# Proposed Data Loss Protection in Electronic Health Record

**Abeer Tariq Maolood, Ph.D(Asst.Prof)**[*] **Rasha Mohammed Mohsen**[*]

**Abstract:** In this paper, proposed data loss protection (DLP) solutions for improving privacy of electronic Health record (EHR). Prevention method is the first phase in DLP system, it used signature-based to provide blocking the well-known attacks, also to improve system accuracy by reducing the number of alerts. On the other hand, detection phase has two levels of detection, online detection, and offline detection. Online detection by using novelty detection used to detect new attacks arrives in the system, while the offline detection detects abnormal user behavior in a period of time by using supervised detection.

**Keywords:** prevention, novelty detection, supervised detection, signature based, whitelisting, decision tree (ID3).

[*] Department of Computer Sciences, University of Technology, Baghdad, Iraq

## 1. Introduction

Nowadays, data security is important, the patient privacy and EHR security are to require security algorithms for improved data security policy. Data loss is a major problem in EHR privacy, especially, the EHR involves sensitive information and any loss of data caused many problems in the patient's privacy.

Data loss protection (DLP) is one of the best solutions to improve EHR security, while it focuses on identifying sensitive data, and it Looks at the content that is considered critical[1, 2] .DLP is providing protect sensitive data from both external or internal threats, also it automatically making sure that sensitive data in EHR not stored, deleted, sent or accessed where should not be in both intentionally or/and unintentionally. The DLP solution contains two models of either allowed prevention methods, or/and detection method, these models can identify, either by expert’s knowledge or by past transactions [3]. Prevention method focuses on external threats and provides blocking of unauthorized (well-known attacks) .In the other hand, detection method focuses on internal threats and provides detect of misused by authorized and detect new attacks

## 2. Related Work

This section discusses important previous works of data loss protection(DLP) solution and cover some of the literature surveys forces in healthcare security:

Bradley MSteve N, John P,[2011], Propose access control system based on a data-driven methodology to generate policies automatically from EHR access logs by using a two-step , network construction, and association rule discovery. First, transforms each transaction to graph, second, convert the graph into rules. This proposed system provides afford great stability in the system and protects from external threats ,but cannot protect from insider threats[4].

Santos R, Bernardino J, Vieira M, Rasteiro D.,2014, developed system based on combined between syntax-centric and results in centric approaches. The author represents the Normal profiles in the statistical distribution, which a query is considered anomalous if does not match with the original probability distribution[5].

Costante E, Etalle S, Fauri D, Hartog J, Zannone N,2016, proposes a hybrid DLP that combines signature-based and anomaly-based, uses an engine that automatically learns the normal user behavior, Typically, this solution automatically builds and update signatures based to block undesired transactions before any damage[6].

Al-Hamdani W,[2016], developed Cryptography access control system for healthcare information systems, by integrating role-based access with cryptography access control,This model is suitable for the client-server environment ,since it's implemented without changes to the client-server protocol.The proposed system is useful for external threats but cannot protect from insider threats[7].

# 3. Theoretical Background

## 3.1 Overview of Data Loss Protection (DLP)

Data loss protection is the practice of detect/prevent sensitive data from being leaked out of an organization's boundary for unauthorized users or "misused" by authorized users. It protects potential data branch incidents in a timely manner and prevents them by monitoring data in an organization's network [2].The basic physical parts of a DLP system present in figure (1) basic data loss protection system[8].

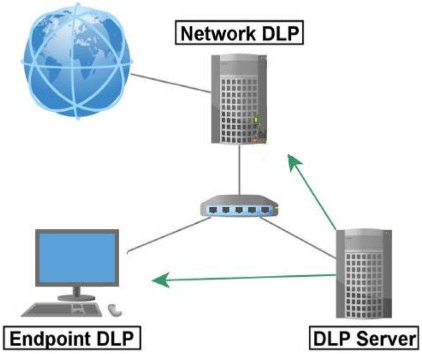


**Figure (1): Basic data loss protection system [8]**

**Endpoint DLP:** this part installed on a workstation or other device be formed as the agent. It Responsible for how data is stored (called data at rest) and who can use it and what (data in use).

**Network DLP (data in motion):** installed between LAN and WAN to monitor network traffic.

**DLP server:** It is responsible for managing previous parts, also for policy management (policy deployment and logging policy violations)

Politics is most important part of the DLP system because it enables to distinguish between different levels of data (public and sensitive).

The determination of these policies depends on the organization's own specifications.

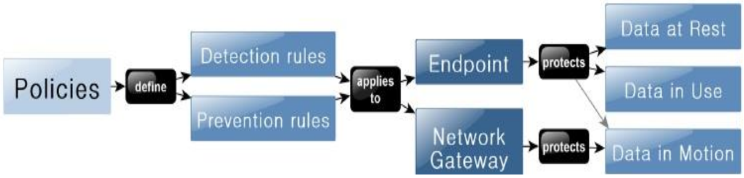


**Figure (2): Policy overview[8]**

After DLP policy creation for any organization, these policies be converted into rules and then the DLP enforce during operation. Figure (2) shows steps of how policies are converted into rules. DLP solutions aim to either prevention rule or detection rule by applying different methods:

### A. Prevention method:

Prevention method is a strategy to make sure that the data is confidential to any organization doesn't move outside of the system. It is the type of security technology that works to secure confidential data from unauthorized people automatically. Prevention system creates policies to automatically make sure that confidential data is not manipulated in terms of stored or access only by authorized users [8]. These policies ensure that the data is not lost and at the same time allows users to use the tools and services they need to perform their tasks. It can also handle policies at different levels of sensitivity and data access control [8,9]. Prevention methods, potential leakage is prevented before they occur through the use of appropriate methods .

### B. Detection method:

The detection method refers the use of a variety of techniques and methods to find patterns that have abnormal behavior in the dataset. These abnormal behaviors are also termed as anomalies. An anomaly means something unusual or unexpected behavior that does not match to the normal behavior [10]. The primary advantage of detection method is its ability to find new attacks. However, false alarm rates are in general very high. A typical detection model is shown in figure (3). It consists of four components [11]: I.Data collection, collects normal user activities and save them. ii.Normal system profile, create normal system profiles. iii. Detection method: determines the abnormal behavior from the normal system profiles and finally, iiii.response. Provides reports on violation information as well as information on the time of the violation.

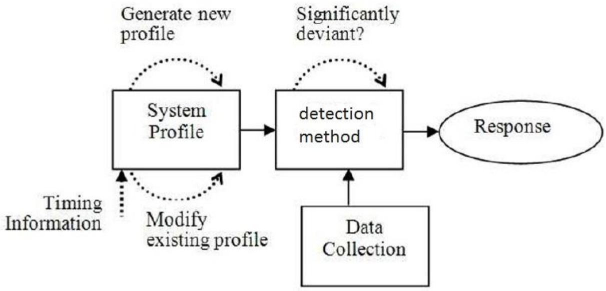


**Figure (3):Typical detection system[11]**

In recent years, detection method was classified according to the learning techniques used:

1. **Supervised detection:** this type of detection based on supervised learning, which learned a model from the training dataset. This learnig can classify into following cases:
   a) one-class classification problem,that only normal data or abnormal data are learned .
   b) two-class classification problem:Both events (normal and abnormal) are learned.

c) multiclass classification problem:Multiple classes of events are learned.

**Outlier detection:** This type of detection based on unsupervised learning, this learning involves both (normal and abnormal) training data without labels [12] .

2. **Novelty detection:** This type of detection based on Semi-supervised learning, which the learning assumes that training data labeled as the normal class only, without needed to labels for the abnormal class [13].

### 3.2 Electronic Health record (EHR)

E-Health systems in many countries offer many services to their citizens such as individual health cards, Clinical decision support system, Telemedicine etc. In addition, e-Health systems provide the electronic health records (EHRs). An EHR includes all patient's health information, and the aim of EHRs is interconnecting the health practitioners(doctors, nurses, specialists,etc.) in common point[14].

The health information in EHR is private, so it should be protected.EHR protection can provide by two important issues ,security and privacy, these issues provide protection for patient health information by protecting integrity and confidentiality from unauthorized access[15].

## 4. Proposed System Architecture

The proposed system in this paper suggests a DLP system divided into two phases, prevention phase and a detection phase. Prevention phase using signature-based techniques to prevent well-known attacks and reduce the number of alerts raised.In other hand, detection phase using novelty detection and modified decision tree to detect unseen attack and rescue of propagating of data loss and the damage they cause. Figure (4) presents the proposed DLP solutions in three stages ,(i)prevention stage ,(ii)novelty detection,and (iii)supervised detection.

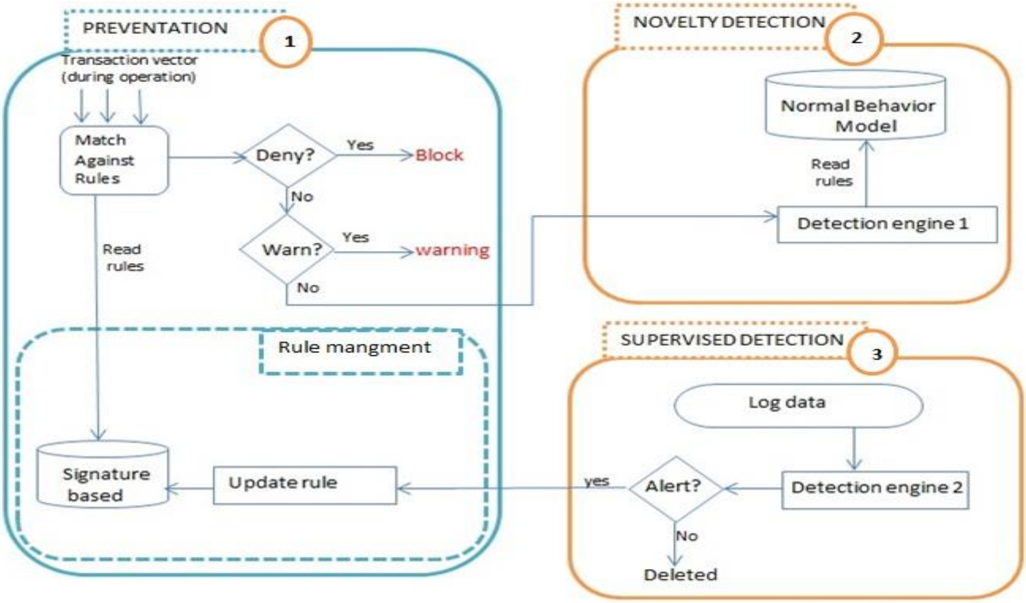


**Figure (4):Proposed data loss protection system**

## 4.1 Proposed Prevention Stage

Propose prevention system based on a signature-based technique to provide the first level of security. Signature-based identifying patterns that describe the abnormal behavior for each user to distinguishing normal behavior from suspicious called blacklisting. The blacklisting building is based on predefined policies that fit with an organization's requirements, where it represented abnormal behaviors for users inside an organization.Table (1) presents an example on blacklisting, This table defines by the administrator to define the blocking and warning rules. Each rule-based generating from these blacklisting is matched with each a new transaction arrives into the system, The response of prevention module depends on the rules matched between them.

The prevention module describes the well-known attacks With few or not produced any false alerts, therefore it was able to block attacks before executing .However, it can not able to describe new attacks.

**Table (1): Example of blacklisting**

| Id | User name | Password | Command | Table list | level | IP address |
|---|---|---|---|---|---|---|
| 1 | Alice | AS12ff*1 | Non | no | 1 | 192.168.1.6 |
| 2 | Bob | Art23%$ | Delete, insert | Table1 | 2 | 192.168.1.5 |
| 3 | Joe | !gh!12HH | Delete, inset | Table1 | 2 | No |
| 4 | Leo | qT8#jSL3 | Update,delete,insert | Table 2 | 2 | 192.168.1.2 |
| … | …. | … | … | … | … | … |

## 4.2 Proposed Novelty Detection Stage

Proposed novelty detection system based on one class classification idea by identifying normal behavior model. The normal behavior depends on identifying the Normal profile of users, when the normal behavior profile(i.e. Whitelisting) is defined, whitelisting rule-based will be generated,  and it learns on normal profiles to detect abnormal behavior by any deviations from these profiles. The aim of novelty detection to solve prevention problem, by detection new attacks using monitoring user activities with the system and finding any deviations from normal profile. The proposed novelty detection is divided into three phases:

### 1) Normal profile(whitelisting) building

During monitoring the database activates ,the  Normal profile or whitelisting is built, the whitelisting building based on the user interaction with the database. Table (2) presents examples of whitelisting.

**Table (2):  whitelisting**

| Id | User name | Table list | Command | Column list | Query length | Time work |
|---|---|---|---|---|---|---|
| 1 | Alice | Table1 | Non | medicine | 200 | 01:00to 04:00 |
| 2 | Bob | Table1 | Delete, insert | treatment | 300 | 8 to 13 |
| 3 | Joe | Table1 | Delete, inset | Medicine,age | 170 | 00 to 7 |
| 4 | Leo | Table | Update,delete,insert | treatment | 400 | 9 to 13 |
| … | …. | … | … | … | … | … |

### 2) Model learning

When the whitelisting for each used is completely described, the whitelisting rules will be generated.These rules learn on the whitelisting.The rules generation is based on the direct method to extract rules from whitelisting without using classification techniques, the sequential covering is one of the effective ways to extract rules directly from the database.This method generates one rule belong to one class by learns each record in the whitelisting.

### 3) Detection phase

The last phase is detection model,during this phase, each new transaction coming into the system will be analyzed and matched with normal profile to be identified if this transaction is normal or not. Detection phase can be explained in set of definitions:

**Definition1** (Transaction **----Tn** )

Transaction Tn is formed from<S, R, C> where:
S, represent the SQL query R, represent result corresponding S and C additional information like time, response code and other information.

**Definition 2 (Feature** vector **--- F).**

The feature vector F = {f1, f2, ..., fn} Contains all characteristics to represent a query. , i.e. Syntax, context and result as shown as table (3).The Feature vector is necessary to define the kind of attacks and user activity monitoring.

**Definition 3** (profile feature **--- Pf** ) .

the profile feature Pf={Pf1,fPF2,fp3 ,....., Pfn}, during the monitoring of creating Feature vector ,the profile feature will be created.

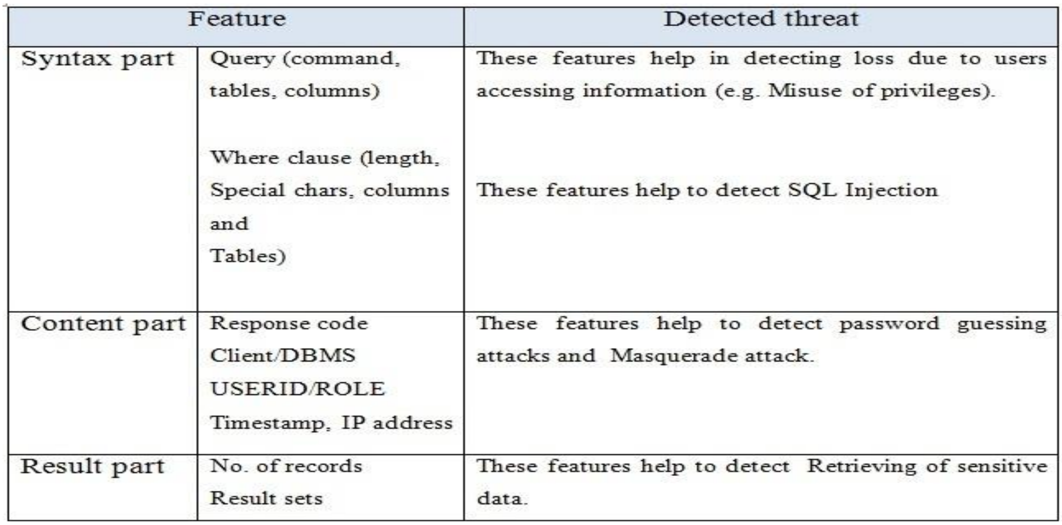
**Definition 4 (**transaction matching **---Tm**)

The transaction matching (Tm) is configured from a set of components <F,Pf,L> the result can determine by matching between giving feature vector F and profile feature Pf and return the result in the L vector, where L is a vector of the result.

**Definition 5 (**detection engine **---M**)

Detection engine M, means using the result from matching L :{l1,l2,l3...ln} and whitelisting rule based detection R, detection engine determines if the transaction is {normal or anomaly}.
M(L, R)={anomaly, normal}.

**Table (3): Feature classification and relationships with attacks**.



| Feature | | Detected threat |
|---|---|---|
| Syntax part | Query (command, tables, columns)<br>Where clause (length, Special chars, columns and Tables) | These features help in detecting loss due to users accessing information (e.g. Misuse of privileges).<br>These features help to detect SQL Injection |
| Content part | Response code<br>Client/DBMS<br>USERID/ROLE<br>Timestamp, IP address | These features help to detect password guessing attacks and Masquerade attack. |
| Result part | No. of records<br>Result sets | These features help to detect Retrieving of sensitive data. |

## 4.3 Proposed Supervised Detection Stage

Proposed supervised detection is the second level in detection phase, it presents a proposal by applying supervised learning ,and chose to use an ID3 decision tree for several reasons will know later. Typically,all user activities with the system is stored in central log,and usually the log data are being unorganized and difficult to deal with directly.Supervised detection is analyzing the log to use for understand End User Behaviors and detect abnormal Behavior. Proposed system review the log data and classify data into unusual and usual events .i.e. Locked accounts,data access,login attempts and other events. The Proposed system based on idea of Training data preparation by using some Steps,

I. Analysis attributes and identify ( conditions and values) for each attribute in the training set.

II.Specifies number of records from values and conditions by applying follow equation :

- ***No.of.record****=P1 in C1* p2 inC2* p3in C3.... pn in Cn* (1)

Where:

Pn: Number of attribute
Cn: Number of class

## 5. Evaluation Methodology

The goal of an evaluation is to evaluate the effectiveness of proposed DLP on EHR security ,also to identify effective each phase on the system and on other phases with in system. For the evaluation ,simulate system with two tables of health care obtained from (PKDD1999),, This database is available on the web at (http://lisp.vse.cz/challenge).First table includes basic information about patients (input by doctors or nurses, about 1300 records), and the second table includes special results Laboratory for patients (input by doctors).

### 5.1 Experiment 1: Prevention phase

The experiment aims to assess proposed prevention module and how it effect on system, first experiment has assessed the impact of this proposed module on reducing the number of alert raised in period time inside the system . Figure 5 shown the measured the effect with and without proposing prevention module over 15 days of analysis. The experiment result shows, the prevention phase reduces the number of alert raising approximately, 29% in healthcare database.

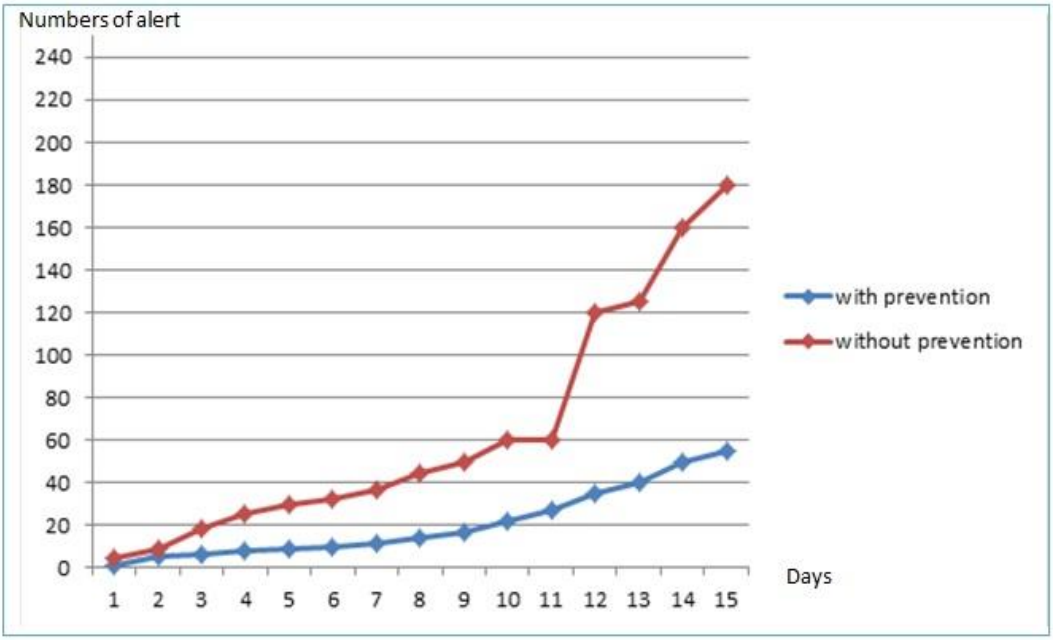


**Figure (5): Number of alerts (with and without a without prevention) in healthcare data.**

## 5.2 Experiment 2: Novelty Detection

This experiment aims to assess novelty detection,by focus on the detection rate for detecting attacks by insider attacks.an attack is classified as two classes:

• Syntax based Attack (ATK-1): this Attack represents set of actions to access parts of the database (e.g. Tables and columns ).

• Result-set-based Attack : (ATK-2): this Attack refers to the actions of a malicious in the result set of queries.

The novelty system architecture depends on handled the query process, Every query issued, it is analyzed by the detection engine before execution.

Now,to compare the efficiency of the proposed novelty detection with similar approaches, the Kamra et al. (2008), Wu et al. (2009), E.

Costante(2014) and E. Constant (2016) . There are two measures using for this :

- Detection rate(DR)=Sensitivity = (true detections) / (all samples), (2)
- Detection time(DT)=MS(milliseconds) per single transaction. (3)

The proposed system evaluation needs applying detection rate(DR) over two types of attacks(ATK1 and ATK2) and compare it with other previous proposals from the literature.To illustrate the result of the comparison can follow the table (4).

**Table (4): DR comparison for novelty detection**

| Authors | Detection system | DR (ATK 1)(%) | DR (ATK2)(%) |
|---|---|---|---|
| Proposed Detection phase | Novelty Detection | 100% | 90.28% |
| E. Costante(2016) | White box | 100% | 88.46% |
| E. Costante(2014) | Whitebox with RS | 100% | 80.77% |
| Kamra et al. | Quiplet | 100% | 76.08% |
| Wu et al. | Standard | 100% | 38.46% |

The average time needed to detect the single transaction and the time measured by a computer for novelty detection and compare with other solutions. Generally, detection time requires around 0,03 ms per single transaction in the novelty proposed system and compared with the other solutions where **Wu et al.**(2009) has reported detection time of 0,08 ms,**Karama et al**.(2008)gave a detection time of 0.06 ms and **Costante E**(2016) has obtained detection time of 0.04 ms.Time analysis using of several approaches is presented in figure (6).

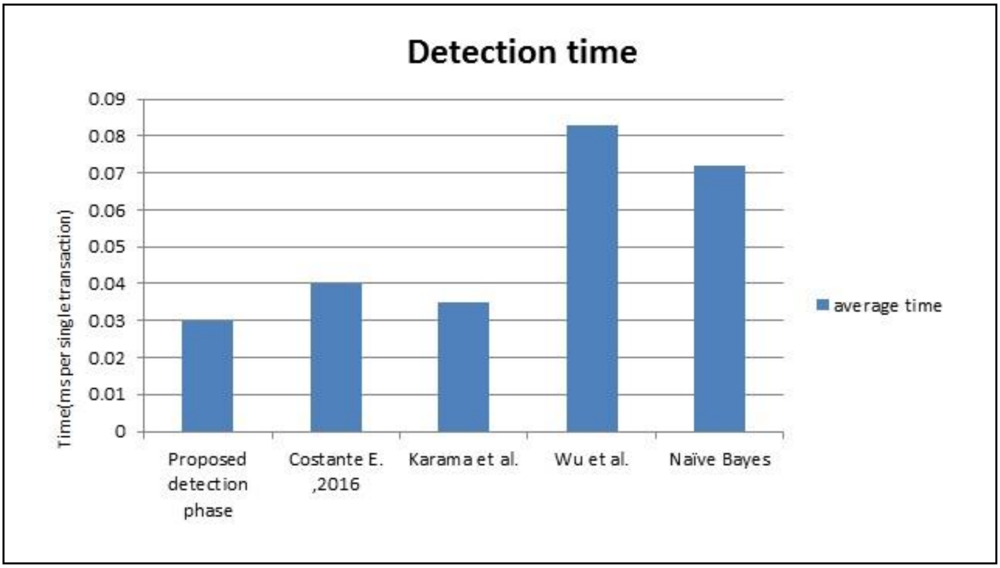
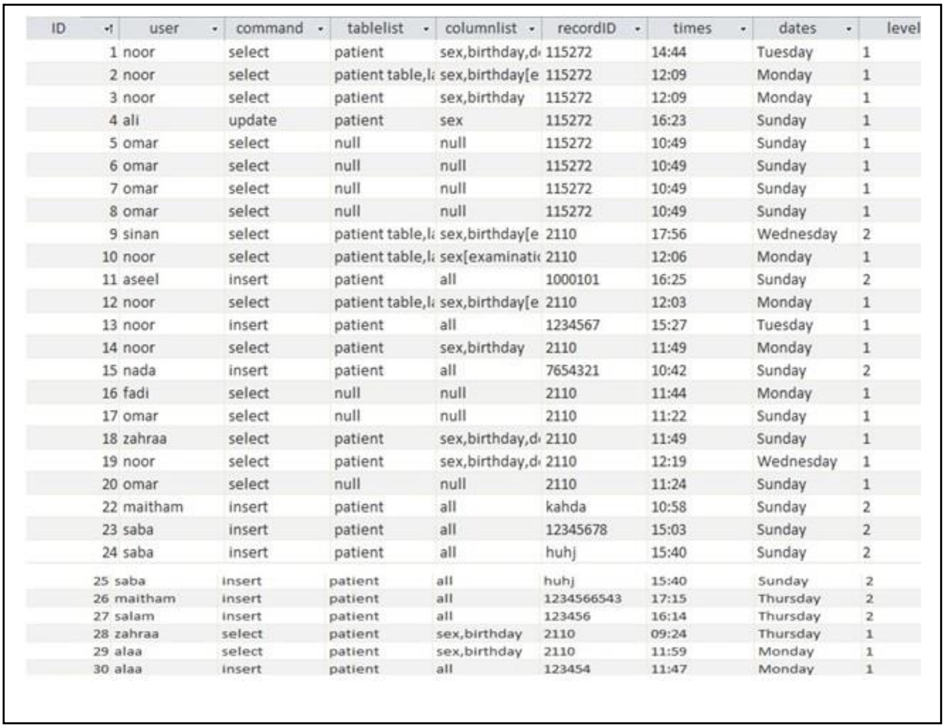


**Figure (6) :Detection time for the different approach**

## 5.3 Experiment Supervised Detection

The experiment of effectiveness the supervised detection can measure how effect on log analysis and for detect new attacks, and then compares this new mechanism with the standard mechanism. The previous approach is novelty detection focused on two types of attacks syntax based and result-set-based attacks, in this approach, focus on one type of attack , context based attack .Typically ,this attack originates from malicious queries i.e.unusual access in the IP address or at time or date.

The aim of proposed supervised detection applies to the system logs will be to classify the traffic, for both normal and anomalous. This approach using ID3 decision tree algorithm with modified training data to analyze a system log . Figure (7) present the snapshot of system log, the attribute view in this log include the command, level,tables, columns,and date and time of the user entering. attribute selection for using in ID3 is very important for effective on detect an anomaly in the log data. The output of a system log analysis is a collected in the central table called security log, the security log will be used as training data to be analyzed and evaluate the detection rate for this approach.



| ID | user | command | tablelist | columnlist | recordID | times | dates | level |
|---|---|---|---|---|---|---|---|---|
| 1 | noor | select | patient | sex,birthday,d | 115272 | 14:44 | Tuesday | 1 |
| 2 | noor | select | patient table,l | sex,birthday[e | 115272 | 12:09 | Monday | 1 |
| 3 | noor | select | patient | sex,birthday | 115272 | 12:09 | Monday | 1 |
| 4 | ali | update | patient | sex | 115272 | 16:23 | Sunday | 1 |
| 5 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 |
| 6 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 |
| 7 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 |
| 8 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 |
| 9 | sinan | select | patient table,l | sex,birthday[e | 2110 | 17:56 | Wednesday | 2 |
| 10 | noor | select | patient table,l | sex[examinati | 2110 | 12:06 | Monday | 1 |
| 11 | aseel | insert | patient | all | 1000101 | 16:25 | Sunday | 2 |
| 12 | noor | select | patient table,l | sex,birthday[e | 2110 | 12:03 | Monday | 1 |
| 13 | noor | insert | patient | all | 1234567 | 15:27 | Tuesday | 1 |
| 14 | noor | select | patient | sex,birthday | 2110 | 11:49 | Monday | 1 |
| 15 | nada | insert | patient | all | 7654321 | 10:42 | Sunday | 2 |
| 16 | fadi | select | null | null | 2110 | 11:44 | Monday | 1 |
| 17 | omar | select | null | null | 2110 | 11:22 | Sunday | 1 |
| 18 | zahraa | select | patient | sex,birthday,d | 2110 | 11:49 | Sunday | 1 |
| 19 | noor | select | patient | sex,birthday,d | 2110 | 12:19 | Wednesday | 1 |
| 20 | omar | select | null | null | 2110 | 11:24 | Sunday | 1 |
| 22 | maitham | insert | patient | all | kahda | 10:58 | Sunday | 2 |
| 23 | saba | insert | patient | all | 12345678 | 15:03 | Sunday | 2 |
| 24 | saba | insert | patient | all | huhj | 15:40 | Sunday | 2 |
| 25 | saba | insert | patient | all | huhj | 15:40 | Sunday | 2 |
| 26 | maitham | insert | patient | all | 1234566543 | 17:15 | Thursday | 2 |
| 27 | salam | insert | patient | all | 123456 | 16:14 | Thursday | 2 |
| 28 | zahraa | select | patient | sex,birthday | 2110 | 09:24 | Thursday | 1 |
| 29 | alaa | select | patient | sex,birthday | 2110 | 11:59 | Monday | 1 |
| 30 | alaa | insert | patient | all | 123454 | 11:47 | Monday | 1 |

**Figure (7): System log**

Now,through applying modified training set algorithm on system log, the resulting attributes are(user level, date and time) can be applying ID3 for generating tree and rule-based.Figure (8) presents decision tree.

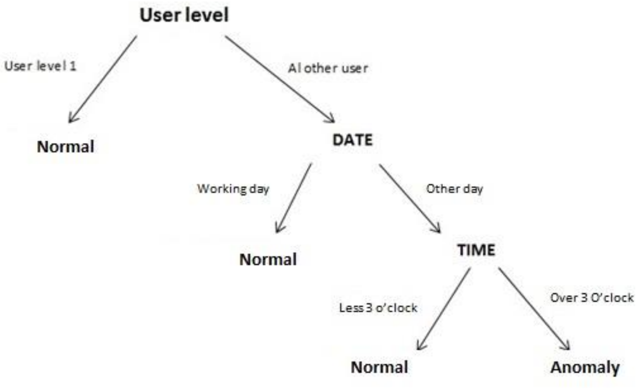
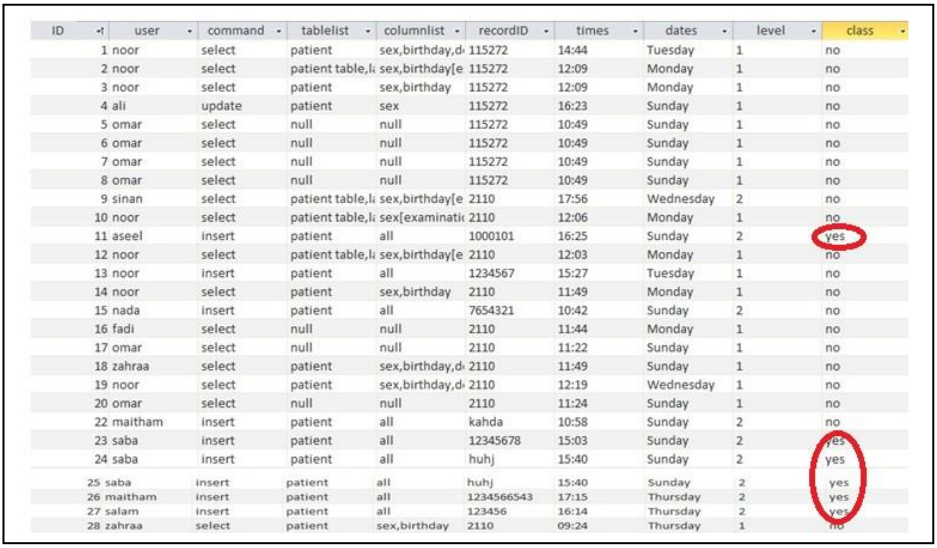


**Figure (8): Decision tree**

Therefore ,this data is detecting the context based attack, These attacks occur over the span of about four days during normal business hours or outside working hours.Figure (9) presents a snapshot of security log which was recorded 30 incidents during that one day. Of these 30 incidents were determined to have been analyzed and then classified to detect any anomaly.



| ID | user | command | tablelist | columnlist | recordID | times | dates | level | class |
|---|---|---|---|---|---|---|---|---|---|
| 1 | noor | select | patient | sex,birthday,d | 115272 | 14:44 | Tuesday | 1 | no |
| 2 | noor | select | patient table,l | sex,birthday[e | 115272 | 12:09 | Monday | 1 | no |
| 3 | noor | select | patient | sex,birthday | 115272 | 12:09 | Monday | 1 | no |
| 4 | ali | update | patient | sex | 115272 | 16:23 | Sunday | 1 | no |
| 5 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 | no |
| 6 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 | no |
| 7 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 | no |
| 8 | omar | select | null | null | 115272 | 10:49 | Sunday | 1 | no |
| 9 | sinan | select | patient table,l | sex,birthday[e | 2110 | 17:56 | Wednesday | 2 | no |
| 10 | noor | select | patient table,l | sex[examinati | 2110 | 12:06 | Monday | 1 | no |
| 11 | aseel | insert | patient | all | 1000101 | 16:25 | Sunday | 2 | yes |
| 12 | noor | select | patient table,l | sex,birthday[e | 2110 | 12:03 | Monday | 1 | no |
| 13 | noor | insert | patient | all | 1234567 | 15:27 | Tuesday | 1 | no |
| 14 | noor | select | patient | sex,birthday | 2110 | 11:49 | Monday | 1 | no |
| 15 | nada | insert | patient | all | 7654321 | 10:42 | Sunday | 2 | no |
| 16 | fadi | select | null | null | 2110 | 11:44 | Monday | 1 | no |
| 17 | omar | select | null | null | 2110 | 11:22 | Sunday | 1 | no |
| 18 | zahraa | select | patient | sex,birthday,d | 2110 | 11:49 | Sunday | 1 | no |
| 19 | noor | select | patient | sex,birthday,d | 2110 | 12:19 | Wednesday | 1 | no |
| 20 | omar | select | null | null | 2110 | 11:24 | Sunday | 1 | no |
| 22 | maitham | insert | patient | all | kahda | 10:58 | Sunday | 2 | no |
| 23 | saba | insert | patient | all | 12345678 | 15:03 | Sunday | 2 | yes |
| 24 | saba | insert | patient | all | huhj | 15:40 | Sunday | 2 | yes |
| 25 | saba | insert | patient | all | huhj | 15:40 | Sunday | 2 | yes |
| 26 | maitham | insert | patient | all | 1234566543 | 17:15 | Thursday | 2 | yes |
| 27 | salam | insert | patient | all | 123456 | 16:14 | Thursday | 2 | yes |
| 28 | zahraa | select | patient | sex,birthday | 2110 | 09:24 | Thursday | 1 | no |

**Figure (9): Security log**

Now, by using DR measure, evaluate proposed a supervised system, The DR result is up to the detection limit 100% for the context-based attack.

## 6. Conclusions and Future Work

This paper develops proposed DLP system from blacklist,whitelist and decision tree ID3 solutions to enhancing security in electronic health record.These techniques used to enable both side prevention and detection in EHRs using these techniques will be obtained several conclusions, can be summarized as follows:

- Prevention method with signature-based have the ability to prevent privilege missuse attack, but unable detect zero-day attacks.Also, the prevention method help to improve the accuracy of the detection method, while in the long term, it reduces the number of alerts is raised in the system.

- The proposed novelty detection improves detection phase in two sides, firstly, detection rate, it detects near to 100% for two types of attacks.Secondly, detection time, the proposed system reduces time needed for detection is around 0.03ms per single transaction.
- The proposed modified training set from system log will improve detection rate, which it gives a high detection rate around 100% on the context-based attack.
- Finally, the proposed DLP solution in detection phase based on three parts of the feature set in queries(Syntax, Content, and Result part), the using of different feature provides system scalability to apply in several domains such as web applications,network-based intrusion detection systems and firewalls.

## References

[1] Koutsourelis d.2014.designing a free data loss prevention system.msc.thesis,university of piraeus.

[2] A. shabtai, y. elovici, and l. rokach.2012. a survey of data leakage detection and prevention solutions, ser. springerbriefs in computer science. springer.new york.92.

[3] Muneer.2011. a new approach for sensitive data leakage prevention based on viewer-side monitoring.m.sc.thesis, al-balqa' applied university,salt-jordan.

[4] Bradley malin, steve nyemba, john paulett.2011. learning relational policies from electronic health record access logs. journal of biomedical informatics.p: 333–342.

[5] Santos rj, bernardino j, vieira m.[2014] approaches and challenges in database intrusion detection. sigmod record .vol. 43, no. 3,p:36-47.

[6] Costante e, etalle s, fauri d, hartog j, zannone n.2016. a hybrid framework for data loss prevention and detection. in proceedings of the workshop on research for insider threats (writ 2016). ieee.

[7] Wasim a al-hamdani.2016. cryptography based access control in healthcare web systems. .researchgate.p:66-79.

[8] Tore torsteinbø.2012. data loss prevention systems and their weaknesses.m.sc.thesis, department of information technology university of agder.87.

[9] Mohd. mahmood ali, mohd. s. qaseem, lakshmi rajamani, a. govardhan.2013.

[10] Extracting useful rules through improved decision tree induction using information entropy. international journal of information sciences and techniques. vol.3, no.1,p:27-41.

[11] Shikha, jitendra ,.2015. survey on anomaly detection using data mining techniques. elsevier ,p: 708 – 713.

[12] Ali a. ghorbani, wei lu &mahbod.2010. Network intrusion detection and prevention: concepts and techniques.springer new york dordrecht heidelberg london.

[13] Dinesh, dikshika .2016 techniques and challenges in building intelligent systems: anomaly detection in camera surveillance. springer, switzerland .vol 2,p:11-21.

[14] Chandola, v., banerjee, a., and kumar, v. 2009. Anomaly Detection: a survey. acm computing surveys, vol. 41, no. 3, article 15,p:15-58.

[15] Joel josé p.c. rodrigues,sandra sendra compte,isabel de la torra diez.2016. e-health systems. iste press ltd and elsevier ltd. great britain and the united states.275.

[16] Sherali zeadally , jes ´us t´ellez isaac , zubair baig.2016. security attacks and solutions in electronic health (e-health) systems. springer science+business media,p:1-12.

# مقترح لحماية فقدان البيانات في سجل الصحة الالكتروني

أ.م.د عبير طارق مولود*

الباحث: رشا محمد محسن*

**المستخلص:** يقدم هذا البحث اقترح حلول حماية فقدان البيانات (DLP) لتحسين أمن سجل الصحة الإلكترونية (EHR). طريقة الوقاية هي المرحلة الأولى في نظام DLP، فإنه يستخدم (signature-based) لتوفير حظر الهجمات المعروفة، وأيضا لتحسين دقة النظام عن طريق الحد من عدد التنبيهات. من ناحية أخرى، مرحلة الكشف لديها مستويين من الكشف،الكشف خلال الاستخدام، والكشف دون اتصال. الكشف خلال الاستخدام يتم باستخدام (novelty detection) المستخدمة للكشف عن هجمات جديدة تصل في النظام، في حين ان الكشف بدون اتصال يتم الكشف عن اي سلوك غير طبيعي للمستخدم خلال فترة زمنية محددة باستخدام ( supervised detection).

**الكلمات المفتاحية:** الوقاية ،الكشف عن الجدة،الكشف تحت الاشراف، القائم على التوقيع ،القائمة البيضاء،شجرة القرار(ID3).

*قسم علوم الحاسوب،الجامعة التكنولوجية،بغداد،العراق