Proc. of the Intl. Conf. on Advances in Computer Science and Electronics Engineering — CSEE 2013

ISBN: 978-981-07-5461-7

Improved Pattern Matching Algorithm for Intrusion Detection System

Shokoufeh Seifi, Hossien Gharaee

Abstract*—IDS is a protective system which identifies the occurring violations on the network and its performance is based on different methods. The most common method of IDS performance relies on patter compatibility in which the ordinary behaviors are recognized so that the abnormal behaviors are identified and specific patterns are found for them. In this paper, through a proposed algorithm, we aim at presenting an improved algorithm from combination of two different algorithms of HC and QWM and clustering the data base of attack signature using decision tree for the operations of pattern compatibility in IDS. The benefit of using this method is that compatibility can be managed in such a way that the extra searches are prevented and the best result is obtained during the overlapping of known attacks with main pattern in high speed.

Keywords*— Intrusion Detection Systems, Pattern Matching, QWM Algorithm, HC Algorithm.

I. *Introduction*

By definition, IDS is a protective system which identifies the occurring attacks on network. Its function is that it can control and report the attacks by diagnosing penetration including the steps of collecting data, surveying ports, and obtaining control of computers. Among other capabilities on IDS is the possibility of diagnosing abnormal traffic from outside to inside the network and announcing it to network manager or closing the suspicious contacts. In general, the tools of this system monitor the situation of network and are aware of connective nodes, traffic between path finders, the type of transmission traffic, message time and most parameters of network.

Shokoufeh Seifi

Department of Computer Engineering University of Tehran Kish International Campus Tehran, Islamic Republic of Iran

Hossien Gharaee

Department of Computer Engineering University of Tehran Kish International Campus Tehran, Islamic Republic of Iran

The commonest method of performance of these systems is based on pattern compatibility between signature database of known attack and received traffic through a network. When the compatibility operation occurs, the existence of known attack is specified. The inactive nature of system diagnosing penetration is what creates the power of conducting intelligent analysis of closed currents. This causes the system of diagnosing penetration to be situated in a good status to confirm the penetration cases. Diagnosing known attacks through signatures and compatibility operations of pattern between the created patterns [1].

styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow. In general, in this method, some data are first collected as sample. Then, the information such as attack characteristics, vulnerable points, penetrations and patterns are extracted from the samples and the data are classified in terms of the type of patterns so that the speed of compatibility operations is increased. Finally, comparing the signs and previously stored signs, we can attain a suitable pattern in the signature database of attacks so that they are identified and the diagnosis operation is terminated [2].

In this paper, we will have a review over the theoretical issues including the algorithms of pattern compatibility, setting up the decision-making tree and system of diagnosing penetration. Then, a brief history of other similar activities is mentioned and the proposed method is explained in details and compared with previous methods. At last, the conclusion is put forward.

II. Historical background

There are presented various methods and algorithms for selecting necessary features to do the pattern compatibility operations for systems of penetration diagnosis. The vest algorithms of setting decision-making tree and pattern compatibility are analyzed here.

A. *Hyper cuts algorithm*

This is the development of AC algorithm with tree feature and used for classifying compatibility operations and linear search among the packages in performance [3]. These algorithms function recursively which results in reduction time for this algorithm is O(nm + |Σ|) [4].

Proc. of the Intl. Conf. on Advances in Computer Science and Electronics Engineering — CSEE 2013

ISBN: 978-981-07-5461-7

B. *Quick Wu–manber algorithm*

This is the improvement of Wu-manber algorithm and can use a compatibility window to do the operations. The benefit of this algorithm is that it divides the similar windows in to smaller ones to reduce the comparisons and the distances between the windows.

While checking each package based on the respective pattern, this algorithm reviews the elementary characters. If there appears the compatibility (located in specific group), a report from the attack and compatibility will be given. The system fails to compare other characters which, in turn, increases the speed of performance (This provides the case of short and long compatibility pattern) [5].

Firstly short mode group P shorter is preprocessed. For every character char that appears in PI, its corresponding location in bitmap EXISTONE with the size of 256 will be tagged l. For every character charlchar2 that appears in P2, its corresponding location in bitmap EXISTTWO with the size of 256 X 256 will be tagged 1. In this way, during the searching stage whether the current character in the text matches a certain pattern in PI or P2 can be quickly determined by whether the value of the corresponding location in the bitmap is 1.

EXISTONE and EXISTTWO can be constructed according to the following method:

C. *Pattern and Text Matching Algorithm*

The Algorithm Process of Pattern and Text Matching in the Short Pattern Group

- 1) Take the current character tc in the text as the index to check EXISTONE (tc).
- 2) Take the two characters tc and tc+10f the current text as index to check EXISTTWO(tc, tc+ 1).
- 3) If both the value of EXISTONE(tc) and EXISTTWO(tc, tc+ 1) are 0, add 1 to the text pointers and tum to step one.
- 4) If the current pattern is case sensitive, examine whether they match under the case sensitive circumstances. If not, add 1 to the text pointers and tum to step one.
- 5) Report the match and finish the search.

D. *The Algorithm Process of Pattern and Text*

Matching in the Long Pattern Group For the long Pattern group P longen first of all, sorting the order of the Pattern; and then construct a bad character skip table BCSHIFT and a prefix hash table Hash.

- 1) Take the current character tc as the index, search for BCSHIFT(tc).
- 2) If the value of BCSHIFT (tc) is greater than 0, the move the text pointer to the right for BCSHIFT (tc) characters and tum to the first step.
- 3) Calculate the Hash function values of the current characters tc and tc+ 1 in the text. Take Hash (tc) as the index to search the Hash table. If the value is 0, add one to the text pointers and start the first step.
- 4) Examine all of the Patterns that have the same prefix with the current Pattern. If there is no Pattern matches the current text, add 1 to the text pointers and move the step one.
- 5) If the current pattern is case sensitive, examine whether they match under the case sensitive circumstances. If not, add 1 to the text pointers and turn to step one.
- 6) Report the match and finish the search. QWM algorithm treats the short mode first after separates it from the mode groups. The minimum mode length is more than 2, i.e. the maximum skip range is at least 3. Thus the time taken to inspect every packet is shortened

significantly and so the performance of pattern matching algorithms is improved.

III. Proposed method

In this method, it is suggested that the signature database of attacks be formed in cluster by using HC algorithm and setting a decision tree based on a suitable parameter. Then, the compatibility operations are performed by applying QWM algorithm. The benefits of using this method are that it is possible to enhance the speed of compatibility operations remarkably. Also, due to the infinite feature, we can apply a great volume of patterns to algorithms. To solve this problem, we can sort the results obtained from compatibility operations in QWM algorithm methods by Fuzzy systems and understand the attack.

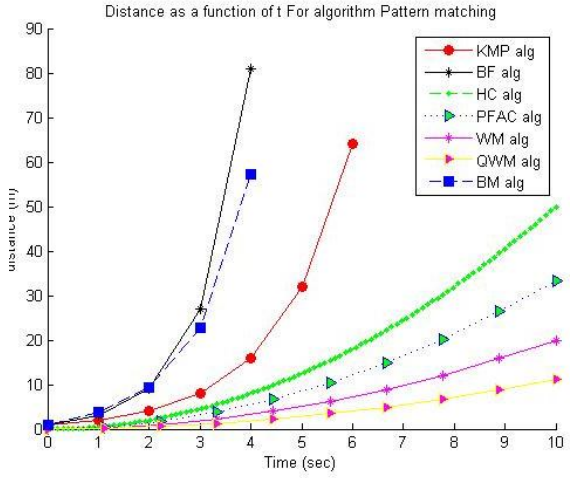

Figure1: Result of the Simulation Algorithm for Pattern matching.

Proc. of the Intl. Conf. on Advances in Computer Science and Electronics Engineering — CSEE 2013

ISBN: 978-981-07-5461-7

IV. Applying these algorithms

After the implementation of some algorithms of pattern compatibility in the similar conditions, the following figure was obtained. It is evident that HC and QWM algorithms are of great quality in pattern compatibility and this combination can increase the speed of operations [6].

In this paper, we have tried to reduce the processing time of packages to some extent through clustering the signature database of attacks and setting similar processes on the packages related to a cluster. To have a greater speed and higher quality in services, IDS's put the packages of similar service quality requirements in a cluster. The separation of packages to streams is dones on the basis of rules in IDS system [7]. These rules are listed in table in the systems of penetration diagnosis based on the priority. Therefore, clustering can result in the best outcome [8]. To prove this, we can compare the performance of the proposed method with combined algorithm iteratively.

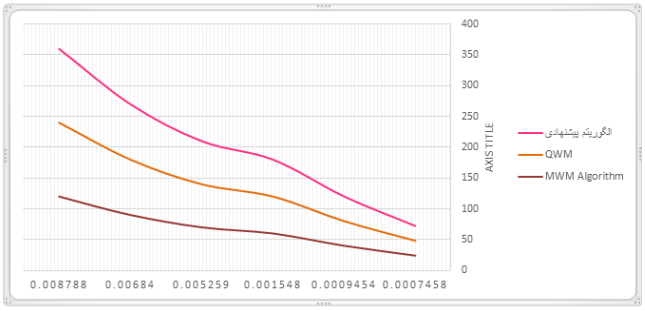

Figure2: Result of the implementation of this algorithms

The following figure shows the results of the implementation of the respective algorithm and the proposed algorithm based on the average values.

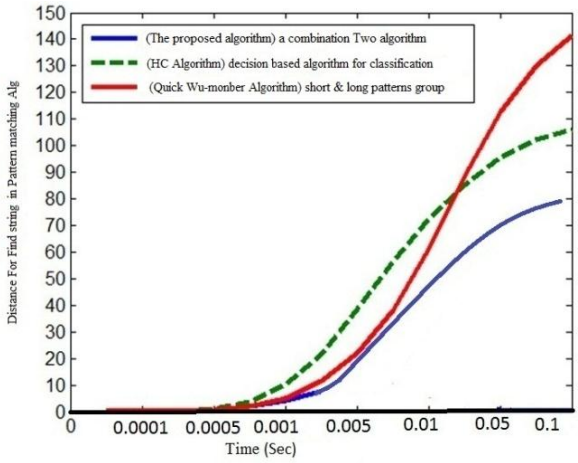

Figure3: Result of the implementation of this algorithm

V. Implementation of the proposal combined algorithm

After the implementation of applied algorithms in the operations of pattern compatibility, we did the simulations in 100, 200 and 300 runs, as the time difference of implementing the compatibility operations in QWM and the proposed algorithm is very small.

VI. Conclusion

In this paper, we tried to investigate the pattern compatibility operations in the systems of penetration diagnosis in a classified manner, so that we can identify the effect of electing this clustering by HC algorithm and the two stage compatibility in the speed of compatibility operations. As it is evident in implementation, the clustering of signature database of attacks has a great effect on the speed of pattern compatibility in using QWM algorithm. It is concluded that the combination of these two methods improves the operations speed in signature database of huge attacks in an acceptable level and brings about positive results for a system of diagnosing penetration based on abuse.

References

[1] Li Shu-zheng, The research of fast pattern matching algorithm, Based on Snort system (Master Thesis), Jilin University, In Chinese, 2009.

[2] Amir H. Payberah, "Auditing Intrusion Detection System using Mobile Agents" (Master thesis), Supervisor: Dr. Babak Sadeghian, Amirkabir University of Technology, Tehran, Iran, 2003.

[3] Alan Kennedy, B.Eng, Energy Efficient Hardware Accelerators for Packet Classification and String Matching, Degree of Doctor of Philosophy thesis, Dublin City University, School of Electronic Engineering , September 2010.

[4] David E. Taylor, Survey & Taxonomy of Packet Classification Techniques, Applied Research Laboratory, Department of Computer Science and Engineering Washington University in Saint Louis WUCSE-2004-24, May 10, 2004.

[5] Zhengqiang, An improved multiple patterns matching algorithm for intrusion detection, IIProc of International Conference on Computer Science and Information Technology ,IEEE 2010, PP. 611 -615.

[6] José M. Bande Serrano, José Hern ndez Palancar, String alignment pre-detection using unique subsequences for FPGA-based network intrusion detection , Contents lists available at SciVerse ScienceDirect, Computer Communications, 2012, PP. 35 (2012) 720–728.

[7] Bo Zhang, T. S. Eugene Ng, On Constructing Efficient Shared Decision Trees for Multiple Packet Filters, Department of Computer Science Rice University, 2010.

[8] Knut Reinert, Sandro Andreotti, Sequence and Structure Analysis, Advanced Algorithms in Bioinformatics (P4), Lecture / Tutorials / Software lab, 2011, PP.1000-1005.