Network Intrusion Detection Using Temporal Association Rules

Vidhu.A, Shibili.T

Abstract— Network security has become the most challenging issue in computer network field. Many researches have been done in detection of network intrusion in different layers of computer network. Most of the transport and network layer intrusion detection is based on previously specified signatures of packet fields without checking the relationship of one packet information with others. But these signatures cannot find novel attacks. Anomaly detection is used to find any behavior deviation from normal behavior. By using anomaly detection along with signature detection strength of IDS can be increased. This paper proposes an idea of mining temporal association rules between packet patterns of normal traffic. This method is making use of the temporal association of one packet pattern with the sequence of previous packet patterns to generate rules. These rules can detect any deviation from the normal activity as well as signature based intrusion. For optimization of rules a Reduced FP (False Positive) rate method is used. This method enables pruning the rules based on FP rate. Finally after several iterations a system with a reduced optimal FP rate will be obtained. Reduced error rate should be optimal so that number of detections is not much affected.

Index Terms— Temporal association rules, anomaly detection, flag pattern, TCP, Mining, packet.

1 Introduction

NETWORK intrusion detection system (NIDS) is the most efficient way of defending against network-based attacks aimed at computer systems. Most of the commercial implementations of NIDS are relatively insufficient and ineffective; it leads to the need for research on more dynamic intrusion detection systems [1].

Basically, there are two main types of intrusion detection systems: 1. Signature-based (SBS) or misuse based and 2. Anomaly-based (ABS) or behavior based.SBS performs intrusion detection by comparing new data with a data base of previously known attacks. It will generate an alarm if signatures are matched. On the other hand ABS system compares new data with a model describing normal behavior of the system. A considerable deviation from the model is identified as an anomaly. Disadvantage of signature based approach is that it cannot find new attack which is not familiar to the system.ABS can detect new unidentified attack and hence increase attack detection rate [2].

In this paper an intrusion detection system is proposed using temporal association rules. Here the proposed system is finding out relationship between continuous packets with in a time interval. Time stamped sequence category of temporality [9] is used in this proposal. Main idea behind our proposal is that some patterns of a packet header depend on its value in the previous packets. This approach finds association of a pattern with a sequence of previous patterns instead of finding association with just previous pattern. For example for a sequence A->B->C->D ,pattern D depends not only on C but on the sequence A->B->C->D. Here TCP flag patterns are considered as matrices for finding this association as 90% of the network traffic are TCP packets. This proposal can be used for finding signature and anomaly based intrusions. It is possible to expand this work on other matrices of packets in different layers.

Rest of the paper is organized as follows. Section 2 deals with related works. Section 3 provides the basics of temporal association rule mining and the mining method used in this approach. Section 4 presents proposed architecture for intrusion detection finally in section 5 conclusion and future works are given.

2 RELATED WORKS

Many studies have been done on intrusion detection systems. Hong Han et.al [3] believed that signatures are almost contained in payload of packet and developed a technique to mine content of network packet and get signatures using an algorithm named signature apriori. Anita Jones and Song Li [4] use system call sequence as signature and established temporal association based on time interval between the calls. In our approach temporal associations are not generated in this way. Instead here associations are found between sequences of continuous events within a time interval.

Xiaolei Li and Xun Li [5] proposed anomaly detection scheme for LAN using association rule mining by maintaining a record of each host, number of communicating host, number of packets send, number of packets received etc.. Estevez-Tapiador et.al. [6] made a stochastic model for anomaly detection by using markov chain and by considering only binary relationship of sequences. But this proposal finds relationship with all preceding patterns and hence is expected to have more accuracy than the stochastic method [6]. Xiaohui Cui et al. [7] proposed multistage attack detection using data mining by using temporal pattern of attacks to predict future attacks. ADAM [8] used apriori algorithm for frequent dataset mining from normal frequent datasets. Then while detection, it executes an on-line algorithm to discover most recent frequent connections, then it compares it with known mined training normal datasets and it rejects those recent connections which look normal. The rest of connections are marked as known attacks, unknown attacks and false positive using a classifier which is previously trained for classifying. The central theme of MADAMID [9] approach is to apply data mining programs to the extensively gathered audit data to compute models that accurately capture the patterns of intrusions and normal activities.

3 Temporal Association Rule Mining

Association rules find correlations between different events or patterns from the large database. An association rule A=>B suggests that presence of item B in the database depends on the presence of item A in the database. Out of four categories of temporality time stamped sequence relationship before and after is used for the temporal association rule presented in this paper. In temporal association, A=>B suggests that presence of B usually occurs after the presence of A.

The main concept of association rule is same for temporal association rule. But algorithms for association rule cannot be directly applied for temporal association rules as in association rules there is no order concept [10].

A time stamped sequential temporal association rule that is used in this proposal is A=>B, A and B are frequent temporal patterns such that A is the preceding sequence of B. Confidence of temporal association A=>B is defined as

$$\mathrm{conf}(\mathrm{A} => B) = \frac{\sigma(\mathrm{AB})}{\sigma(A)}$$

Support of the pattern is defined as [10]

$$\sigma(\mathrm{X}) = \frac{\mathrm{F_X}}{\mathcal{D}}$$

Where $\mathrm{F_X}$ is the frequency of the event in the database and $\mathcal{D}$ is the total number of events in the database. A type of negative temporal association rule method is used in this work. Sequences that are having confidence less than the min-confidence are selected as rules.

4 Proposed Method

Out of four categories of temporality [10] sequence relationship before and after is used for the temporal association rule presented in this paper. Temporal associations are established from the patterns generated after preprocessing of the network data. The generated temporal association rules are used to define anomalies. For optimizing these rules a Reduced FP rate method will be used.

A. System Architecture

Normal Network data for initial training will be taken from global repository. Later input will be taken directly from the environment using a packet analyzer. Several iterations with different input sets of these types will be conducted for optimizing our rules and hence making the system ready for intrusion detection. This will make the system suitable for the behavior of the environment. Packet analyzer will give the required packet information from the network. It is possible to filter the TCP header information. From the normal flow patterns temporal association rules are made by using the temporal association mining algorithm on them. Rules are finalized by finding rules with confidence value less than a pre specified min-confidence value in each step of processing. Anomaly detector will check the new TCP flag patterns with rules. Those patterns which are matched with the rules are marked as anomaly. Later, for pruning of the rules a Reduced FP rate method can be used.

The error rate can be reduced by changing the min support and min confidence values which were selected manually during first iteration of training.

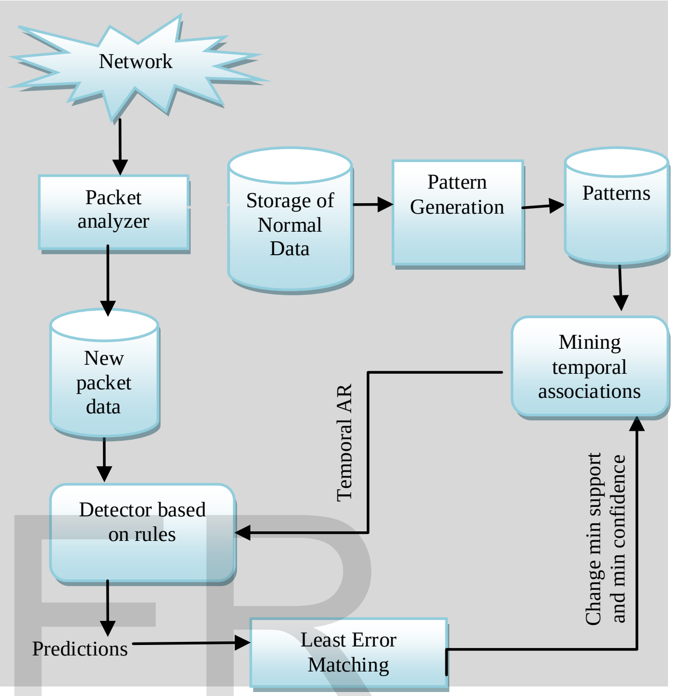

Fig.1.System Architecture

B. Pattern Generation

Table.1.flag patterns

	flag	Binary	Pattern
CWR	Congestion window reduced	10000000	P128
ECE	ECN- Echo	01000000	P64
URG	Urgent	00100000	P32
ACK	Acknowledgement	00010000	P16
PSH	Push	00001000	P8
RST	Reset	00000100	P4
SYN	Syn	00000010	P2
FIN	Fin	00000001	P1

Table.1. shows patterns for different flag combinations like P1, P2, P3…P256.It shows patterns in which only one flag is set per pattern, but all combinations of flag patterns can be made as mentioned above. For example pattern P3 will be with SYN and FIN flags set. From these 256 patterns, temporal associations should be mined as flag pattern of each TCP packet depends on flag patterns of previous sequences.

C. *Mining rules from patterns*

Let K = {t1, t2, t3…..tn} be the set of traces captured with each trace having sequence of patterns

Let S_0 = {P1, P2, P3….P256} be the set of possible Patterns

Let R= {{Rule set1}, {Rule set2}, {Rule set3}…, {Rule set k}}

Where k is the number of stages

Each Rule set is initialized as Rule set i=Φ, where i ∈ {i…k}

Temporal association rules are mined in different stages from the patterns and all traces of capture.

Stage 1: In stage 1 total number of occurrences of each pattern in S_0 in every traces of capture will be found out and kept as the frequency. Then support of each pattern can be calculated by the formula [10]

$$\sigma(\mathrm{X}) = \frac{\mathrm{F_X}}{\mathcal{D}}$$

Where, $\mathrm{F_X}$ is the frequency of pattern and $\mathcal{D}$ is the total number of traces of capture.

Now support value of each pattern will be compared with the min-support which is specified manually. Those patterns which have support value less than this specified min-support will be added to Rule set1 as we need to find negative temporal associations for making rules. Flag patterns that are having very low frequency in training data will come under this Rule set 0.Flag patterns that are not allowed as per the usual signature based IDS will also be come under Rule set1 because their frequency will be very low or null. Hence patterns containing in Rule set1 will be able to detect signature based anomalies also. These patterns will be removed from the pattern set S_0 as they will not occur in any frequent sequence of higher order. So they should be eliminated in this stage itself.

Stage 2: Stage 2 will find out all binary relationships in the form P_i=>P_j from S_0 and put them in a set S_1. After that support should be calculated as in the first stage. Using the formula for confidence in rich temporal association rule, confidence of each association is calculated

$$\mathrm{conf}(\mathrm{X} => Y) = \frac{\sigma(\mathrm{XY})}{\sigma(X)}$$

All associations which are having confidence less than the min-confidence are added to Rule set 2. For example, if P2=>P3 is having confidence less than min-confidence, then

{Rule set2}= {Rule set2} U P2=>P3

These associations should be removed from S_2 as it will not form part of any frequent higher order sequence. During detection if a sequence gets matched with a rule in Rule set 2, then it will be marked as anomaly. So there is no need to find higher order sequence which will start with these rules.

Stage 3: In stage 3, algorithm will find ternary association of the form A=>B=>C from S_1 and S_0 and will be kept it in set S2. After that confidence values are to be found. Then rules are made with the patterns having confidence less than min-confidence and remove those from set S2 as in stage 2.

If P1=>P3=>P5 is having confidence less than min-confidence, then

{Rule set3}= {Rule set3} U P1=>P3=>P5

This process will continue in more stages until a stage is reached where the number of sequences that are having confidence value greater than or equal to min-confidence becomes less than a limiting factor. Value of this limiting factor is also given manually which can be optimized after several iterations. It is very difficult to find out exact minimum values for min-support, min-confidence and for limiting factor. Strength and accuracy of our rules depend upon these minimum values. For that first minimum values are set manually, then for refining it Reduced FP rate method can be used. FP Rate is given by the

$$FP\ Rate = \frac{Number\ of\ false\ positives}{Total\ number\ of\ training\ patterns}.$$

The values of min-support, min-confidence and limiting factor are chosen in such a way that minimum error rate and acceptable rate of error detection are obtained. These final values can be used in mining the accurate temporal association rules. For detecting anomaly each incoming flag pattern will be checked with rules of these Rule sets in order i.e. first it will be matched with all rules in Rule set1 and if they don't match then it is matched with Rule set2 and so on. Those sequences which match with the rules will be marked as anomalous.

5 Conclusion

This paper has proposed a new method for anomaly detection using temporal association rules on packet patterns. As the values of some matrices like flag patterns depend on a sequence of previous patterns, sequence temporal associations are mined. Rules generated by our system can identify both signature based and behavior based anomalies of flag patterns. For optimizing our rules a Reduced FP rate is used. In future it can be extended to find more matrices of layer 3-4 packets that show temporal association so that strength of this system can be increased.

.

REFERENCES

[1] Mohammad Sazzadul Hoque,, Md. Abdul Mukit and Md. Abu Naser Bikas, "*An Implementation of Intrusion Detection*", International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012

[2] Anna Sperotto, "Flow Based Intrusion Detection", CTIT Ph.D.-thesis Series No. 10-180,Centre for Telematics and Information Technology,University of Twente

[3] Hong Han,Xian-Liang Lu,Li-Yong Ren,"*Using Data Mining To Discover Signatures In Network-Based Intrusion Detection*", Proceedings of the First International Conference On Machine Learning and Cybernetics,Beijing,4-5 November 2002

[4] Anita Jones, Song Li, "*Temporal Signatures for Intrusion Detection,*" Computer Security Applications Conference,2001.ACSAC 2001.Proceedings 17th Annual.

[5] Xiaolei Li,XUn Li,"*Local Area Network Anomaly Detection using Association Rules Mining*", Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09.

[6] Jusn M.Estevez-Tapiador,Pedro Garcia-Teodoro,Jesus E.Diaz-Verdejo,"*Stochastic Protocol Modelling for Anomaly Based Network Intrusion Detection*",Proceedings of the First IEEE International Workshop on Information assurance(IWIA'03).

[7] Rajeshwar Katipally,Xiaohui Cui,Li Yang, "*Multi stage attack Detection system for Network Administrators using Data Mining*",CSIIRW '10 Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research

[8] Daniel Bara,Julia Couto Sushil Jajodia Nnngning Wu" *ADAM:A Testbed for Exploring the Use of Data Mining in Intrusion Detection"* Newsletter ACM SIGMOID Record Volume 30 Issue 4 December 2001

[9] W. Lee and S. J. Stolfo. "*A framework for constructing features and models for intrusion detection systems.*" ACM Transaction on Information and System Security, 3(4):227–261, Nov. 2000

[10] John F. Roddick,Myra Spiliopoulou, "*A Survey of Temporal Knowledge Discovery Paradigms and Methods*", IEEE Transactions on knowledge and data engineering, VOL. 14, NO. 4, July/August 2002

[11] Edi Winarko, John F. Roddick,"*ARMADA – An algorithm for discovering richer relative temporal association rules from interval-based data*", Data & Knowledge Engineering. Volume 63 Issue 1,October 2007

.

.