ABSTRACT

As the primary vehicle for most organized cybercrimes, botnet has become one of the most serious threats to cyber security. Today, there is a high incentive for cybercriminals to engage in malicious, profit-oriented illegal activity on the Internet. A popular tool of choice for digital criminals today is botnets. This has led to a surge in the number of new botnet threats and has created several major challenges for the cyber security.

Botnet has affected desktop computers and mobile devices alike. Microsoft Windows is the most widely used consumer operating system worldwide, mostly used in desktop computers. Hence a vulnerability identified in one system could compromise a large number of systems. On the other hand, the ubiquity of mobile devices has led to a rowing adoption of mobile applications in a variety of application areas and 60% of Android malware includes elements of large or small botnets confirming both the popularity of this mobile OS and the vulnerability of its architecture. Hence in this thesis, a study and analysis of botnets affecting the above mentioned operating systems is carried out to find a better solution to detect botnets and mitigate the attacks.

The traditional methods of protecting a system through signature based detection are beset with drawbacks justifying the need to look for better and efficient solutions. This thesis aims to address the problems related to traditional botnet detection methods and arrive at solutions that are more effective. The work in this thesis investigates the botnet at different stages to design efficient detection mechanisms and to overcome the drawbacks of traditional botnet detection methods both in Windows and Android platform.

The first work focuses on the analysis of botnets to understand the behavior and lifecycle mechanism of botnets which would be helpful for future study of thwarting botnet communications. Zeus, BlackEnergy, Spyeye, Athena, TDL4, Rbot, Virut.n etc., are some of the various botnets that are analyzed in a controlled environment to identify its possible attacks, their topology, C&C structure and communication mechanisms. Also a brief survey of some of the existing botnet detection techniques along with their advantages and limitations is considered. This analysis helps us to identify the significant features of various botnet which could be used in the detection techniques.

The next work focuses on the detection of HTTP botnets in the host level. Most of the communications of web based HTTP botnets are based on the TCP connection. The data collected through the Management Information Base of Simple-Network Management Protocol (SNMP-MIB) becomes an important entity since it provides information about the change in the host directly. Hence TCP connection related SNMP-MIB variables are used as features to model the system behavior at host level using Hidden semi-Markov Model (HsMM). Several experiments are conducted by using HTTP botnets to validate our model. The proposed model is efficient with high detection accuracy and low false positive rate and also the model is light weight, and real time.

In the next work, a HTTP botnet detection at network level is proposed. The relative and direct features of TCP connections are extracted from the network traffic. The extracted TCP features are passed to the Multi-Layer Feed Forward Neural Network training model which uses Bold Driver Back-propagation learning algorithm to detect the botnets in the network level. The algorithm has the advantage of dynamically changing the learning rate parameter during weight updating process. The performance of the proposed method is compared with that of C4.5 Decision Tree, Random Forest and Radial Basis

function network. Results show an improvement in detection accuracy with neural network when compared to other classification techniques.

In the next work, a method is proposed to detect botnets irrespective of their command and control structures, based on network traffic flow behavior analysis and machine learning techniques. Botnet characteristics have been analyzed in a controlled environment, based on their behavior, four traffic flow features have been extracted in different time windows. After identifying the significant features, bots can be detected in advance before it launches some attack. To accomplish this task, individual flows are split into multiple parts using time windows W_T in seconds. The characteristics of a given flow are observed in a given time window. After the feature extraction, flow vectors are formed to classify the traffic flows into botnet or normal flows by applying machine learning techniques namely, Boosted decision tree ensemble classifier, Naive bayesian (NB) statistical classifier and support vector machine (SVM) discriminative classifier. The experimental results show that the proposed approach is capable of detecting the known and new botnets effectively irrespective of their structures and also it achieves high detection accuracy and low false positive rate.

Smart phone device usage has expanded at a very high rate and Android has surpassed other mobile platforms as the most popular whilst also witnessing a dramatic increase in botnet targeting the platform. Android botnets are analyzed and a detection mechanism is designed using machine learning algorithms. Unique patterns (i.e combinations of requested permissions and used features) based on malicious activities of botnets are generated by using Apriori association rule mining algorithm and information gain method is used to select the most significant patterns in order to provide a better detection. The selected unique patterns are passed to the machine learning framework to classify the applications as benign or botnet. We have used diverse sources of Android

botnet datasets such as, Android Malware Gnome project, Drebin, Droid Analytics, ISCX Android Botnet dataset and dataset from Beijing Jiao-tong University of China. Experiments on real-world benchmark datasets show that the selected patterns produce high detection accuracy compared with prior state-of-art works.

The various works mentioned above tried to find a rational solution to the important problem of botnet detection. Analysis of many real bots revealed significant features and machine learning techniques have been used to identify botnet.