Vol 4 Issue 7 April 2015

ISSN No : 2249-894X

Monthly Multidisciplinary Research Journal

Review Of Research Journal

Welcome to Review Of Research

RNI MAHMUL/2011/38595 **ISSN No.2249-894X**

Review Of Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Advisory Board

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.ror.isrj.org

Review Of Research ISSN : 2249-894X Impact Factor : 3.1402(UIF)
Volume-4 | Issue-7 | April-2015 Available online at www.ror.isrj.org

DETECTION OF ENCRYPTED BOTNETS

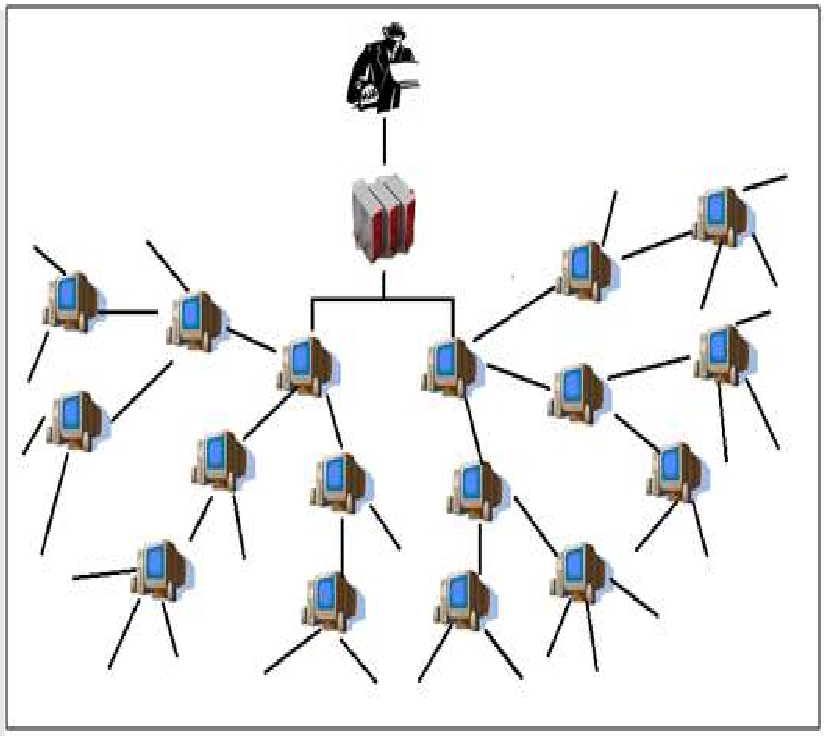
Andrea Noreen D'silva
PG scholar, Dr AIT,Bangalore.

Short Profile

Andrea Noreen D'silva is a PG scholar at Dr AIT,Bangalore.

Co-Author Details :

Vidyarani H. J.
Asst. prof, Dr AIT, Bangalore.



ABSTRACT:
In recent years, botnet is one of the major threats to network security. Many approaches have been proposed to detect botnets by comparing bot features. Usually, these approaches adopt traffic reduction strategy as first step to reduce the flow to following strategies by filtering packets. Botnets have started usingInformation obfuscation techniques include encryption to evade detection. In order to detect encrypted botnet traffic, in this paper we see detection of encrypted botnet traffic from normal network traffic as traffic classification problem. After analyses features of encrypted botnet traffic, we propose a novel meta-level classification algorithm based on content features and flow features of traffic. The content features consist of information entropy and byte frequency distribution, and the flow features consist of port number, payload length and protocol type of application layer. Then we use Naive Bayes classification algorithms to detect botnet traffic.

KEYWORDS
monitoring behaviors and transmission information , Machine learning Classification.

1. INTRODUCTION :

With the rapid development of the network hardware and software, the network speed is enhanced to multi-gigabit. A variety of Internet services, such as web search engines, entertainments, and others, have been provided to people. Therefore, Internet security has become an important role to protect activities on Internet. Botnet has become one major threat to Internet users in recent years. A botnet consists of a large number of bots that are networked computers compromised by malicious attackers. Usually, an attacker controls the bots to launch various types of attacks such as phishing and spamming with a botnet, and thereby receives benefits from a variety of aspects such as economy and social security. Therefore, detecting bots and preventing users from being infected is critical to network security experts and researchers.Most of methods detect bot's activities based on predefined patterns and signatures retrieved from well-known bots Because the most obvious characteristics of botnet is collaborative control, botnets must transmit control command and information by command control communication system, such as sending spam, sending back the data theft to c&c server, downloading updated version of bot program and so on, monitoring behaviors and transmission information in a network candetect botnets. But recently, based on research of various botnets samples, we find some botnets have started using Information obfuscation techniques include encryption to evade detection. Examples include Nugache and Sinit, two p2p-based botnet begin use cryptographic transformations with asymmetric keys RSA e.g. . In more recent versions zues botnet, a http-based botnet, the data sent between the bot and the command and control server is encoded using RC4. All the (C&C) traffic in Storm botnet is encrypted using XOR [5]. This makes it difficult to detect encrypted botnet traffic whether in locality or in gateway detection.

In this paper we explore encrypted botnet detection scheme. In this scheme, firstly, we will analysis which features will vary when the botnet traffic is encrypted. Secondly, we will analysis other botnet flow-based characteristics. Thirdly we will build botnet data model based on encrypted features and flow characteristics. Finally we use machine learning techniques to find out the botnet encrypted traffic.

2 ENCRYPTED BOTNET TRAFFIC

When a botmaster of a botnet-http based send a control command hided in http post message in plaintext like this:

Post/v55/index.phpHTTP/1.0..HOST:Sppa.net..content-Type:text/xml..Content-Length:45.<Retur n a lists emails via http >

If botmaster encrypts the payload with base64,Post/v55/index.phpHTTP/1.0..HOST:Sppa.net.. content-Type:text/xml..Content-Length:45<UmV0dXJuIGEgbGIzdHMgZW1haWxzIHZpYSBodHRwI A==>

We can see that the command payload was obfuscated. The change of entropy can illustrate the disorder of the encrypted botnet traffic. Entropy is actually a measure of disorder

$H(p) = - \ \Sigma \ \_ \ \_ \ \_p\_ \ \log p\_$

In generally we use statistical frequency as estimation value, obtain sample entropy estimator:

H(s) = - Σ _ _ _f_log f_ = Σ _ _ _ _log _ _

H(s)~H(p) is valid when N }Ù m

H(s)~H(p)is valid when N }Ù m . In the previous example, H(s) of payload in plaintext is 3.601,H(s) of encrypted payload with base64 is 4.86251, the value of H(s)
increases when payload is encrypted.

Entropy will help us to assess the randomness of botnet message. Shannon entropy tells us what is the minimal number of bits per symbol needed to encode when the information in binary form (if log base is 2). H(s) converges in probability to H(p) when the length N of s tends to , as soon as each character of string is drawn independently according to the distribution. In the case of the uniform distribution, H(p) ≤logm , this means that

H(p) tends to logm. When characters are bytes, m = 256, so H(p) tends to 8.

Other varied feature of the encrypted botnet traffic is BFD (byte frequency distribution). The message contents are sequence of bytes, and a byte has 256 unique patterns (0~255), thus, counting the occurrence of byte patterns that is often referred as byte frequency distribution. In

an n length message M_ composed of n
characters Σ = (_,_ _. ,,_ _ _,itis) subset
of alphabet Σ = (_,_ _. ,,_ _ _._) For each
b_ ∈Σ = {b_ , b_, . . , b___), it appears in the
message followingp_. We can count its
occurrences f_, and obtain byte frequency
distribution for message
BFD_ : _n__/n , n__/n , ... , n____/n_

1.1 Bias Analysis in Entropy Estimation

The various feature of the encrypted botnet traffic, include increasing entropy and decreasing byte frequency. In fact, encrypting data could raise the randomness of data, andrandomness will increase with the levels of integrity of the encryption.So we could distinguish encrypted message from non encrypted message through evaluating randomness of message,and could distinguish different levels of intensity encryption.

As mentioned, entropy and byte frequency are two math tools to evaluate randomness, this leaves us with a problem, we need to fin reasonable estimators of them.

2 Machine learning Classification

Detectionof botnet traffic from normalnetwork traffic, it could be look as traffic classification problem. According to specific classification goals, classifying traffic based on features passively observed in the traffic.Several methods exist for classifying data and all of them fall into two broad classes: deterministic and probabilistic. Probabilistic classification methods classify data by assigning it with probabilities of belonging to each class of interest. Class assignment is done by considering the class with the largest probability. Machine learning technique is a very significance method for classification problem. The most significant of machine learning is data mining. Machine learning can often be effectively applied to establish relationships between multiple features of internet traffic, improving the efficiency of detect botnet traffic.

Traffic Feature Vector The application of a maching learning classifcation scheme requires the parameterizations of the objects to be classifed. Using these parameters the classier allocates an object to a class. Due to their ability to allow discrimination between classes.

In this section, we will extraction some traffic characteristics as feature vector for machine learning-based classifiers. We will choose characteristics that may be related to the botnet. Characteristics can be divided into two categories, network characteristics and payload characteristics.Andt hese characteristics will be important with botnet,According to this principle we choose network characteristics includeip source address , ip destination address, source port, destination port packet size, and choose payload characteristics include payload entroy, payload byte frequency distribution, as the table1 shows.

	Feature	description	Type
	$I^{JKLO}\_(P)$ entropy	distance from $H\_(u)$	Payload-based
	Q_{RST}	Byte frequency distribution	Payload-based
3	dstIP	IPdestination address	Flow-based
4	srcIP	IP source address	Flow-based
5	srcPrt	source port	Flow-based
6	dstPrt	destination port	Flow-based
5	protocol	application layer protocol	Flow-based
6	PLength	Payload length	Flow-based

Table 1 Traffic Feature Vector

3 Methods

Data sets and Experiment Tools

We use ISOT dataset [9] in our experiments. TheISOT dataset is the combination of several existing publicly available malicious and non-malicious datasets. Malicious traffic include the Zeus, Storm, and Waledac botnets, non-malicious traffic comes from two different datasets, one from the Traffic Lab at Ericsson Research in Hungary [10] and the other from the Lawrence Berkeley National Lab (LBNL)]. The Ericsson Lab dataset contains alarge number of general traffic from a variety ofapplication,including HTTP webbrowsing behaviour.

Detection experiment framework

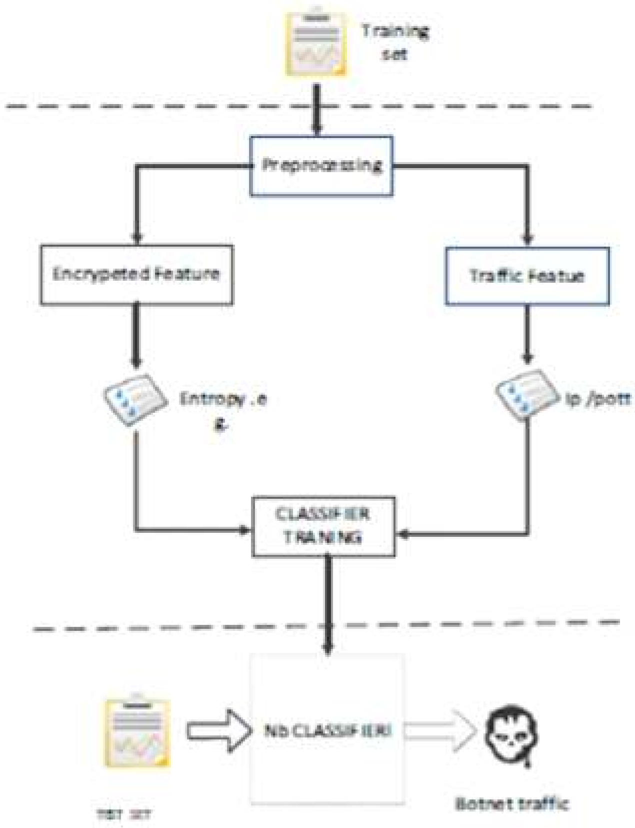


Figure 1 Detection Experiment Framework

Data Preprocessing Stage

The work in this section include extraction of triffic packets from pcap, parsing packets, getting TCP protocol Layer data, eliminating the wrong error data packets, preparing for extraction of feature. The program for this work is done by python.

Then the program exctract key information of packet including ip source address, ip destination address, source port, destination port packet size, packet payload and compute entropy and byte frequency distribution of payload.

After the above steps, we get feature vector composed of sting type (e.g., ip address port numbers, protocol type) and numeric data (e.g. entropy and BFD). We will convert the string type to enumeration type and normalize the value of numeric data.

CONCLUSIONS

For against that botnets have started using Information obfuscation techniques incliud encryption to evade detection. We present a botnet detect scheme that could detect the encryption botnet traffic. In this scheme, fisrtly we established traffic profile modle of encryption botnet traffic, there are import features in the modle include entropy and BFD that could reflect the encryption level and other traffic feature that are the key parractier for the communication pattern of botnet traffic. Then we see detection of Botnet traffic from normal network traffic as traffic classification problem. We use machine learning techniques for classification detector and enter feature vector of traffic profiling into Naive Bayes classifier. ISOT datasets is used for our experiment.

REFERENCES

[1].S. ChangT. E. Daniels, "P2P botnet detection using behavior clustering & statistical tests," in Proceedings of the 2nd ACM Workshop on Security and Artificial Intelligence, 2009, pp. 23-30.
[2].P. Wang, S. Sparks,C. C. Zou, "An advanced hybrid peer-to-peer botnet," Dependable and Secure Computing, IEEE Transactions on, vol. 7, pp. 113-127, 2010.
[3].T. Holz, M. Steiner, F. Dahl, E. Biersack, and F. Freiling. Measurements and mitigation of peer-to-peer-based botnets: a case study on storm worm. In Proceedings of the 1st UsenixWorkshop on Large-Scale Exploits and Emergent Threats, pages 9:1–9:9, Berkeley, CA, USA, 2008. USENIX Association.
[4].C. E. Shannon. A mathematical theory of communication. The Bell system technical journal, 27:379–423, July 1948.
[5].MILLER, G. A. Note on the bias of information estimates. In Information Theory in Psychology; Problems and Methods II-B (Glencoe, IL, USA, 1955), H. Quastler, Ed., Free Press, pp. 95– 100.
[6].J. Olivain and J. Goubault-Larrecq. Detecting subverted cryptographic protocols by entropy checking. Technical report, LaboratoireSpcifica- tionetVrification, June 2006.

Publish Research Article
International Level Multidisciplinary Research Journal
For All Subjects

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Books Review for publication,you will be pleased to know that our journals are

Associated and Indexed,India

- Directory Of Research Journal Indexing
- International Scientific Journal Consortium Scientific
- OPEN J-GATE

Associated and Indexed,USA

- DOAJ
- EBSCO
- Crossref DOI
- Index Copernicus
- Publication Index
- Academic Journal Database
- Contemporary Research Index
- Academic Paper Databse
- Digital Journals Database
- Current Index to Scholarly Journals
- Elite Scientific Journal Archive
- Directory Of Academic Resources
- Scholar Journal Index
- Recent Science Index
- Scientific Resources Database

Review Of Research Journal
258/34 Raviwar Peth Solapur-413005,Maharashtra
Contact-9595359435
E-Mail-ayisrj@yahoo.in/ayisrj2011@gmail.com
Website : www.ror.isrj.org