Chapter 99
Survey of Unknown Malware Attack Finding

Murugan Sethuraman Sethuraman
Wolkite University, Ethiopia

ABSTRACT

Intrusion detection system(IDS) has played a vital role as a device to guard our networks from unknown malware attacks. However, since it still suffers from detecting an unknown attack, i.e., 0-day attack, the ultimate challenge in intrusion detection field is how we can precisely identify such an attack. This chapter will analyze the various unknown malware activities while networking, internet or remote connection. For identifying known malware various tools are available but that does not detect Unknown malware exactly. It will vary according to connectivity and using tools and finding strategies what they used. Anyhow like known Malware few of unknown malware listed according to their abnormal activities and changes in the system. In this chapter, we will see the various Unknown methods and avoiding preventions as birds eye view manner.

INTRODUCTION

This chapter surveys proposed solutions for the problem of Unknown Malware attack appearing in the computer security research literature. After describing the challenges of this problem and highlighting current approaches and techniques pursued by the research community for insider attack detection, suggest directions for future research.

Recent news articles have reported that Every year to year time to time an enormous increase of known and unknown malware variants . This has made it even more difficult for the anti-malware vendors to maintain protection against the vast amount of Unknown threats. Various obfuscation techniques, such as reverse engineering, honeypot, and intelligence intrusion detection prevention, contribute to this trend. The ongoing battle between malware creators and anti-virus vendors causes an increasing signature, which leads to vulnerable end-systems for home users as well as in corporate environments.

DOI: 10.4018/978-1-5225-5643-5.ch099

Data Mining Basics

Recent progress in scientific and engineering applications has accumulated huge volumes of data. The fast growing, tremendous amount of data, collected and stored in large databases has far exceeded our human ability to comprehend it without proper tools. Coverage and volume of digital geographic data sets and multidimensional data have grown rapidly in recent years. These data sets include digital data of all sorts created and disseminated by government and private agencies on land use, climate data and vast amounts of data acquired through remote sensing systems and other monitoring devices. It is estimated that multimedia data is growing at about 70% per year. Therefore, there is a critical need for data analysis systems that can automatically analyze the data, to summarize it and predict future trends. Data Mining is a necessary technology for collecting information from distributed databases and then performing data analysis.

The process of knowledge discovery in databases is explained and it consists of the following steps:

- Data cleaning to remove noise and inconsistencies.
- Data integration to get data from multiple sources.
- Data selection step where data relevant for the task is retrieved.
- Data transformation step where data is transformed into an appropriate form for data analysis.
- Data Analysis where complex queries are executed for in depth analysis.

The following are different kinds of techniques and algorithms that data mining can provide:

Association Analysis involves discovery of association rules showing attribute value conditions that occur frequently together in a given set of data. This is used frequently for transaction data analysis.

A popular algorithm for discovering association rules is the Apriori method. This algorithm uses an iterative approach known as level-wise search where k-itemsets are used to explore (k+1) itemsets. Association rules are widely used for prediction.

Classification and Prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. The basic techniques for data classification are decision tree induction, Bayesian classification, and neural networks. These techniques find a set of models that describe the different classes of objects. These models can be used to predict the class of an object for which the class is unknown. The derived model can be represented as rules (IF-THEN), decision trees or other formulae.

Clustering involves grouping objects so that objects within a cluster have high similarity but are very dissimilar to objects in other clusters. Clustering is based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. Due to a large amount of data collected, cluster analysis has recently become a highly active topic in Data Mining research. As a branch of statistics, cluster analysis has been extensively studied for many years, focusing primarily on distance based cluster analysis. These techniques have been built into statistical analysis packages such as S-PLUS and SAS. In machine learning, clustering is an example of unsupervised learning. For this reason, clustering is an example of learning by observation.

A database may contain data objects that do not comply with the general model or behavior of data. These data objects are called outliers. Most Data Mining methods discard outliers as noise or exceptions.