A Machine Learning Model for Temporal Anomaly Detection in Venture Cyber-Security

Geethumol P.V[1] Sithara Sasidharan[2]
[1,2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Mount Zion College of Engineering, Kadammanitta, Kerala, India

Abstract— Anomaly detection is emerging as an important complement to signature-based methods for venture network defence. In this paper, discuss an persevering structure to describe anomaly detection in large enterprise networks. This structure provides structure in the basis of a regression-based anomaly detection method. This paper describes Markov-Chain model for detecting anomalies in network.

Key words: Machine Learning, Stationary Markov Chain Model, Anomaly Detection

I. INTRODUCTION

As the complexity of cyber-attacks against venture networks increases, the usage of currently using defence mechanisms based mainly on signature matching techniques becomes the utility of existing defences based primarily on signature matching methods becomes increasingly uncertain. This statement is readily proved to by both the increasing number of importance and openly reported attacks, and the presence of advanced persistent threats [1].

The new methods are required to match existing defences. There is an evolving research community developing statistical and machine learning methods for anomaly detection in venture network monitoring and defence [2]. Cultured attackers are familiar of a range of defences, and appear to be able to pierce signature-based defence rather easily. Instead of concentrating on precise strong signatures of attack behavior, to use anomaly detection to find unusual behaviors with respect to a "normal" background. These anomalies can then be ranked and triaged for further investigation.

The focus of this paper, sketch a simple machine learning procedure for identifying anomalous temporal behavior from a selection of initiative network data behavior observed to be common to these data sources.

This paper is structured as follows. Section II describes the data set. The persistent structure observed in the data and the machine learning methodology is discussed in Section III.

II. DATA SET

This research utilizes the recently released "comprehensive, multi-source cyber-security events" data set [4], from Los Alamos National Laboratory. This data set collects together event data from a variety of collection mechanisms over the same period of time.

The data set particularly with two of these sources: Flow and Auth. The Flow data relates to Netflow, a router based protocol that provides summaries of IP flow events between devices. In this case, the data only reports internal traffic. The Auth data relates to authentication events collected from log files.

The data is anonymized, and for either data source are only concerned with two elements: timestamps, and identifiers .In the case of Flow data, each record reports a timestamp and two device identifiers, corresponding to their IP address. With authentication events, each record features a timestamp, a user id, and two computer identifiers. In both data sources, events are detailed to second accuracy.

The study of data set uses three types of interaction derived from Auth and Flow. For the Auth data, have divided on user/user interactions (Auth-U) as well as on user/computer interactions (Auth-C) in the recorded events.

III. METHODOLOGY

In this paper, discuss the anomaly detection using Markov Chain model. This is a machine learning technique for anomaly detection.

A. Markov Chain Model

A discrete-time stochastic process specifies how a random variable changes at discrete points in time. Let X_t denote a random variable representing the state of a system at time *t*, where $t = 0, 1, 2 \ldots$s. A stationary Markov chain is a special type of discrete time stochastic process with the following assumptions:

- The probability distribution of the state at time $t+1$ depends on the state at time *t*, and does not depend on the previous states leading to the state at time *t*;
- A state transition from time *t* to time $t+1$ is independent of time.

Let p_{ij} denote the probability that the system is in a state j at time t+1 given the system is in state t at time t. If the system has a finite number of states, 1, 2, … s, the stationary Markov chain can be defined by a transition probability matrix :

$$P=\begin{bmatrix} P_{11} & P_{12} \ldots & P_{1S} \\ P_{21} & P_{22} \ldots & P_{2S} \\ \vdots & \vdots & \vdots \\ P_{S1} & P_{S2} \cdots & P_{SS} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

An initial probability distribution is,

$$Q= [q_1 , q_2 \ldots q_s] \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where q_i is the probability that the system is in state *i* at time 0, and

$$\sum_{j=1}^{j=s} pij \ =1 \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The probability that a sequence of states $X_1, \ldots, X_T$ at time 1, …, *T* occurs in the context of the stationary Markov chain is computed as follows:

$$P(X1, \ldots XT) = qx1 \prod_{t=2}^{T} Pxt - 1xt \ldots\ldots (4)$$

The transition probability matrix and the initial probability distribution of a stationary Markov chain can be learned from the observations of the system state. Provided

with the observations of the system state X_0, X_1, X_2, …, X_{N-1} at time $t = 0$, …, N-1,

To learn the transition probability matrix and the initial probability distribution as follows:

$$P_{ij} = \frac{N_{ij}}{N_i} \quad \text{..... (5)}$$

$$q_i = \frac{N_i}{N} \quad \text{..... (6)}$$

In this study, used a Markov chain instead of a high-order stochastic process model to represent the temporal behavior of the host machine. It also made the stationary assumption that is, assuming that the user's action sequence was not related to the time of using the host machine.

For intrusion detection, to build a long term norm profile of temporal behavior, and to compare the temporal behavior in the recent past to the long-term norm profile for detecting a important difference. Using formulae (5) and (6), built a stationary Markov chain mode of temporal behavior as the long-term norm profile by learning the transition probability matrix and the initial probability distribution from a stream of events that was observed during the normal usage of the host machine.

To define the temporal behavior in the recent past by opening up an observation window of size N on the continuous steam of audit events to view the last N audit events from the current time t:

$E_{t-(N-1)=t-N+1}$, …, E_t, where E stands for event.

For the audit events E_{t-99}, …, E_t in the window at time t, we examine the type of each audit event and obtain the sequence of states X_{t-99}, …, X_t appearing in the window, where X_i is the state (the type of audit event) that the audit event E_i takes. Using formula (6), we compute the probability that the sequence of states X_t99, …, X_t occurs in the context of the normal usage, that is, the probability that the Markov model of the norm profile supports the sequence of states X_{t-99}, …, X_t.

$$p = (X_t - 99, X_1, \ldots X_t) = q_{xt-99} \prod_{i=98}^{t} PX_i - 1X_i$$

It is possible that a sequence of states from a window of the testing data presents an initial state and some state transitions which are not encountered in the training and thus have the probabilities of zero in the initial probability distribution or the transition probability matrix of the Markov model. While using formula (4) to infer the probability of support to the sequence of states, the probabilities of zero would dominate the final probability result from formula (4) and make it zero, regardless of the number of nonzero elements in the computation using formula (4). In this study we assigned the small probability to the initial state and state transitions which did not appear in the training data, while using formula (4) to infer the probability of support to a sequence of states. This replacement of zero with a small probability value took place during the inference and after the learning of the Markov model from the training data was completed.

IV. CONCLUSION

The paper proposed a simple machine learning anomaly detection procedure based on a persistent behavior observed through different enterprise network data types. The method is able to identify periods of time in which the network advances from normal behavior. Identifying such periods is a first step in attempting to identify sophisticated intrusion behavior. Future work involves development of the methodology. First, improvements to the statistical models to capture more structure. Second, creating streaming versions of the methodology that updates models automatically as new data arrives.

REFERENCES

[1] Friedberg, F. Skopik, G. Settanni, and R. Fiedler, "Combating advanced persistent threats: From network event correlation to incident detection." Computers and Security, vol. 48, pp. 35–57, 2015.
[2] N. Adams and N. Heard, Data analysis for network cyber-security. Imperial College Press, 2014.
[3] N. Adams, and N. Heard, Dynamic networks and cyber-security. World Scientific, 2016.
[4] A. D. Kent, "Comprehensive, multi-source cyber-security events," Los Alamos National Laboratory, 2015.
[5] A. D. Kent, "Cybersecurity data sources for dynamic network research," in Dynamic Networks in Cybersecurity. World Scientific, 2016.
[6] R. Fisher, Statistical methods for research workers. Oliver and Boyd, 1925.
[7] M. J. M. Turcotte, N. A. Heard, and J. Neil, "Detecting localised anomalous behaviour in a computer network." in Advances in Intelligent Data Analysis XIII, 2014, pp. 321–332.

