Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 4 (2012) 506 – 514

Procedia Technology

C3IT - 2012

Intrusion Detection System using Bayesian Network and Hidden Markov Model

Nagaraju Devarakonda[a], Srinivasulu Pamidi[b], Valli Kumari V[c], Govardhan A[d]

[a]*Department of cse, Acharya Nagarjuna University, Guntur- 522510, India*
[b]*Department of cse, V R Siddhartha Engineering College, Vijayawada-520007, India*
[c]*Department of CS & SE, Andhra University, Visakhapatnam-, India*
[d]*Department of CSE, J N T University, Hyderabad-, India*

Abstract

Across the globe, billions of dollars are spending every year to provide security to the network systems to prevent the intrusions. Some consider the disruption of the vital systems as a serious threat which disables the work of hospitals, banks, military and various internet services across the world. To avert this impending threat, there are many possible solutions: one of these solutions is intrusion detection systems (IDS). The paper proposes to discuss the IDS model in its elaboration using Bayesian Network and the Hidden Markov Model (HMM) approach with KDDCUP dataset. The IDS framework has been designed with various levels of processing such as model learning with training data and constructing the Bayesian Network and this structure has been used as HMM state transition diagram. The preprocessed KDDCUP dataset has been used to train and test the model. The IDS model has been trained and tested for normal and attack type connection records separately. The results evince that the performance of the model is of high order for classification of normal and intrusions attacks.

Keywords: IDS, State Transition Probabilities, Observations, Training, and Evaluation.

1. Introduction

An intrusion detection system (IDS) [1][5][15][16] is used to monitor network traffic, check for suspicious activities and notifies the network administrator or the system. In some instances, the IDS might also react to malicious or anomalous traffic and will take action such as barring the user or perhaps the IP address source from accessing the system.

IDS [8] are available in many different types and will approach the mission of uncovering shady traffic in various ways. Several types of IDSs exist such as Network Intrusion Detection System (NIDS), Host Based Intrusion System (HIDS), and Hybrid Based Intrusion Detection System. These systems use either statistical anomaly-based IDS or Signature based IDS for intrusion detection. It is impossible for

2212-0173 © 2012 Published by Elsevier Ltd.
doi:10.1016/j.protcy.2012.05.081

IDS to be perfect because network traffic is so high. The objective of the IDS is to minimize false positive rate and maximize true positive rate.

This paper aims at investigating the capabilities of HMM and Bayesian Network for building IDS. The Bayesian Network will be constructed with training data. Using the Bayesian network conditional probabilities can be estimated and the dependencies among the variables can be found. Based on network information the state transition probabilities and emission probability matrices can be initialized and these parameters will act as HMM parameters for model building. The structure of paper is as follows: Section 2 gives brief introduction to the concepts of Bayesian Network and HMM, section 3 covers kddcup99 dataset description, section 4 describes the experimental setup for building IDS using Bayesian Network and HMM Model, and last section 5 covers the concluding remarks.

HMM-based classifiers are capable of detecting intrusion attempts on network systems [9]. IDS using HMM based predictive model capable of discriminating between normal and abnormal behaviors of network traffic [5]. This paper investigates the problem of intrusion detection while reducing the number of false positives.

2. Bayesian Network and Hidden Markov Model (HMM)

Hidden Markov Model (HMM) approach can be used to various kinds of applications, such as speech recognition, Speech synthesis, Gene prediction, Crypt analysis and many more. The incorporation of HMM's for intrusion detection system is still in its infancy. Hidden Markov models are generative models in which the joint distribution of observations and hidden states are equivalent prior to the distribution of hidden states (the transition probabilities). This is based on conditional distribution of observations and given states (the emission probabilities) [4].

2.1 Bayesian Network

A Bayes net also called a belief network is an augmented directed acyclic graph, represented by G(V, E) where V is a set of vertices and E is a set of directed edges joining vertices. In Bayes net no loops of any length are allowed. Each vertex in V contains the name of a random variable and probability distribution table indicating how the probability of this variable's values depends on all possible combinations of parental values.

The following procedure has been used in building the Bayes Net:

1. Choose a set of relevant variables from training dataset. These variables are state variables in HMM
2. Choose an ordering for them.
3. Assume they're called X_1, X_2, ... , X_m (where X_1 is the first in the ordering, X_2 is the second, etc)
4. For i = 1 to m:
   a) Add the X_i node to the network
   b) Set Parents(X_i) to be a minimal subset of {$X_1 \ldots X_{i-1}$ } such that we have conditional independence of X_i and all other members of {$X_1 \ldots X_{i-1}$ } given Parents(X_i)
   c) Define the probability table of P(X_i = k | Assignments of Parents(X_i)).

2.2 HMM Architecture

The figure 1 shows the general architecture of an instantiated HMM. Each oval shape represents a random state variable that can adopt any number of values. The random variable $Z(t)$ is the hidden state at time t, where $Z(t) \in \{ Z_1, Z_2, Z_3 \ldots Z_M \}$. The random variable $X(t)$ is the observation at time t with