPiccolo
Kristian Samuelsson
Ann Say
Mattias Sjöström
Torbjörn Svangård
Linus Wollentz