

Fuzzy Frequent Episodes for Real-Time Intrusion Detection

Jianxiong Luo, Susan M. Bridges, Rayford B. Vaughn, Jr.
Department of Computer Science
Mississippi State University
Box 9637 Mississippi State, MS 39762 USA

Abstract

Data mining methods including association rule mining and frequent episode mining have been applied to the intrusion detection problem. In other work, we have introduced modifications of these methods that mine fuzzy association rules and fuzzy frequent episodes and have described off-line methods that utilize these fuzzy methods for anomaly detection from audit data. In this paper we describe another extension that uses fuzzy frequent episodes for near real-time intrusion detection. We first define fuzzy frequent episodes and then describe experiments that explore their applicability for real-time intrusion detection. Experimental results indicate that fuzzy frequent episodes can provide effective approximate anomaly detection.

I. INTRODUCTION

The pervasiveness of computer networks in modern life and recent high-profile attacks against major Internet sites have prompted wide interest in improved security of computer networks including better methods for intrusion detection. Approaches to network intrusion detection are typically categorized as misuse detection or anomaly detection. Misuse detection systems look for known patterns of attack while anomaly detection systems look for deviations from normal behavior. A variety of data mining methods have been used in an attempt to automate the acquisition of patterns for both misuse and anomaly detection. Many research groups now advocate the fusion of information from a number of different detection modules [e.g. 2, 4]. We describe an anomaly detection approach based on fuzzy frequent episodes that can provide evidence of intrusions in near real-time.

Frequent episodes have been used by several research groups to represent frequent sequential patterns in temporal data streams [4, 6]. When the sequential patterns involve quantitative values, some sort of quantization is used to yield discrete attribute values. The quantization process can produce an artificial sharp boundary problem. Kuok, Fu, and Wong [3] addressed this problem for association rule mining by developing methods for mining fuzzy association rules. We have extended their approach to mine fuzzy frequent episodes [5]. In our previous work, we have used fuzzy association rules and fuzzy frequent episodes for off-line anomaly detection. In this paper, we explore the applicability of fuzzy frequent episodes for near real-time intrusion detection from network audit data. We first present a brief overview of frequent episodes and then describe our fuzzy extension to this data mining method. We then describe our method for real-time intrusion detection using fuzzy frequent

episodes and present preliminary experimental results that demonstrate the effectiveness of the approach.

II. THEORY

Frequent Episodes

Mannila and Toivonen [6] described an algorithm to discover simple serial frequency episodes from event sequences based on minimal occurrences. In Mannila and Toivonen's method [6], suppose $S = \{E_1, E_2, \dots, E_n\}$ is an event sequence of n events and $A = \{a_1, a_2, \dots, a_m\}$ is the set of all the event attributes. Each event $E = \{E.a_1, E.a_2, \dots, E.a_m\}$ in S consists of m values for all the event attributes. E is also associated with a timestamp denoted by ET . A simple serial episode $P(e_1, e_2, \dots, e_k)$ represents a sequential occurrence of k event variables where each e_i ($1 \leq i \leq k$) is an event variable and for all i and j ($1 \leq i < j \leq k$), $e_i T < e_j T$. Usually, k is much smaller than n , so $1 \leq k \ll n$. We use e^q to represent an event variable consisting of q event attributes, i.e., $e^q \{attr_1=v_1, attr_2=v_2, \dots, attr_q=v_q\}$ where $\{e^q.attr_1, e^q.attr_2, \dots, e^q.attr_q\} \subseteq A$ and $1 \leq q \leq m$ and each v_i ($1 \leq i \leq q$) is a value from the domain of attribute $attr_i$. An event variable e^q is said to have an occurrence in an event E if for all i ($1 \leq i \leq q$), $E(e^q.attr_i) = e^q.v_i$.

Mannila and Toivonen [6] further an episode $P(e_1, e_2, \dots, e_k)$ as occurring in interval $[t, t']$ if $t \leq e_1 T$ and $t' \geq e_k T$. An occurrence of $P(e_1, e_2, \dots, e_k)$ in interval $[t, t']$ is defined as minimal if there does not exist another of occurrence of $P(e_1, e_2, \dots, e_k)$ in a subinterval $[u, u'] \subset [t, t']$. Given a threshold of *window* (representing timestamp bounds), the frequency of $P(e_1, e_2, \dots, e_k)$ in the event sequence S is the total number of its minimal occurrences in any interval smaller than *window*. In order to find "frequent" episodes, a second threshold, *minfrequency*, is used. An episode $P(e_1, e_2, \dots, e_k)$ is called frequent if $frequency(P)/(n-k+1) \geq minfrequency$. Since in our domain $k \ll n$, an episode is considered frequent if $frequency(P)/n \geq minfrequency$. Mannila and Toivonen's algorithm [6] for mining frequent episodes is similar to the Apriori algorithm [1] except for the difference between calculating episode frequencies and calculating itemset supports. Like association rules, episode rules can be directly established from frequent episodes. Given a frequent episode $P(e_1, e_2, \dots, e_k)$, there will be a $k-1$ non-empty ordered superepisodes $P(e_1, e_2, \dots, e_i) \supset P$ where $1 \leq i \leq k-1$. Given another

* This material is based upon work supported by the Army Research Laboratory under Contract No. DAAD1701C00110101005 and by DOD DEPSCoR under Contract No. N000140110678.

threshold *minconfidence*, a simple serial episode rule of the form $P_i @ Q_i, c, s, w$ can be constructed where:

$$\begin{aligned} P(e_1, e_2, \dots, e_i) &\hat{=} P, \\ Q(e_{i+1}, e_{i+2}, \dots, e_k) &= P - P_i \hat{=} P, \\ s &= \text{frequency}(P) \hat{=} \text{minfrequency}, \\ c &= (\text{frequency}(P) / \text{frequency}(P_i)) \hat{=} \text{minconfidence}, \text{ and} \\ w &= \text{window}. \end{aligned}$$

In our work, P_i and Q_i are required to be within the same time window w . The last episode rule $P_{k-1} @ Q_k$ is of the most interest since it can be used to predict the k^{th} event given the previous $k-1$ events. We use this form of episode rules for real-time intrusion detection.

Fuzzy Frequent Episodes

The need to develop fuzzy frequent episodes comes from the involvement of quantitative attributes in an event. We have modified the methods of Mannila and Toivonen [6] to incorporate the representation of quantitative attributes via fuzzy sets. Extending the notation of [6] in a manner similar to [3], given the set of event attributes $A = \{a_1, a_2, \dots, a_m\}$, each attribute a_j ($1 \leq j \leq m$) may be categorical or quantitative (fuzzy). Suppose $f(a_j)$ represents the maximum number of categories (if a_j is categorical) or the maximum number of fuzzy sets (if a_j is fuzzy), and $m_{a_j}(l, v)$ represents the membership degree of v in the l^{th} category or fuzzy set of a_j . If a_j is categorical, $m_{a_j}(l, v) \in \{0, 1\}$. If a_j is fuzzy, $0 \leq m_{a_j}(l, v) \leq 1$. For an event attribute, its categories or fuzzy sets can be mapped to consecutive integers. This allows an event variable e^k to be expressed as $e^k \{attr_1 = c_1, attr_2 = c_2, \dots, attr_k = c_k\}$ where $e^k \{attr_1 = c_1, attr_2 = c_2, \dots, attr_k = c_k\} \subseteq A$ and for all j ($1 \leq j \leq k$), $1 \leq c_j \leq f(a_j)$. We define two event variables $e^p \{attr_1 = c_1, attr_2 = c_2, \dots, attr_p = c_p\}$ and $e^q \{attr_1' = c_1, attr_2' = c_2, \dots, attr_q' = c_q\}$ to be homogeneous, if $\{e^p.attr_1, e^p.attr_2, \dots, e^p.attr_p\} = \{e^q.attr_1', e^q.attr_2', \dots, e^q.attr_q'\}$, which also indicates that $p = q$. It is obvious that an event variable is homogeneous to itself.

Given an event $E = \{E.a_1, E.a_2, \dots, E.a_m\}$, $E.a_j$ ($1 \leq j \leq m$), represents a value of the j^{th} attribute and can be mapped to $\{(l, m_{a_j}(l, E.a_j)) \mid \text{for all } l, 1 \leq l \leq f(a_j)\}$. However, if a_j is fuzzy, the sum of all the memberships may not always equal one. A normalization process is used as follows:

$$m'_{a_j}(l, E.a_j) = \begin{cases} \frac{m_{a_j}(l, E.a_j)}{f(a_j)} & \text{if } a_j \text{ is fuzzy} \\ \sum_{l=1} m_{a_j}(l, E.a_j) & \\ m_{a_j}(l, E.a_j) & \text{if } a_j \text{ is categorical} \end{cases}$$

For an event variable $e^k \{attr_1 = c_1, attr_2 = c_2, \dots, attr_k = c_k\}$ where $1 \leq k \leq m$, its occurrence in E is no longer restricted to the values 0 or 1. Rather, it is defined as:

$$\text{occurrence}(e^k, E) = \prod_{j=1}^k m'_{e^k.attr_j}(e^k.c_j, E.(e^k.attr_j)).$$

The minimal occurrence of an episode is the product of the occurrences of its event variables. This means an event E may support several event variable occurrences due to the introduction of fuzzy sets. However, a side effect may arise due to very small membership values. For example, consider the event sequence $\{E_1, E_2, E_3\}$ within the *window* threshold. A, B, C , and D are event variables in which A and B are homogeneous but $A \neq B$, and C and D are homogeneous but $C \neq D$. Suppose we have the following occurrences of the event variables:

$$\begin{aligned} \text{occurrence}(A, E_1) &= 0.8, \\ \text{occurrence}(B, E_1) &= 0.2, \\ \text{occurrence}(A, E_2) &= 0.1, \\ \text{occurrence}(B, E_2) &= 0.9, \\ \text{occurrence}(C, E_3) &= 0.9, \text{ and} \\ \text{occurrence}(D, E_3) &= 0.1. \end{aligned}$$

Then the minimal occurrence of episode $\{A, C\}$ will become 0.09 based on $\{E_2.A, E_3.C\}$ and not 0.72 that would be the value based on the non-minimal event sequence $\{E_1.A, E_3.C\}$. This means that the occurrence of an event variable with a very small membership may change the minimal occurrence of an episode in the event sequence.

To address this problem, we introduce another user-specified threshold *minoccurrence* to represent the smallest allowable occurrence of an event variable. So, given an event variable e^k , if $\text{occurrence}(e^k, E) < \text{minoccurrence}$, it will not be considered to have occurred in E . The following normalization process will be conducted to account for the deletion of these low membership occurrences:

$$\begin{aligned} &\text{if } (\text{occurrence}(e^k, E) < \text{leastoccurrence}) \\ &\quad \text{occurrence}(e^k, E) = 0 \\ &\text{else} \end{aligned}$$

$$\text{occurrence}(e^k, E) = \frac{\text{occurrence}(e^k, E)}{\sum_{e^q} \text{occurrence}(e^q, E)}$$

Here every e^q is homogeneous to e^k and $\text{occurrence}(e^q, E) \hat{=} \text{minoccurrence}$. For instance, if $\text{leastoccurrence} = 0.2$, E_1 will contribute 0.8 to A and 0.2 to B , E_2 will contribute 1 to B , and E_3 will contribute 1 to C . If leastoccurrence is set above 0.5, for any event, only one event variable will be counted with a normalized occurrence of one.

Other than the difference in calculating the frequency (or minimal occurrence) of an episode, the algorithm for mining fuzzy frequent episodes is similar to Mannila and Toivonen's algorithm [6] for mining frequent episodes. More details are available in [5].

III. REAL TIME INTRUSION DETECTION

Lee, Stolfo, and Mok [4] describe the use of frequent episodes to select features that will be significant for construction of real-time intrusion detection systems. We have previously described the use of fuzzy frequent episodes

in a similar fashion [5]. In addition, we have investigated the use of fuzzy episode rules directly for near real-time intrusion detection. In the Time-based Inductive Machine (TIM) proposed by Teng, Chen, and Lu [7], a sequential pattern with 100% certainty can be used to detect anomalies. For example, given a normal pattern like $A \rightarrow B \rightarrow C \rightarrow D$ ($D=100\%$), the sequence $A \rightarrow B \rightarrow C \rightarrow E$ will be marked as an anomaly since it is believed that $A \rightarrow B \rightarrow C$ is always followed by D . In a similar fashion, we introduce the idea of using fuzzy episode rules with *high* confidence (e.g., ≥ 0.8) for anomaly detection.

Consider the event sequence $S=\{E_1, E_2, \dots, E_n\}$, the current event E_n following S , and a fuzzy episode rule of the form $R: e_1, \dots, e_{k-1} @ e_k, c, s, w$, where $k \geq 1$ and all $e_i (1 \leq i \leq k)$ are event variables. For the episode $\{e_1, \dots, e_{k-1}\}$, if its minimum occurrence in S is x ($x > 0$), it then can be predicted, with confidence c , that $\{e_1, \dots, e_{k-1}, e_k\}$ will also have a minimal occurrence in $S+\{E_n\}$ with the constraint that e_k occurs in the event E_n . Suppose the minimal occurrence of the episode in the sequence $S+\{E_n\}$ is y , $y/x \geq c$ should also hold. In this case, event E_n is said to match the episode rule R . On the other hand, like TIM, if the episode $\{e_1, \dots, e_k\}$ has no minimum occurrence in S , or if $x = 0$, the episode rule R is said to be mismatched and indicates an anomaly. Our experiments demonstrate that a large *window* threshold (e.g., $15 \text{ seconds} \leq w \leq 30 \text{ seconds}$) decreases the probability of mismatches.

The set of episode rules mined from training data is used to represent normal patterns. If the current event E_n does not match any episode rule, it will be marked as an anomaly with some degree of belief. The confidence of an episode rule is usually less than 1 and provides a measure of the strength of the evidence of an anomaly. This approximate intrusion detection method evidence for anomalies that can be combined with evidence from other detection modules.

IV. EXPERIMENTAL RESULTS

The experiments described in this section were designed to demonstrate the applicability of fuzzy episode rules in real-time intrusion detection. In the first experiment, intrusions of the probing type were simulated by use of *mscan* in the network of the Computer Science Department of Mississippi State University. *Mscan* is a software tool which can be used to scan multiple systems. Fuzzy episode rules were mined from 3 hours of training data with no intrusions for the feature of PN (the number of different destination ports during last 2 seconds). Since the simulated intrusions usually take 1 to 1.5 minutes, test data sets were established by collecting network traffic data for 3 minutes, with the goal of covering the duration of every simulated intrusion. Six test data sets were collected during 13:00—14:00, Tuesday, 30 March, 1999 and 13:00—14:00, Wednesday, 31 March, 1999. Among the six test data sets, T1', T2', and T3' represent normal data sets, and T4', T5', and T6' represent data sets

with simulated *mscan* intrusions. The anomaly percentage of every test data set is calculated as follows. Suppose we are given a sequence of n events for testing. An event will be marked as an anomaly if, in the set of episode rules representing normal behavior, there is no episode rule that it matches. If the total number of anomaly events is m ,

$$\text{anomaly percentage} = \frac{m}{n} * 100\%.$$

Figure 1 shows results of the first experiment. Rule sets were mined using *minconfidence* = 0.8, *minsupport* = 0.1, *minoccurrence* = 0.3 and *window* = 15s.

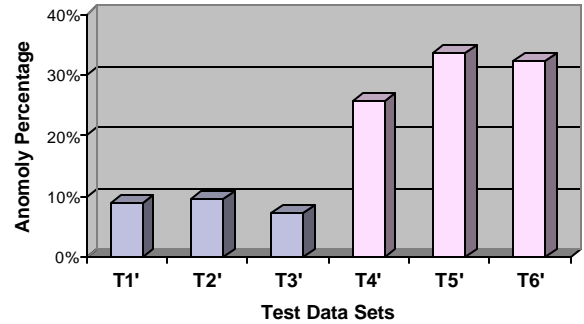


Fig. 1 Anomaly Percentages of Different Test Data Sets

These results demonstrate clear differences in the anomaly percentages of normal data and intrusion data. Since there are no simulated intrusions in T1', T2', and T3', the anomaly percentages for this data actually represent false positive error rates. Further analysis of the results from T4', T5', and T6' shows that all have false positive error rates below 10%. However, the false negative error rates are relatively high (about 40%). There are several reasons. First, only one feature, i.e., PN, is used. Second, the simulated intrusions are not evenly distributed along time according to this feature.

A second experiment was conducted in order to compare the intrusion detection performance, especially the false positive error rate, between fuzzy episode rules and non-fuzzy episode rules (by use of intervals). The same training data set and six test data sets as in the first were used. Both fuzzy episode rules and non-fuzzy episode rules were mined from training data for the feature of PN (the number of different destination ports during last 2 seconds), which was divided into three fuzzy sets (for fuzzy episode rules) or three intervals (for non-fuzzy episode rules): *LOW*, *MEDIUM*, and *HIGH*. Figure 6.10 shows a comparison of the false positive error rates on test data sets between fuzzy episode rules and non-fuzzy episode rules.

The experimental results demonstrate that the false positive error rates from fuzzy episode rules are less than for non-fuzzy episode rules. That is to say, the error rate of predicting a normal behavior as an intrusion is much lower for fuzzy episode rules than non-fuzzy episode rules.

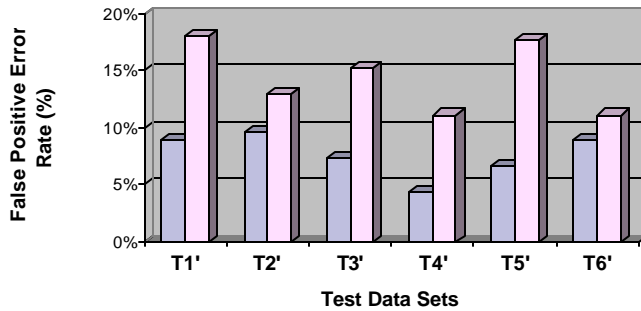


Fig. 2 Comparison of False Positive Error Rates of Fuzzy Episode Rules and Non-Fuzzy Episode Rules

We have also conducted experiments to determine the effect of the *minconfidence* threshold on the false positive and false negative error rates. Higher *minconfidence* values were found to result in higher false positive error rates and lower false negative error rates. Our experiments have also shown that a higher *minconfidence* value will cause many more mismatches. This is because a higher *minconfidence* value will reduce the number of episode rules learned from training data, resulting in rule numbers so low that the rules are not be able to cover patterns representing all normal behavior.

V. CONCLUSIONS AND FUTURE WORK

Intrusion detection is an important but complex task for a computer system. Many AI techniques have been widely used in intrusion detection systems. Data mining methods are capable of extracting patterns automatically and adaptively from a large amount of data. Association rules and frequent episodes have been used to mine training data to establish normal patterns for anomaly detection. We have shown that when the data involve quantitative attributes, the higher level of abstraction offered by fuzzy the fuzzy counterparts of these methods generates more abstract and flexible patterns for anomaly detection.

There are two main reasons for introducing fuzzy logic for intrusion detection. First, many quantitative features are involved in intrusion detection. Fuzzy set theory provides a reasonable and efficient way to categorize these quantitative features in order to establish high-level patterns. Second, security itself is fuzzy. For quantitative features, there is no sharp delineation between normal operations and anomalies. Fuzzy episode rules allow one to create the high-level sequential patterns representing normal behavior.

We have modified the procedure of Mannila and Toivonen [6] for mining frequent episodes to learn fuzzy frequent episodes. We use fuzzy frequent episodes to extract patterns for temporal statistical measurements at a higher level than the data level. We present a real-time intrusion detection method that uses fuzzy episode rules and demonstrate that the

use of fuzzy frequent episodes reduces the error rate as compared to non-fuzzy frequent episodes. Dickerson and Dickerson [2] have also developed methods for using fuzzy logic to profile networks behavior. Their methods differ from ours in that they use data mining primarily for audit date reduction and their fuzzy rules are handcrafted.

In order to further evaluate the utility of this method, we plan to conduct experiments with additional audit data features. In other work, we have exploited genetic algorithms to tune the membership functions for fuzzy association rules and for feature selection. We will also apply these methods to the membership functions for fuzzy frequency episodes. The architecture of our intelligent intrusion detection system as described in [5] supports the fusion of results from multiple detection modules by a decision module. We are exploring the use of fuzzy cognitive maps for this fusion process.

REFERENCES

- [1] Agrawal, R., H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, p 307-328. AAAI/MIT Press, 1996.
- [2] Dickerson, John E. and Julia A. Dickerson. 2000. Fuzzy network profiling for intrusion detection. *Proceedings of NAIFIPS 2000*. 301-306.
- [3] Kuok, C., A. Fu, and M. Wong. 1998. Mining fuzzy association rules in databases. *SIGMOD Record* 27(1): 41-6.
- [4] Lee, W., S. Stolfo, and K. Mok. 1998. Mining audit data to build intrusion detection models. In *Proceedings of the fourth international conference on knowledge discovery and data mining held in New York, New York, August 27-31, 1998*, edited by Rakesh Agrawal, and Paul Stolorz, 66-72. New York, NY: AAAI Press.
- [5] Luo, Jianxiong and Susan M. Bridges. 2000. Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection. *International Journal of Intelligent Systems* 15: 687-703.
- [6] Mannila, H., and H. Toivonen. 1996. Discovering generalized episodes using minimal occurrences. In *Proceedings of the second international conference on knowledge discovery and data mining held in Portland, Oregon, August, 1996*, by AAAI Press, 146-51.
- [7] Teng, H., K. Chen, and S. Lu. 1990. Adaptive real-time anomaly detection using inductively generated sequential patterns. In *Proceedings of 1990 IEEE computer society symposium on research in security and privacy, Oakland, California, May 7-9, 1990*, 278-84.