Nathanael Asaam

September 22017

Abstract

This paper is an investigation into building an anomaly threat detection system using behaviour models. Such behaviour models are the heart of machine learning, and evolutionary computing. Some other methods of building such anomaly threat detection systems include the use of statistical models such as time series models, univariate models and mean and standard deviation models. In this paper we describe how to build a usage profile of a system. The aim of building these usage profiles is to be able to detect unusual behavior on the system. As a result, the usage profile will be used as the basic building block for the development of the anomaly threat detection system. This paper uses regression to determine the usage profiles of a system by studying the relationship between relevant system variables that will be used to formulate the usage profile. The dependent and independent variables for the usage profile can be determined from an audit trail. Additionally, the paper applies hidden markov models to study the various states a computer system can fall into and the various stage transition in order to be able to predict unusual behaviour in the system. Unusual behaviour in this case may be a particular state or a transition from one state to another or the manner in which a particular state transition occurred. With this usage profile which is composed of the usage profile equation and a mean and standard deviation model that capture average usage and its standard deviation and the markov chain model that captures the various states of the system and the various state transition it becomes possible to detect anomaly on the system. Using linear and nonlinear programming, the usage profile equation can be maximized to determine states of the system and points at which the system is optimal. This can help improve the system’s usage. Also using differential coefficient of the usage profile equation and other statistical models such as the mean and standard deviation model, a threat profile of the system can be developed. When the threat profile equation is minimized using linear and nonlinear programming, it will help prevent threats on the system. The benefit of this research is its application to the development of anomaly threat detection systems and risk analysis systems that can be used for performing computer security risk assessments and analysis.

BUILDING AN ANOMALY THREAT DETECTION SYSTEM USING BEHAVIOR MODELS

**Introduction**

If a usage profile of a system can be built, it will become possible to detect unusual behaviour on the system. The method for building such usage profiles involved determining factors of the system that are critical to the system. These factors can be seen as critical system variables that affect the system’s usage. The other thing to consider is determining the way in which you can obtain an abstract representation of the usage profile. The abstract representation of the usage profile can be achieved by the application of behaviour models such as statistical models, machine learning models and cognitive based models.

The first goal of this research paper is to investigate techniques for building a usage profile of a computer system. The aim of building the usage profile is to be able to have a working model that describes the systems behaviour. The second goal of the research is to be able to detect unusual behaviour on the system. Unusual behaviour will be detected as deviation from the usage profile model built. The last goal is to be able to build an anomaly threat detection system and a risk analysis system that can be used for detecting threats and performing risk analysis on a system.

**Research Model**

The research model is given as Y=f(Xi) where Y is the usage of a system, and Xi are the various independent variables that can be related to a dependent variable Y to form the equation Y=f(Xi).

**Research Methodology**

The research methodology for developing the threat detection system is given below.

* Study machine learning techniques and algorithms, linear and nonlinear regression and intrusion detection system to know the extent to which the research model can be used to represent the usage of a system.
* Experiment the usage profiles that can be built by using programming as a tool.
* Apply rates of change, linear and nonlinear programming to detect threats that may occur in the system.
* Develop an anomaly threat detection system and a risk analysis system using the result of the experiment.

### BUIDING THE USAGE PROFILE

To build a usage profile, we use a mathematical model that captures the behaviour of the system and a markov chain model that captures various states and transitions in the system. The mathematical model is made up of a usage equation composed of a dependent and independent variables and a statistical model that captures average usage and its standard deviation. The usage equation of the system can be summarized as Y=f (Xi, Ci), where Y is our systems’ usage and Xi are the various independent variables of our system that constitutes the normal usage or behavior of the system.

The markov chain model can be built using various states or points of occurrences in a system and their associated probabilities. The mathematical model can be developed using regression. This paper will describe how to build a usage profile using an authentication model.

### BUILDING THE AUTHENTICATION USAGE MODEL

The critical system variables for an authentication usage model of a system on a network may include the download speed on the network, the upload speed on the network, the size of data sent to the server during authentication, the size of data sent to the client during authentication and the time it takes for a successful authentication. The size of data sent and received from the server are request data and response data respectively. To build the usage profile for the authentication data, we will capture data for all the critical variables at equal time intervals say every 10 minutes while the authentication system is being used. After having sample of sample size of about 10 we will try to determine the relationship between the dependent variable and the independent variable. As already stated the relationship can be determined using simple or multiple linear regression. In addition to the relationship, we will also determine other statistics that describe the behavior of the authentication system. These statistics are the averages for the various critical variable captured.

### BUILDING THE MARKOV CHAIN MODEL FOR THE AUTHENTICATION SYSTEM

Hidden markov models are machine learning models that are used to model states in a system, the sequence in which they occur and the associated probabilities for each state transition. When a system has a set of states in which it usually falls and it can be predicted or established that each new state is dependent on the previous states, then hidden markov models can be used to learn the state transitions that usually happens in the system. To build the markov chain model we will determine states on the authentication usage model and their associated probabilities. Some of these states include the average usage of the authentication system. This may be abstracted as the average time it takes for a successful authentication. Other states include the minimum and maximum recorded time for a successful authentication and the average time it takes for a failed authentication or the maximum and minimum recorded time for failed authentications. With this information and their associated probabilities of occurrence during a normal day we have more information about the behaviour of the authentication system.

**BUILDING THE ANOMALY THREAT DETECTION SYSTEM**

To build the anomaly threat detection system we will be finding occurrences or states in the system that deviate from the behaviour models we have built which are the regression based model and the markov hidden model. Any programmed system that infers from the two behaviour models built and alerts or blocks occurrences that are not in line with the behaviour seen in the two behaviour models is an anomaly threat detection system that prevents threats on the authentication system. As such, if for some reason the data from the independent variables don’t match up with their dependent variable when substituted into the regression equation obtained then an alert can be signal to show that there is a threat in the authentication system. Also if for some reason the probabilities for a particular state say the average time it takes for a failed authentication or the average time it takes for a successful authentication is increasing then we can signal an alert that there is a threat in the authentication system. Normally, after a model is built, and agreed upon we expect the behaviour of the system in question to conform to the built model when being used. This becomes the basis for signaling an alert.