American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

A Technique for Classifying and Retrieving of Malware Detials in Signtures Based

Salah Albermany*

Department of Computer Science Faculty of Computer Science and Mathematics /Kufa University /Iraq
Email: salah.albermany@uokufa.edu.iq

Abstract

Signature based is one of the common techniques that are used to detect malware attack. The problem of the signature based is about the management of large database that has a new signature, in this paper we will create a new method to classify and fast retrieve malware of database, the size of database increase database is dependent on the number of signatures that are based on malware file , to solve classify database by using the concept of room based, we use this concept “room based” to manage the database. Each room based that has content Prohibition privileges of signature based on malware files, or pattern of collections of signature based of malware files.

Keywords: Type *signature base malware detection; database management; Role Base Access.*

1. Introduction

Computer security is the main concern of many individual organizations, as attacks on network and computer system have recently increased and the main motivation of the attackers recently have been toward financial matters gain. Therefore, there is a need for a better and more robust detection and prevention system. However, there are many existing detection and prevention system such as IDS and IPS, and both systems are based on signature and anomaly technique. Signature based, is a technique that receives a packet of data (signatures) then the signatures through the malware analysis will be removed. Then comparing the signatures and matching with existing signature data base and storing it in the signatures dictionary [1] many commercial anti malware system are using this technique because it is fast in scanning time and also it gives few false positive [1].

---

* Corresponding author.

But on the other hand this technique is not able to detect a new malware which does not exist on the data base and also is not able to deal with simple obfuscation technique such as hiding polymorphic and zero day attack [1]. And the other weakness of this technique is that the data base of these techniques is updating continually the signatures in order to detect a new malware attack that have a new signature [2]. In this paper we will discuss the subject of Probations privileges to study malware software and the extent of its request for accesses, read, update, move and other. For each malignant program and non-malignant program they have their own set of requirements. and we will discuss the granted the requirement or reject it depending on the purpose of the program and in which group that this program is linked to it such as clean clear software group and this group is called a clean software , false positive group this type of software that test incorrectly to know, is found on the existence of the signing of a malicious program in a file is not infected and the problem of this that the anti-virus reject file that doesn't content malware because the anti- virus cannot Prevention identified and excellence [3] false negative group is a negative of the false positive that false negative can give a wrong report about the file or some time it's not able to know if the file infected or not and this is the worst case because it may accept an infect file that contain malicious [3]. Bugs this type of files is under specific certain condition and malware group. [3]. In general the software is categorized to group of levels and we will name it on our paper as rooms, each room has ratio risk percentage. A clean room the ratio risk percentage of this room will be nearly to zero. And the malignant room the ratio risk percentage will be nearly to one. For specifying the room we should know the Probations privileges and then based on it the distribution of the software into rooms.

The purpose of the most detection malware algorithms, is to test the unknown file to know its type. And this type will main categories in there sections: the known files, is the files that are easy to know there signature and category. And these categories are clean, malware, false positive, false negative and bug. Second, the unknown files or new files. Third the encryption files the unknown main signature [3].

The important problem in signature-base malware detection is that the large database and the way of the fast retrieving from database, especially the number of signature are generated and increasing every time .therefore, the normal way of the storing in database is not suitable. In this paper we use the concept of room based that contains the entire non privilege file that we call Prohibitions. And each software want a group of Prohibitions that will be in malware software. If the software doesn't have any Prohibitions this mean that this software is clean (clear) and we don't need to study their signatures, the Prohibitions room cannot contain all the Prohibitions but it can contain most of Prohibitions and this room which contain most of the Prohibitions in this room will be at Maximum level of malware and room which contain one Prohibitions this room will be at Minimum level malware. We will categorize this research to three main parts , in this paper we will creates the room base in group of prohibited privileges and create signature based on patterns recognize[4] for signature and located in room based and here we will divided it in two cases the first case that the weight of the prohibited privileges is equal value and the second case is the prohibited privileges that has weighted .

2. Rooms base

2.1. What is Rooms Base?

In this paper we used meaning of Rooms Base to the database management of signature-base of malware, and to reduce the increasing storage in the database because the creation of the new malware. The Rooms Base is the home of decision represented by single object (user, file, data, pattern, ...), single group or multiple groups, in our paper each room Base could have one or group of patterns signature-base of malware, such as the patterns in the same room are Convergent properties of signature-base code.

If n represent the number of the objects (user, files, data, pattern, ...) in group U where n >1 then:

$$n=n_1+n_2+n_3+...+n_r \qquad (2\text{-}1)$$

such that r represent the number of division n from the objects to group gathering depended on the levels of patterns of signature base malware code, and all gathered level one its group n>ni object in granted level i, symbolize of this group of patterns close in Ri, and we named it as a Room Base which represented Role connection.

2.2. Why Rooms Base?

The need of applied for presence room Base: Since each pattern own group of prohibited privileges to store in database. Therefore, the need for flexible method to manage the huge number of the of signature base that have Prohibition privileges of the files, user. Especially, for large database which is called by the protection of the environment. (Role –Based Connection) [7], as we plan to create sessions of the patterns with keeping on the prohibition privileges level which donated to the prohibition privileges, therefore we need to specify the suitable group from pattern. With the prohibition privileges level that is specified before for the pattern. Ablility to create from pattern's group session within. This group which created from the patterns without the effect of the protection of action creator. The reason of the failure in using Rule, and using Room Base.

2.3. Number of Room Base

The Number of Rooms Base wanted (denoted by r) depended on the number of the objects (files of malware, ...) n in U, and the number of Prohibition privileges denoted by P, then the size of P is the number of the possible Prohibition privileges. There are two cases to determine the number r:

1- Predetermine r for any number of objects (files of malware,...) n in U, and the number of Prohibition privileges p in P, and. This case has the disadvantage that it does not specify the Prohibition levels for the (files malware, patterns,...) well in a room base, and this selection is considered in restriction (files malware, patterns,...) in rooms base.

2- The number r of rooms base, its determination without restrictions, depends on the Prohibition privileges for the file of malware, r its optimize number satisfies that each room base have maximum Prohibition privileges of the malware file belong to same room base.

2.4. Size of Room Base

The number of malware files (or patterns) ni at room base i (see the equation (2-1)), and There are two cases that either pre-determined and thus its restriction represents the largest number of malware files (or patterns) in a room base, or determines through the greatest number of beneficiaries in the room base.

2.5. Prohibited privileges

Since malware is compromising the systems' confidentiality, integrity, and availability [5], where each malware, it has Prohibited privileges, where Prohibited privileges is a group of prohibitions is not conferred to malware.

2.6. Weight of Prohibited privileges

In normal case the element of group P and its element equal weights in the Room base, there is no difference between them, all the weight is equal to one. If the elements on P are different by weight for Prohibited privileges, such as the large weight value represent the maximum level of Prohibited privileges, other side for small weight value represent the maximum level of Prohibited privileges, denoted to the weight of pi by W(pi) , i=1,..., p, where

$$W(p^1) < W(p^2) < ... < W(p^p)$$

3. Mathematical model

In this paper we will focus to create rooms base for a limit group of malware files or pattern n, such as each malware files or pattern owns a partial group of Prohibited privileges within group of all prohibited privilege (p) available in the signature base code, in condition there is no malware files or pattern that has all the prohibited privilege that is offered in general, and also the malware files or pattern will not be with the group of the malware files or pattern. if this malware files or pattern has not owned any prohibited privilege within the set of prohibited privileges that available the signature base code, and it has an external Prohibited privileges (external Prohibited privileges means that the malware files or pattern has external Prohibited privileges) which is not related to the Prohibited privileges group that is offered from Known malware traces [6] and does not affect the group, and incase a group of malware files or pattern that has a similar number and weighted Prohibited privilege that donated to them, so they will be considered as one malware files or pattern with the references to them and the number of the malware files or pattern. The malware files or pattern that have entitles the biggest weight of Prohibited privilege will be within maximum level of Prohibited privilege that is offered. On the other hand the malware files or pattern that has a small number of entitles of Prohibited privilege will be the minimum level of the Prohibited privilege that is offered. We will divide the Prohibited privilege offer in two parts based on the weight of the Prohibited privilege. The first type is the equal Prohibited privilege Wight for the malware files or pattern. The second type is the non-equal Prohibited privilege weight in which two Prohibited privilege has different weight.

The goal of this paper is to discuss the mechanism of distribution the malware files or pattern rooms based on the offer Prohibited privilege to them, as two Prohibited privilege type , to create a sessions from (p) partial from room based, where we will discuss how to create the room based.

3.1. The problem

There is a group of Prohibited privilege (p) in database and this Prohibited privilege can of the malware files or pattern, where all the group of Prohibited privilege is stripped. There is no difference, this mean that all of them are equal in the weight (Prohibited privilege weight), and it is represented as high Prohibited privilege, and different word, the Prohibited privilege that this weight is one, is the fundamental, for all the Prohibited privilege within the group of the group of the Prohibited privilege (p), this means each Prohibited privilege weight is related to the others. Prohibited privileges of the malware files or pattern Fi, the problem is how to create a room based, each room based is an optimal numbers of malware file or pattern, with condition the intersection between complete Prohibited privilege of malware files or pattern as a minimum, and it is be better if it's zero or near to zero.

3.2. The problem Formulation

In this paper we will discuss how to form the problem of the weights of the Prohibited privilege of equal value, we will discuss the different weights of the privilege the Prohibited privilege of equal value. We will form the problem by using zero-one-matrix A= (aij), where the rows of the matrix A represent of the malware files or pattern. and the columns of the matrix A is represent of Prohibited privileges, and in general each malware files or pattern has a group of Prohibited privileges, if the malware files or pattern Fi have the Prohibited privilege pj, then 1 =aij, and if the malware files or pattern do not have Prohibited privilege pj for the malware files or pattern Fi then 0= aij, and this type of the matrix is called as Incidence matrix and we can write it by:

$$a_{ij} = \begin{cases} 1 \; if \; malware \; i \; have \; \text{Prohibited privilege j} \\ 0 \quad Otherwise \end{cases}$$

To solve this Incidence matrix A, there are several algorithms to solve A, and create room based.

4. Mathematical programming

Several suggestions have been made to solve (to formulize) the problem, the most common of these suggestions is by using mathematical programming. And we are going to explain the most common suggestion which is integer programming. That using incidence matrix to find the solution, and its features that are using by all mathematical programming. And also other famous examples for mathematical programming is quadric programming and integer programming [8-3]. And to solve mathematical programming we can use several programming language such as Matlab or Lindo, because it is one of the good programming language that deal with linear and non-linear programming case.

4.1. integer programming method

This method suggests that all the Weight of Prohibited privileges are equal. It means that there is (incidence matrix), and there element is zero and one. Before we are formulizing. Linear example of programming for the integer number 0, 1 we create similarity deal matrix (Z):

$$Z_{ij} = \sum_{k=1}^{p} \theta(a_{ik}, a_{jk}) \qquad (4-1)$$

Where

$$\theta(a_{ik}, a_{jk}) = \begin{cases} -1 & if \quad a_{ik} \neq a_{jk} \\ 0 & if \ a_{ik} = a_{jk} = 0 \\ 2 & a_{ik} = a_{jk} = 1 \end{cases} \qquad (4-2)$$

or other mathematics form :

$$\theta(a_{ik}, a_{jk}) = (-1)^{a_{ik} + a_{jk}} (a_{ik} + a_{jk})$$

suppose that Xij integer number (zero, or one), then (malware files, or pattern,...) i in room base j then the value of Xij is 1 else the value of Xij is 0. For above we can formalize integer programming as:

object function:

$$Maximum \ \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij} X_{ij} \qquad (4-3)$$

The constraints on objective function is:

$$\sum_{j=1}^{n} X_{ij} = 1, \quad \forall i = 1, 2, ..., n \qquad (4-4)$$

$$X_{ij} \leq X_{jj}, \quad \forall i = 1, 2, ..., n \ \ and \ \ j = 1, ..., n \qquad (4-5)$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} X_{ij} \geq X_{jj}, \quad \forall \ j = 1, 2, ..., n \qquad (4-6)$$

$$X_{ij} = 0 \ or \ 1 \quad \forall \ i, j = 1, 2, ..., n \qquad (4-7)$$

Constraint in (4-4) show that each (malware file, or pattern,...) that has to be in one room base, the constraint (4-5) show that there if the (malware file, or pattern,...) j in room base j where that Xjj=0, then the room base j is empty and it's not in the solution. The constraint (4-6) show that (malware file, or pattern,...) j in room base j show that minimum there is one other (malware file, or pattern,...) that is in room base and this constraint suppose that at least two (malware file, or pattern,...) as a minimum in one room base. The constraint show that it is in effect on the solution highlights, especially in the case that require the best solution of the (malware file, or pattern,...) in one room base. The constraint (4-7) show that Xij=1 if (malware file, or pattern,...) i will be in room base j otherwise Xij=0, we note in this constraint and this does not specify the Prohibited privileges in room base, and this process does not affect because it is important to specify the (malware file, or pattern,...), and it is easy to know the specification of the Prohibited privilege in room base.

4.2. Implementaion

The System of database that content 6 (malware file, or pattern,...) , Prohibited privilege, incidence matrix (table 4-1).

Table (4-1)

	1	2	3	4	5	6
1	1	1	0	0	1	0
2	0	0	1	0	1	1
3	1	1	0	0	0	0
4	0	0	0	1	1	1
5	0	0	1	1	1	1
6	0	0	0	1	1	1

By using integer programming , we go to explain the function between (4-3) to (4-7). We can write the functions in n=6 and p=6 and the matrix in table (4-1):

Max {$6X11 -2X_{12}+3X_{13} -2X_{14} -3X_{15} –2X_{16}$

$-2X_{21} +6X_{22} -5X_{23} +2X_{24} +5X_{25} +2X_{26}$

$-3X_{31} -5X_{32} +4X_{33} -5X_{34} -6X_{35} -5X_{36}$

$-2X_{41} -2X_{42} -5X_{43} +6X_{44} +5X_{45} .6X_{46}$

$-3X_{51} -5X52 -6X_{53} + 5X_{54} + 8X_{55} +5X_{56}$

$-6X_{61} +2X_{62} –5X_{63} +6X_{64} +5X_{65} +6X_{66}$}

Subject to:

$X_{11}+X_{12}+X_{13}+X_{14}+X_{15}+X_{16} = 1$

$X_{21}+X_{22}+X_{23}+X_{24}+X_{25}+X_{26} = 1$

$X_{31}+X_{32}+X_{33}+X_{34}+X_{35}+X_{36} = 1$

$X_{41}+X_{42}+X_{43}+X_{44}+X_{45}+X_{46} = 1$

$X_{51}+X_{52}+X_{53}+X_{54}+X_{55}+X_{56} = 1$

$X_{61}+X_{62}+X_{63}+X_{64}+X_{65}+X_{66} = 1$

$X_{21}-X_{11} \leq 0$

$X_{31}-X_{11} \leq 0$

$X_{41}-X_{11} \leq 0$

$X_{51}-X_{11} \leq 0$

$X_{61}-X_{11} \leq 0$

$X_{12}-X_{22} \leq 0$

$X_{32}-X_{22} \leq 0$

$X_{42}-X_{22} \leq 0$

$X_{52}-X_{22} \leq 0$

$X_{62}-X_{22} \leq 0$

$X_{13}-X_{33} \leq 0$

$X_{23}-X_{33} \leq 0$

$X_{43}-X_{33} \leq 0$

$X_{53}-X_{33} \leq 0$

$X_{63}-X_{33} \leq 0$

$X_{14}-X_{44} \leq 0$

$X_{24}-X_{44} \leq 0$

$X_{34}-X_{44} \leq 0$

$X_{54}-X_{44} \leq 0$

$X_{64}-X_{44} \leq 0$

$X_{15}-X_{55} \leq 0$

$X_{25}-X_{55} \leq 0$

$X_{35}-X_{55} \leq 0$

$X_{45}-X_{55} \leq 0$

$X_{65}-X_{55} \leq 0$

$X_{16}-X_{66} \leq 0$

$X_{26}-X_{66} \leq 0$

$X_{36}-X_{66} \leq 0$

$X_{46}-X_{66} \leq 0$

$X_{56}-X_{66} \leq 0$

$-X_{11}+X_{21}+X_{31}+X_{41}+X_{51}+X_{61} \geq 0$

$X_{12}-X_{22}+X_{32}+X_{42}+X_{52}+X_{62} \geq 0$

$X_{13}+X_{23}-X_{33}+X_{43}+X_{53}+X_{63} \geq 0$

$X_{14}+X_{24}+X_{34}-X_{44}+X_{54}+X_{64} \geq 0$

$X_{15}+X_{25}+X_{35}+X_{45}-X_{55}+X_{65} \geq 0$

$X_{16}+X_{26}+X_{36}+X_{46}+X_{56}-X_{66} \geq 0$

By implemented the above function, and specify the privilege of room base we find the solution:

4.3. Add new malware files or pattern

In the case of entering Add new malware files or pattern to the group divided in room based simply set the malware files or pattern added in a room based suit in case of proven rooms based and not to change the prohibited privileges , otherwise you will have to re-distribution of malware files or patterns methods to create rooms based, but if you delete the malware files or patterns prefers rework Rooms based again.

We note that the add and delete malware files or patterns influence the nature of the distribution of rooms based opposite prohibited privileges.

Table (4-2)

	1	2	4	5	3	6
1	1	1	0	1	0	0
3	1	1	0	0	0	0
4	0	0	1	1	0	1
6	0	0	1	1	0	1
2	0	0	0	1	1	1
5	0	0	1	1	1	1

5. Analysis and conclusion

5.1. Analysis

Because the time to that the methods to solve it as complex, the solution is limited between the number of the malware files or pattern less than 1000 malware files or pattern. but in our paper we do not need more than 30 malware files or pattern, with the use of add method. The room based technique is dynamic technique for classifying and retrieve malware details where compare with role database. There are several reasons that explain the usefulness of Rules and the useful benefit of Room Base and they are:

1- Since the Role specifies from the server or the database administrator based on the given authorize to users . To achieve the information, but not to take into account in Role, because it might be there two Roles Convergence to each other and the different between them, on privilege only. Both users can be one session without any effect on privileges, because the session is based on privileges type[2]. While Room Base are composed Automatic, without specify from the server.

2- Role is a name but is not restricted [8] but it can be restricted through the privilege granted which is under specify Role.

3- The intersection between two the roles is not necessary to be the low intersection (the low number of privilege that can be) and this is not there on the Room Base, it gives the lowest possible value of the intersection for two Rooms Base.

5.2. Conclusion

This paper has given the common detection technique which is using the signature based to detect attack. In this research paper we investigate the ability of the signature based detection for detecting the new malware attack that has new signature, and the aim of this investigation is to assess the performance and the false positive and false negative and the accuracy of the detection of the signature based in order to detect a new attack. The findings of this investigation show that the signature based is not able to detect a new malware attacks , and also

the database of the signature based is in need to be updating frequently before the new malware signature is created to detect it , and also the classic way of store the signature in data base is not more useful (in way of searching and size) therefore, the idea of our research paper is to improve the size and the searching by manage the database , to do this, we collect all the Prohibitions privilege of the malware and add them to room, each room will content Prohibitions privilege and also the signatures of the all attack, and also well will add an algorithms to generate the signatures and add it to the room, this generation algorithm will help us to detect the a new signature and this signature will be based on the pattern recognize. This aim of the research paper is to contribute the new signature detection. The limitation of the current investigation is focus only on the pattern recognition of the signature. It is recommended that the research will be undertaken in the following area such as the main parts, first part : create the Prohibitions room in group of Prohibitions and create signature based on patterns recognize for signature and located in room and here we will divided in two cases the first case that the Prohibitions is equal and second case is the Prohibitions that has beenweighted, Second Part: Solves the software that content encryption signature. Third part: creates an algorithm to add a new signature and then add it to the Prohibitions room from unknown file and a new file and this we call updating the Prohibitions room.

References

[1] Ammar Ahmed E. Elhadi, Mohd Aizaini Maarof and Ahmed Hamza Osman (2012). Malware Detection Based on Hybrid Signature Behaviour Application Programming Interface Call Graph. American Journal of Applied Sciences 9 (3): 283-288, 2012

[2] Intrusion Detection System (2006) , Lecture Notes for Internet Security , Syracuse University

[3] Karin Ask (2006) Automatic Malware Signature Generation ,kth royal institute of technologey, sweeden

[4] Christopher M. Bishop (2006) Pattern Recognition and Machine Learning Springer Science Business Media, LLC

[5] M. Szczepanik and I. Jozwiak, “Detecting New and Unknown Malwares Using Honeynet,” WroclawUniversity of Technology, Institute of Informatics, Poland, 2010, p 173.

[6] Apel, M. ; Bockermann, C. ; Meier, M. " Measuring similarity of malware behavior ", Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on Digital Object Identifier:10.1109/LCN.2009.5355037 Publication Year: 2009 , Page(s): 891 - 89*8*

[7] Matunda Nyanchama & Sylvia Osborn,"Access Rights Administration in Role-Based Security System ",The Department of Computer Science, University of Western Ontario, London Ontario,Canada,1994.

[8] Abraham Silberschatz,Henry F. Korth,S. Sudarshan,"Database System Concepts ",McGraw-Hill ISBN 0-07-228363-7,2001.