Hybrid Botnet Detection Mechanism

Katha Chanda
Computer Science and Engineering Department,
Amity School of Engineering and Technology, India

ABSTRACT
Botnets have emerged as one of the biggest threats to internet security in the recent years. They have confounded security researchers because of their mobile and secretive behavior. A Botnet is a network of zombie machines remotely controlled by a command server or a Botmaster. These compromised host machines may be used for sending spam, launching DOS attacks, spying or stealing information. As botnets have evolved, so has the detection techniques changed. A number of different techniques have been suggested yet no technique is completely foolproof. While some are based on detecting anomalies, others focus on DNS queries [Choi et al., 2007] or DNSBL [Ramachandran et al., 2006] queries etc. This paper analyzes layouts of different detection techniques. The paper tries to find features that, when combined together, complement each other's strengths and eliminate the weaknesses and suggests a framework consisting of a combination of those features which, theoretically, should overcome most of the common problems faced by detection techniques.

General Terms
Botnets, Detection techniques, Security threat, Networks, hybrid botnets

Keywords
Signature based detection, Anomaly based detection, Hybrid Method, Protocol Independence.

1. INTRODUCTION
As mentioned above, a botnet consists of an army of compromised machines, made to do the bidding of the Botmaster. Initially, bots were harmless. They were used to monitor chat rooms, IRC channels etc. Later on, they began to be used for malicious purposes. The size of a botnet may vary from tens and hundreds to a few thousand. Usually, the host machines are unaware of the fact that they are compromised. In fact, the very PC or laptop that we are using for our everyday work at our offices or homes could be a part of a botnet and we would not even know! Depending upon the changing features of botnets, researchers have come up with a variety of techniques. In the following sections, a few pre-existing works have been explored, emphasizing their advantages and weaknesses. Then they have been combined to design a structure, which should improve the accuracy of botnet detection, thus eliminating the chances of overlooking a bot or flagging benign data as malicious.

2. RELATED WORK
There have been a number of works on the detection of botnets. Different researchers have tried tackling the issue via different methods. Each technique has been explored and a number of theories and ideas have sprung up from each. There are signature based, anomaly based, host based, network based and data mining based techniques which have all been explored thoroughly. The earliest bot detection theories used signature based techniques. They were very useful but it quickly lost popularity when it couldn't detect the unknown bots. A Ramachandran **[1]** used DNSBL counter intelligence to monitor DNSBL reconnaissance activity by bots. Li et al **[2]** studied and identified a number of common botnet features and used them to check whether a host is infected or not. It is very efficient for detecting known botnets but accuracy rate for unknown bots were a little lower. Zeidanloo et al **[3]** studied network flow characteristics and used similar behavior of hosts to classify and identify those behaving suspiciously. The method is very effective at detecting unknown bots but fails for encrypted data packets unless a decryption mechanism is provided. Chois et al **[4]** presented a paper which implemented the monitoring of group activities in DNS traffic for detection. It presented a solution that was structure independent, could overcome encryption and could also detect server migration. However it could be bypassed by deliberate DNS spoofing. A more comprehensive list of existing techniques is provided in Table 1 and a comparative analysis of the discussed papers in Table 2.

3. APPROACHES TO BOTNET DETECTION

3.1 Signature Based Detection
This method maintains a database of known attacks and compares the features of the network traffic against that database. It is very efficient while detecting known bot attacks. The chances of a false positive are almost zero and detection occurs immediately. It requires less use of system resources. However it cannot detect unknown bots. This uses the signatures of current Botnets for its detection. For instance, Snort, which is an Intrusion Detection System, is capable of monitoring network traffic to find signature of existing bots.

3.2 Anomaly Based Detection
This method focuses on studying the normal behavior and statistics of the system and monitoring any behavior or pattern that is unusual or abnormal. It studies characteristics like high volume of data, high network latency, traffic on unusual ports and many more. So, it focuses on normal behavior to detect unknown attacks. This method is very efficient in detecting unknown bots. It has two phases- Training and Detection phase. In the training phase, the normal behavior system (in the absence of an attack) is observed and a profile is created, using machine learning techniques. In the detection phase, the current behavior of the system is compared to the created

Table 1. A few existing techniques

TITLE OF PAPER	AUTHOR	TECHNIQUE USED
A Detection Method for Botnet based on Behavior Features	Weiming Li, Songlin Xie, Jie Luo,Xiaodong Zhu	Detection method is based on botnet features such as accessing backup DNS server, scanning, null TCP connections.
BotNet C&C Control Behavior Analysis Using HoneyPot and Reverse Hacking Techniques -	*Han-Wei Hsiao, Yu-Han Lin*	Used reverse hacking technique to monitor internal information of botnet and IRC server
Botnet Detection by Abnormal IRC Traffic Analysis	Gu-Hsin Lai, Chia-Mei Chen, and Ray-Yu Tzeng, Chi-Sung Laih, Christos Faloutsos	Detects abnormal IRC traffic to identify suspicious behavior of hosts.
Botnet Detection by Monitoring Group Activities in DNS Traffic	Hyunsang Choi, Hanwoo Lee, Heejo Lee, Hyogon Kim	Group activities in DNS traffic is used to detect Botnets.
Botnet Detection by Monitoring Similar Communication Patterns	Hossein Rouhani Zeidanloo, Azizah Bt Abdul Manaf	Studies network flow characteristics and used similar behavior to classify and identify suspicious hosts.
Revealing Botnet Membership Using DNSBL Counter-Intelligence	Anirudh Ramachandran, Nick Feamster and David Dagon	Performed counter intelligence based on the fact that Botmasters may perform DNSBL checkups in case their bots are blacklisted.
Identifying Botnets Using Anomaly Detection Techniques Applied to DNS Traffic	Ricardo Villamarín-Salomón and José Carlos Brustoloni	Uses an anomaly based method. Proposed a Bayesian approach based on similarity of DNS traffic to detect bots.

profile. However, it may use a lot of system resources as it has to constantly update the user and system profiles and it also generates a high false positive alarm **[5]**.

3.3 Data Mining Based Detection

This method uses machine learning, data clustering and classification to detect botnets. Bothunter**[6]**, BotSniffer and Botminer**[7]** rely on data mining techniques to detect botnets. BotSniffer is based on the fact that bots in the same Botnet display similarities in their responses and activities. Botminer clusters similar communication traffic and malicious traffic and performs cross cluster communication to identify which machines have both similar communication patterns as well as malicious traffic patterns. It has a low false positive rate.

4. PROPOSED DETECTION SYSTEM

No method is foolproof. There are certain advantages and disadvantages of each technique and moreover it is near impossible to keep up with the dynamic nature of botnets. Each new botnet is launched with new features and characteristics which makes the task of detection much harder. This work attempts to study the existing technology and create a detection mechanism that will, using that technology, eliminate a number of weaknesses that individual techniques face. It tries to combine detection mechanisms that complement each other with their best features and minimize the drawbacks. The focus is on the most common obstacles that botnet detection systems face and clusters different features from different mechanisms to try and eliminate them. The following are a few ways those obstacles could be removed.

Objectives of proposed detection system:

1. To overcome encryption
2. To detect botnets independent of the protocol
3. To be able to detect unknown bots
4. To improve the accuracy of detection(reduced high false positives and negatives)

Table 2. Features of a few discussed papers

	Protocol Independence	Detect known	Detect Unknown	Reduce false +ve	Reduce false –ve	Overcomes encryption
Signature (general)	No	Yes	No	Yes	No	No
[3] (Anomaly)	Yes (different mechanisms provided)	Yes	Yes	No	Yes	No
[2] (partly signature)	Yes	Yes	No (low efficiency)	Yes	No	--unmentioned but possibly yes
[4] Anomaly	Yes	Yes	Yes	Unmentioned		Yes

[2] proposes an IDS like mechanism to develop 6 specific components to detect 6 most important and most common botnet characteristics. It is protocol independent as the factors used to determine the abnormalities do not depend upon the protocol used. The mechanism successfully detected IRC, HTTP and P2P bots. According to the published results, false positives were zero in all cases. For known bots, false negatives ranged from 5-20% and for unknown bots from 25-45%. The detection rate for known bots is very high (74-94%) however the detection rates for unknown bots is comparatively lower (56-70%)

[3] describes a detection mechanism, which depends on the protocol utilized but provide mechanisms for detecting bots using all three of the commonly used protocols, namely IRC, HTTP and P2P. The application classifier divides the traffic into IRC and HTTP, by examining the packet contents. IRC messages contain keywords like NICK, JOIN, USER and HTTP messages keywords like GET and REQUEST. The remaining traffic is considered to follow P2P protocol and is filtered separately. There are chances of false positives as it is an anomaly based detection method. Even a non malicious anomaly can be misconstrued as a bot characteristic. False negatives are reduced as even the smallest anomalies are detected and will classify the machine as a suspicious bot. The method detects known bots as well as unknown bots. The paper states that the proposed mechanism does not need prior knowledge of existing bots, like their signatures.

[4] monitors DNS group activities to detect bots. It makes use of the information contained in the IP headers to collect the DNS names queried hence it can overcome encryption. It also is protocol independent. It can successfully detect both known and unknown bots and incorporates an algorithm for detecting bots even in case of C&C server migration. Table 1 summarizes the positives and the shortcomings of the three papers mentioned above.

5. METHODOLOGY

The basic structure follows that which is described in **[3]** with extra features incorporated from other works. Network traffic is passed through a filter. Those which are bound for Whitelisted servers like Yahoo or Google can be removed. It will reduce the traffic load. Then the traffic is passed through a signature based detection mechanism. This detects all known bots and reduces the false negative rate. Now the network possibly consists of unknown bots and some legitimate traffic. The traffic can be classified as encrypted and non-encrypted.

5.1 Non Encrypted Traffic

Traffic can be separated into IRC based, HTTP based and P2P based using packet data analysis. Different protocol detection techniques have been specified in different papers. The authors of **[3]** use a few known keywords to identify IRC and HTTP traffic. They suggest the use of BLINC to detect P2P traffic **[8]**. This data can be analyzed and suspicious data can be streamlined using the technique specified in **[3]**. A malicious activity detector for scanning and spamming can be used on P2P traffic. The traffic is also monitored by capturing network flow and recording particular data (Source IP, Destination IP, Source Port, and Destination Port) as seen from the traffic. This is responsible for identifying hosts that may be a part of a bot when the hosts begin an attack. The results of the malicious activity detector and the traffic monitoring system are combined and the hosts common to both are likely bots.

The IRC traffic has a one-to-many architecture between the IRC server and the bots; hence the bots all have a similar pattern of communication. Therefore, traffic is inspected and similar characteristics of the network flow are captured. The aim is to identify hosts that are a part of a botnet before the attack begins, i.e. when the server is updating or commanding the bots.

This is an anomaly based technique and has been seen to reduce false negative rates efficiently. By passing the data

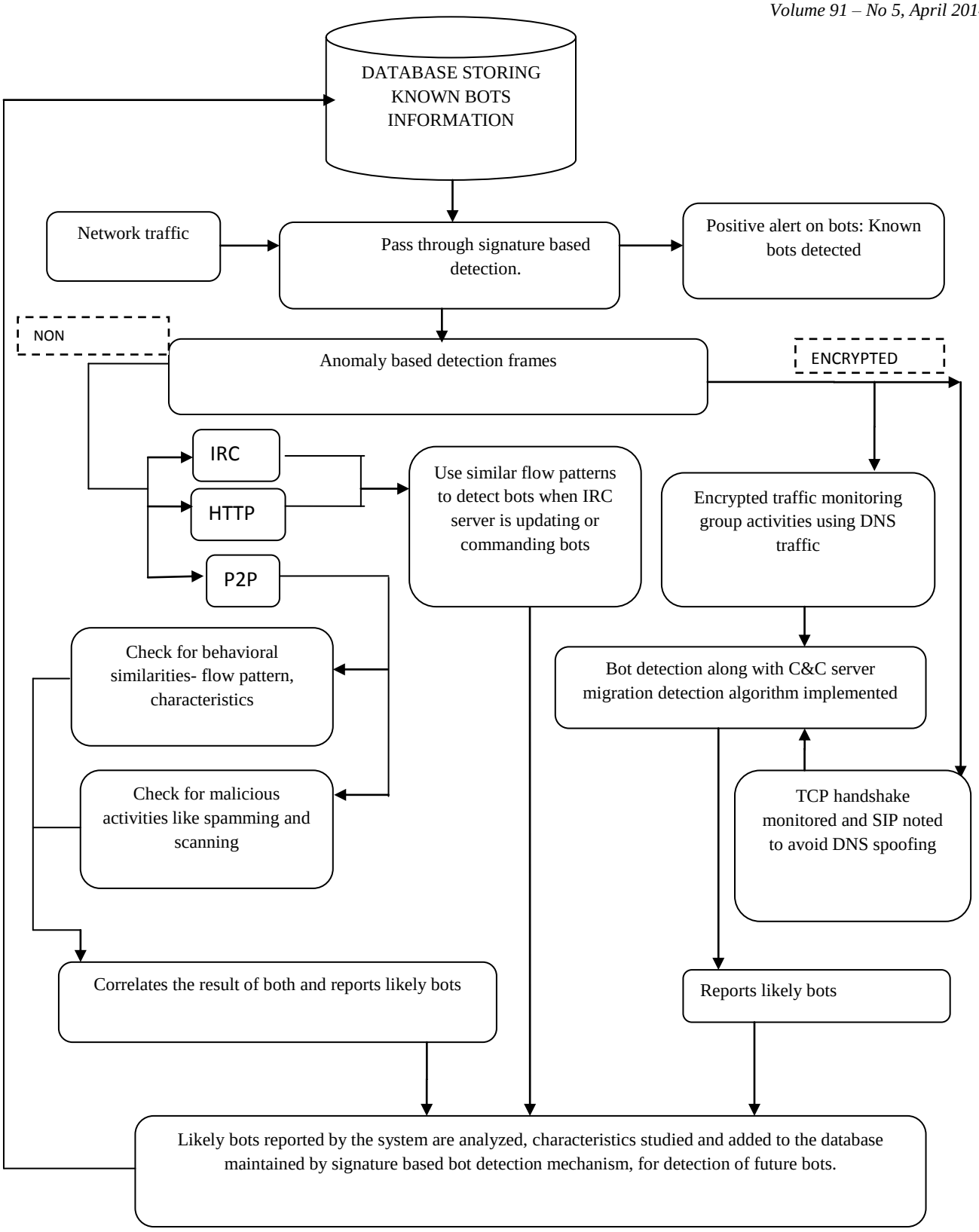

Figure 1. Framework in detail

first through a signature based detection system and then through an anomaly based detection system, the chances of both false positive and false negatives can be reduced as both methods complement each other's advantages. This should improve the general accuracy of botnet detection.

5.2 Encrypted Traffic

H Chois et al **[4]** provide an efficient method for detecting botnets by monitoring DNS group activities, their detection framework can overcome encrypted packets because it analyses the IP headers only. This makes the technique independent of structure as well as encryption. The algorithm provides a mechanism for detecting botnets even in the case of C&C migration, so by gathering DNS information, botnets are detected. Incorporating this algorithm solves the issue of encrypted data and also may detect bots that were missed by the previously described detection methods. However, the papers state that the algorithm could fail if the DNS query was spoofed using a fake IP address. They have suggested a way to overcome it. By monitoring the TCP handshaking and comparing the IP addresses used in the handshake to the ones in the DNS queries, fake queries could be detected. This added feature can cover all bases.

Once an anomaly based detection mechanism identifies a bot, its features could be studied and added to the database of the signature based detection mechanism. This would make the signature based mechanism much more efficient. The bot characteristics need not be studied separately and added to the mechanism. Figure 1 describes the proposed framework in detail.

6. CONCLUSION

Over the years, many techniques have been proposed, implemented and improved upon, to detect and remove botnets. Yet, it is clear that no method is completely foolproof. No technique is perfect, they have exists vulnerabilities that can be exploited and eventually, malicious users will exploit them and render them ineffective. Instead of opting for one particular method, it seems that an amalgamation of existing techniques can be carefully chosen so that most of the weaknesses can be eliminated. As is demonstrated above, the methods employed by a few existing works have been combined to form a framework for a detection system that should eliminate most of the major known weaknesses of anomaly based detection systems and signature based detection systems. There lies the potential to construct many such frameworks from existing technology that can improve the accuracy and efficiency of botnet detection systems.

7. REFERENCES

[1] N. Feamster, A. Ramachandran and D. Dagon, "Revealing botnet membership using dnsbl counter-intelligence*," in The 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI '06),* 2006.

[2] W. Li, S. Xie, J.Luo, Xiaodong Zhu,A Detection Method for Botnet based on Behavior Features, *Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM-2013),* 2013

[3] Hossein Rouhani Zeidanloo,Azizah Bt Abdul Manaf, "Botnet Detection by Monitoring Similar Communication Patterns",*(IJCSIS) International Journal of Computer Science and Information Security*,Vol.7, No.3, March 2010,pp.36- 45.

[4] H. Choi, H. Lee, and H. Kim. Botnet detection by monitoring group activities in dns traffic. In *proceedings of the 7th IEEE International Conference on Computer and Information Technology(CIT'07),* Washington, DC, October 2007**.**

[5] Robiah Y, Siti Rahayu S., Mohd Zaki M., Shahrin S., Faizal M. A., Marliza R.;A New Generic Taxonomy on Hybrid Malware Detection Technique. *(IJCSIS) International Journal of Computer Science and Information Security*, Vol. 5, No. 1, 2009

[6] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee. BotHunter: Detecting malware infection through ids-driven dialog correlation. In *Proceedings of the 16th USENIX Security Symposium (Security'07)*, 2007.

[7] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol- and structure independent Botnet detection," *in Proc. 17th USENIX Security Symposium*, 2008.

[8] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC:multilevel traffic classification in the dark," In *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 229-240, Philadelphia, Pennsylvania, 2005