International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 10 Issue 9 - September 2022 - Pages 49-56

DETECTING BOTNET TRAFFIC BY USING MACHINE LEARNING

K.BAVTHA[1*], B.JENEFA [1] ,P.SEETHA[2*],

1.Assistant Professor Department computer science and Engineering
Annai Vailankanni College of Engineering, Kanakumari-629401.
bavitha-ai@avce.edu.in
2.Assistant Professor Department computer science and Engineering
Annai Vailankanni College of Engineering, Kanakumari-629401.
jenifa-cs@avce.edu.in
3.Assistant Professor Department computer science and Engineering
Annai Vailankanni College of Engineering, Kanakumari-629401.
seetha-ai@avce.edu.in

ABSTRACT

Many cyber security mishaps have been recorded globally in recent years as a result of distributed denial of service assaults. Many of these assaults were carried out via a botnet, which is often made up of hacked computers, cell phones, or IoT devices. This research suggests a machine learning technique for detecting botnet traffic. First, we used Malware Capture Facility Project datasets. The files comprise network traffic data acquired from the target machine's victim. Botnet traffic and regular traffic are both included in the network traffic data. Second, we preprocessed the traffic data and retrieved information such as source and destination addresses, ports, and packet sizes. Third, we used a machine learning system to distinguish between botnet and regular traffic. The botnet detection module is trained using a single huge dataset that includes botnet and regular traffic records. After the trained model has achieved high accuracy, new dataset is loaded into the module for detection. The suggested method can detect botnet traffic with high accuracy.

Keywords: Botnet.Machine Learning,Network,Network Security

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

INTRODUCTION

One of the most vexing cyber-security threats today is the use of very large, coordinated groups of hosts for brute force attacks. These large groups of hosts are assembled by turning vulnerable hosts into so-called zombies, or bots, after which they can be controlled from afar. A collection of bots, when controlled by a single command and control(C2) infrastructure, form what is called a botnet. Botnets obfuscate the attacking host by providing a level of indirection and separating the assembly of the botnet and its use for attack by an arbitrary amount of time. Botnets, often involving thousands of hosts, are increasingly being used to launch highly-effective cyber-attacks.

We use two-stage approach. In **Stage I**, we classify communication flows (i.e., TCP connections) observed as either chat or non-chat flows. Here, the underlying assumption is that the C2 flows of chat-based botnets indeed resemble real chat flows. We argue that this is the case for two reasons:

(1) botnet C2 channels have, to date, been observed to exchange text-based commands that are of comparable length to that of real chat messages, and

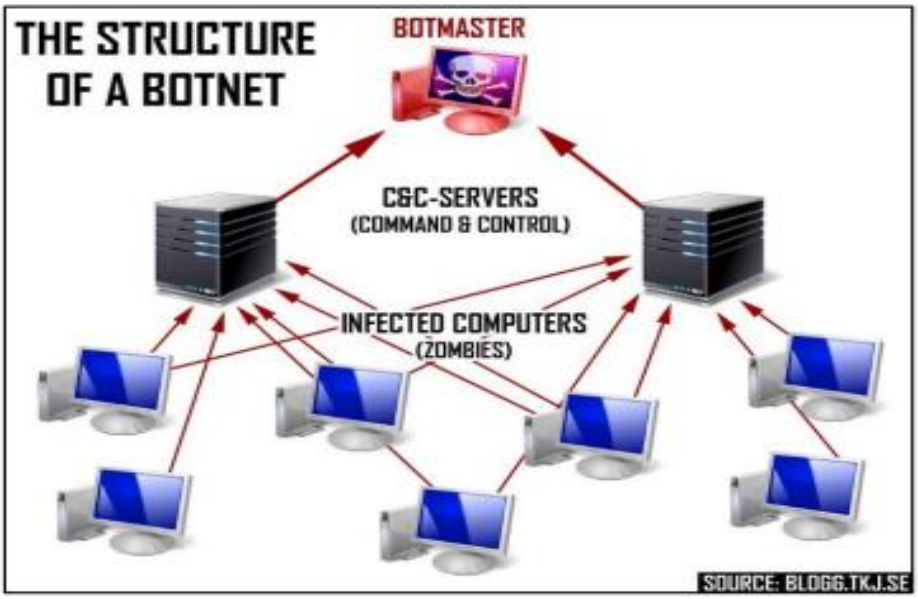
(2) in order for botnet flows not to alarm chat server administrators, they must roughly resemble real chat connections. Thus, we expect that a classifier that is trained to differentiate between chat and non-chat flows will classify botnet chat flows as chat flows.

Using classification in **Stage II** is trickier. We are currently investigating whether machine learning-based classifiers can be used to distinguish between botnet and real IRC flows. Here, the underlying assumption is that botnet IRC flows are, in some subtle respects, different than real IRC flows. The challenge here is to accurately label IRC flows as either botnet or real IRC flows. Because the traces available to us are anonymized and their payloads have been discarded, labeling IRC traffic as either botnet or real IRC traffic is challenging. Nevertheless, we are exploring the possibility of labeling flows as either suspicious or non-suspicious using telltales of hosts being compromised.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 10 Issue 9 - September 2022 - Pages 49-56



3.1. Existing System:

Although there are many botnet detection techniques, the following four techniques explained are widely used.

3.1.1. Signature-based vs Anomaly-based Botnet Detection:

In signature-based detection technique, a malware is identified by its characteristics or signatures. The signatures can be examinable through internal visible behaviors, which is host-based botnet detection, and if the behavior is externally visible, it is a network-based signature detection technique. To detect botnets using signature-based techniques, a signature list needs to be securely maintained by each machine. This list has to be updated periodically as new threats are discovered to make this approach effective against clever malware which attempts to tamper with the secure list. Signature-based botnet detection approaches yield a very low number of false positives, but are not effective against unknown threats.

Abnormal behavior can be detected through statistical models to check abnormal behavior in a system which include high network latency, unusual port traffic, high CPU memory or high disk usage. One of the advantages of this type of detection is that it is strong against new/emerging threats. There are two weaknesses with this kind of approach:

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

• It can yield high rates of false positives because it tries to capture new/emerging threats using system anomalies.

• Overhead of knowing what constitutes the normal behavior and what does not to capture data for new/emerging threats.

3.1.2 Host-based Signature vs Network-based Signature Botnet Detection:

Host-based botnet detection technique examines a host's behavior, resources and internal state to detect malware. A few of the differentiating characteristics could be:

• A collection of system calls performed by a binary

• Instruction patterns in a binary

• Malwares that attempt to connect to an IP address

• What times the malware is most active during a given time of a day

• Low/high processor utilization with particular conditions

• File/accessed or modified file sets by the binary

Signatures can be created with the following characteristics: modification of privileges, system resources, registries, the interaction between processes, and process lifetimes.

3.1.3 DNS-based Botnet Detection:

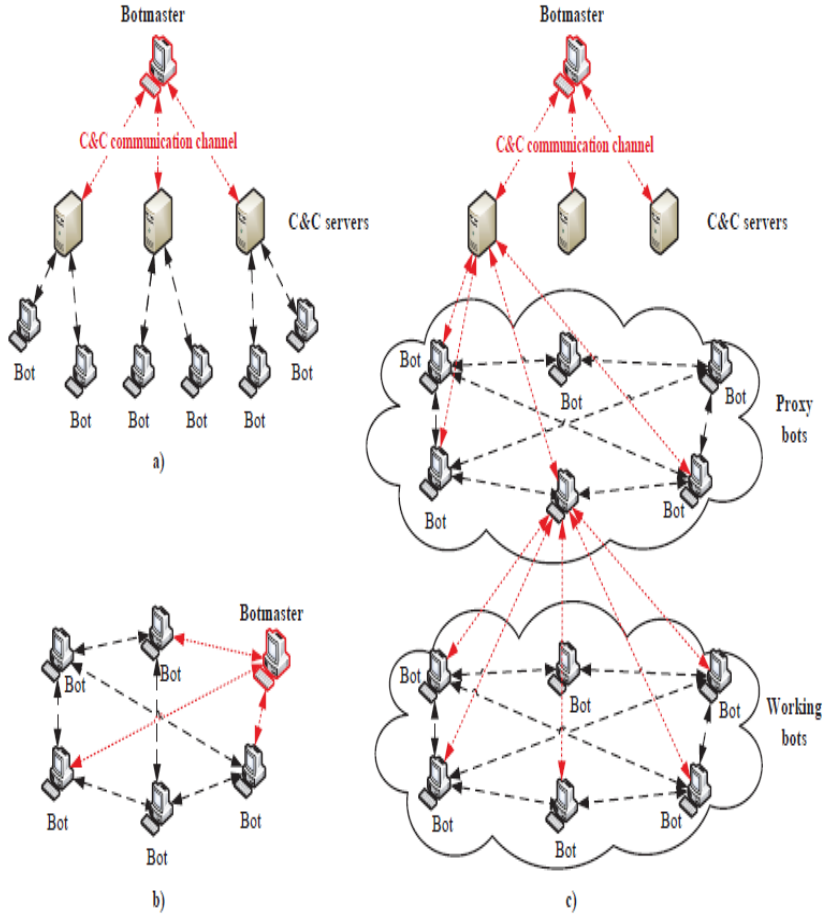
This technique uses DNS (Domain Name Space) request and response queries. DNS query and response traffic can be considered as main elements in network behavior. A number of botnet detection and identification methods are available based on DNS activities and behavior in the network infrastructure. Unlike other approaches or techniques, this technique does not require specific knowledge about botnet protocol or its behaviour.

In signature-based detection technique, botnets are identified by their host or network characteristic signatures. These signatures are maintained in the repository and need to be updated periodically. This technique is not efficient to identify unknown threats. The machine learning technique is used to address this problem. For example, to detect a botnet, the classification model collects the normal traffic behavior data. If the data does not match a normal traffic pattern, the algorithm identifies it as a botnet. Anomaly-based detection technique checks the abnormal behavior of the system. The advantage of this technique is to identify new emerging threats, but it has high false positive rates. The SVM classifier algorithm divides data into multiple classes based on their different

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

characteristics. The decision function of SVM classifier can be optimized to reduce false positive rates.

3.4. Architecture Diagram:



3.2. Proposed System:

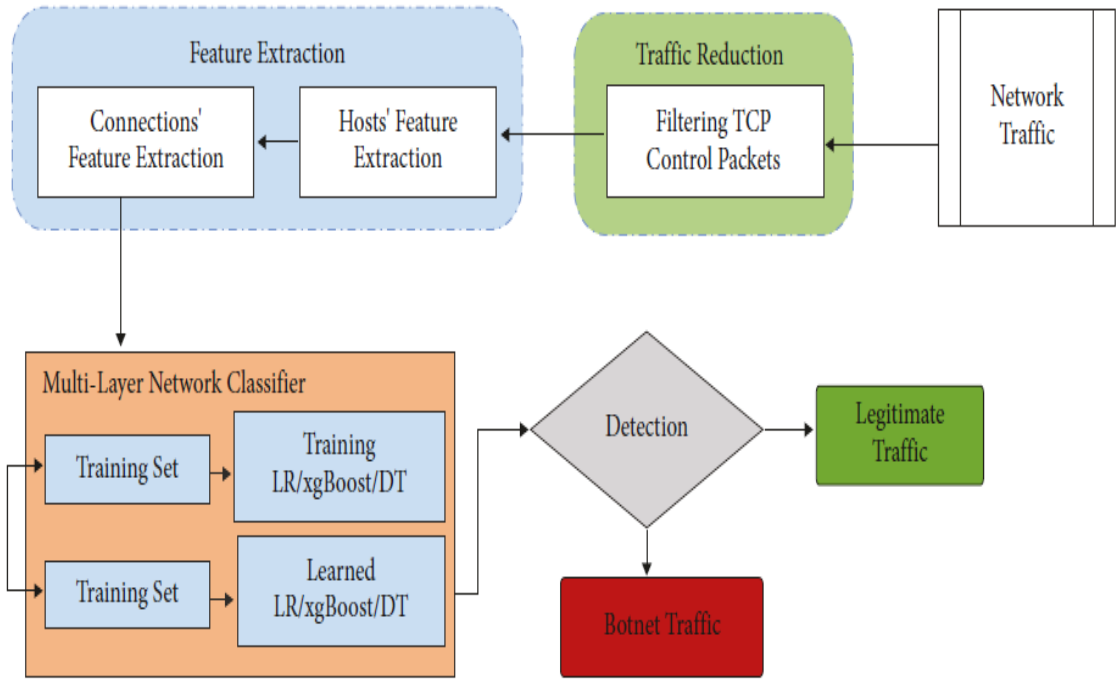
3.2.1. Dataset Description:-

The CTU-13 datasets from Malware Capture Facility Project were used in this project. These datasets are labeled as normal, background and botnet traffic inflows. It was extracted using the NERIS software at CTU University, Czech Republic in 2011. The botnet was run for 6 hours and 15 minutes on HTTP and C&C channel to send malicious

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

contents and perform click-fraud. Argus Software [9], a system and network monitoring software, was also used to monitor and visualize inflows in the network.

The main objective of data capturing was to store and accumulate a big amount of real botnet traffic combined with the ordinary traffic. The data contains 13 captures which are 13 scenarios of different botnet samples. Each scenario was executed using a different malware with several protocols. The data in each scenario/instance was captured in PCAP format file which consists of all the packets of traffic. Later on, the file was processed to obtain attributes, such as source address, destination address, port, and packet size, extracted from packet information.



CONCLUSION:

With the spread of botnet attacks and diversified behaviors of the botnets, implementation of a robust detection technique is necessary. The implemented technique in this project can mitigate complications of existing techniques by yielding

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

precise outcomes through refined methods to identify botnet traffic. From the result analysis, SVM classifier showed better results in comparison with random forest and linear regression. The network traffic was logged and analyzed in the proposed solution to identify botnet traffic from normal traffic.

FUTURE ENHANCEMENT

In future, characteristics like API calls, registry changes by system procedure calls, and system stats can also be interpreted and evaluated to identify botnet traffic. This method can be extended to build robust algorithms and computations for precise performance. The current method can plot and generate graph for two attributes. Extending this to plot more than two attributes in a multi-dimensional plane can yield more accurate results for the detection.

REFERENCES:

1. Botnet. [Online]. Available: https://en.wikipedia.org/wiki/Botnet. [Accessed: January 2017].
2. Botnet Detection and Removal: Methods & Best Practices. [Online]. Available: https://www.alienvault.com/blogs/security-essentials/botnet-detection-and-removal-methods-best-practices. [Accessed: January 2017].
3. E. Alparslan, A. Karahoca, and D. Karahoca, "BotNet Detection: Enhancing Analysis by Using Data Mining Techniques," *Advances in Data Mining Knowledge Discovery and Applications.* Turkey: InTech, 2012.
4. S. Anwar, J. M. Zain, M. F. Zolkipli, and Z. Inayat, "A Review Paper on Botnet and Botnet Detection Techniques in Cloud Computing," in *Proceedings of the ISCI*, 2014, pp. 28-29.
5. H. S. Nair and V. Ewards, "A Study on Botnet Detection Techniques," *International Journal of Scientific and Research Publications*, vol. 2, issue 4, 2012.
6. Kang Brent Byung Hoon, "DNS-based Botnet Detection," in *Encyclopedia of Cryptography and Security*, 2011, pp. 362-363.
7. Malware Capture Facility Project. [Online]. Available: https://mcfp.weebly.com/mcfp-dataset.html. [Accessed: March 2017].
8. Python Programming. [Online]. Available: http://www.codingdojo.com/what-is-python-programming. [Accessed: March 2017].
9. Argus Software. [Online]. Available: https://argus-sec.com/argus-connectivity-protection. [Accessed: May 2017].
10. Support Vector Machine. [Online]. Available: http://www.ipcsit.com. [Accessed: June 2017].

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 9 - September 2022 - Pages 49-56

11. V. Jakkula, "Tutorial on Support Vector Machine (SVM)," School of EECS, Washington State Univ., Pullman, WA, USA, 2006.