PREVENTING DNS BASED BOTNET COMMUNICATION FOR DDOS ATTACK

K.Suganya , Asst.Prof C.Gomathi/CSE
University College Of Engineering, Trichirappalli

Abstract

The collection of infected systems called botnet which is controlled by the one named as bot master. Botnets are capable to initiate many DoS attacks. Denial of Service attack (DoS) is attempted by the attacker to prevent legal user and in Distributed Denial of Service attack (DDoS) the attacker sends more number of unwanted request at the same time to infect particular host. The botnet controller sends queries to bot by using command and control channel to infect the host. Domain Name Service (DNS) used for better botnet command and control channel. Domain Name Service (DNS) provides the distributed infrastructure for storing and updating data. The user intension based query attacks are detect by statistical analysis method. The machine intension based query attacks are detect by the Naive bayesian classification process.DNS based query attacks are enhanced to provide privacy preservation. The small query analysis mechanism also integrated with the query attack detection system.

Index Terms – Network security,botnet detection,command and control

1.Introduction

A botnet is a collection of infected system which affect the security by sending malicious content. Bot receiving the commands from the botnet controller to infect the host. Examples of communication channel to send infected queries are E-mail, HTTP (Hyper Text Transfer Protocol) mechanism and IRC (Internet Relay Chat). Now DNS playing a major role to transfer infected queries to attack a host. The botnet is a collection of infected system that is controlled by one party called botmaster. Botmaster can send a command to the bots for affecting the system. Botnet operator or the botnet controller provides instruction that the way to affect system online. A botnet controller sends out the viruses to infect the user computer by malicious application. The bot in a infected machine logs into particular command and control server. In a financial profit the spammer collects the services of the botnet from operator and instructs them to send those messages. Botnets are capable to initiate many DoS attacks and involves in various activities like theft of credit card numbers and collects financial information.

The botnets are termed as a malware, botnets are using same malware and controlled by different entity. The collection of computers running malicious software is termed as botnets. The controller or operator manage the group as command and control server. The botnet operator protocol includes both client program and server program for an operation. The bot runs the program and communicate with command and control server. Botnet communication consists of several controllers developed with command hierarchies. Botnet architecture using different topology for a command and control. The different topologies are star, multi server, hierarchial and random. The bots are receiving the command from botmaster and act for their instruction. In the command includes the type of request and the time for sending the request that to be hiden. Botnets are using DNS as a better command and control channel for sending the command from botmaster to bots.

DoS attack means that the attacker to prevent from using resources, submitting request from one to one computer. DDoS attack means that the large number of host are combined to send a useless service request , packets at the same time, attacks are generated in a many to one. One software called Spyware which is used to send user activities to creators like credit card number, password and so on. Email spam are the email messages differed from the message in malicious, if the user visits the websites unknowingly create traffic. Fast Flux is technique followed by DNS for hiding the malware. Scareware is software used to install malware programs.

Passive OS fingerprinting used to identify the attacks, the administrator configure firewall to take necessary steps using the information. The Rate based Intrusion Detection system used in a specific hardware. Network intrusion detection system used to monitoring network in an effective. Host based techniques and network based techniques are used to identify the bot activities. Classification is the process of finding a model describing and differentiates the classes. The data mining system can be categorized with the techniques. The techniques are described with user interaction by query based systems or the methods of data analysis like machine learning and statistics. Bayesian classifiers are the statistical classifiers, predict the class members probability, the probability belongs to the same class. The high accuracy and fast computation is achieved in this classifier. Bayesian classifiers are known as the Naïve Bayesian classifier, the attribute value on a class is independent to the other attribute values.

2. Related Work

Exploring DNS protocol as a practical C&C channel and identifying its limitations have not been scientifically studied. Various proof-of-concept botnet C&C systems via unconventional media exist, such as via bluetooth and social networks. In comparison, our work is useful beyond the specific DNS-based communication channel studied in two aspects:

- We present new quantitative techniques and evaluation regarding the detection and construction of general-purpose distributed stealthy communication systems, including temporal strategies for making stealthy communication and statistical content analysis.
- We give a practical technique that is useful in domain flux from the attacker’s perspective, namely MC-based domain name generation.

For DNS-based anomaly detection, Karasaridis et al. [7] described the use of the Kullback-Leibler distance to measure byte distribution in DNS datagrams. Dagon [3] proposed to quantify how anomalous the number of queries for each domain name during an hour in a day with Chebyshev’s inequality and distance measures previously used for examining anomalous payloads. DNS-based anomaly detection approaches are presented in [2] for detecting botnet C&C activities. One method is to detect dynamic domain names whose query rates are abnormally high or temporally concentrated using outlier detection metrics such as Chebyshev’s inequality. Our work describes stealthy DNS behaviors whose querying patterns are hard to distinguish with legitimate domains, which make the counting-based detection less effective.

Stone-Gross et al. [1] observed the use of domain flux in Torpig botnet, where new communication domains are generated periodically and registered by the C&C server. Torpig bots communicated with the server over HTTP, after resolving the domain name. Patterns of fast-flux botnets are measured and analyzed in [9]. For example, renewal using piggyback method was proposed to piggyback cached DNS records to DNS queries to refresh expired cached records [6]. Related domains may also be piggybacked in DNS queries [5], e.g., to include i.cnn.net in the DNS packet for www.cnn.com as they are likely to be requested together by the browser. Covert channel has been heavily analyzed in the context of traffic-analysis prevention [4] and routing anonymity [8]. Our work differs from them in that we focus on designing practical covert channels across the Internet. Our work is complementary to host-based malware detection and prevention solutions, such as the cryptographic provenance verification technique.

3. DNS queries for Command and Control

DNS queries for command and control, provides the distributed infrastructure for storing, updating data.DNS as a high traffic channel and data can be easily hide in it. In an automatic domain flux, the domains can be changed frequently and it can be detect by Markov chain property, the comparison of normal traffic and DNS traffic can be detect by statistical method.DNS tunneling mechanism used for communication, tunneling for bypass firewalls. The communication used for botmaster and bot as, Codeword Mode and Tunneled Mode.

3.1. Codeword Mode

One way communication between botmaster and bot, provides the attack commands. The codeword are generated in random manner. Botnet controller decides upon from those code words. The codeword represents the type of attack. The client queries DNS server, in server that is controlled by botnet operator contains the command information. If the codeword representing DOS attacks, response as the target of DOS attacks. If the codeword representing update, response as the IP address client may contact.

3.2. Tunneled Mode

Two way communication between botmaster and bot, transferring random number between client and server. They are upstream communication, downstream communication. The client submit their stolen data to server, client submit the data as a query. Encode the data and encode to construct the host name. The server issuing commands to the clients, server reacts upon the query from clients. Encode the response in a certain format and returns the host name. Prevent DNS caching, the server set a Time To Live, DNS protocol does not allow the server to initiate the connection .Always the client initiate the connection from the server. Straightforward querying patterns are identified easily, the total count of DNS queries for unique domains are identified. For hiding query activities, exponentially distributed query and Piggybacking Query. In an Exponentially Distributed Query, the bot distributes its query. The bot send a query, computing the time interval t. The bot send a query after its time t. In an Piggybacking Query the bot listens normal queries and mix with their bot queries. More number of sites contains content from the multiple domains like legal sites containing advertisements. The bot listens the legitimate DNS queries by traffic sniffing. Evaluating these strategies, use KS test method. Testing for DNS logs for anomaly detection, collect the details from the host and listen for particular time period. In a test proves that small bot queries are not to be detectable.

3.3. Domain Names

The long lived domains are easy to manage and cheaper. In a Domain flux, the short lived domains are used in botnet command and controller. The bots and botmaster generate the domains periodically. Domain names are generated in, hash chain based method and automatic text generation. In a Hash based method, domain names are generated in random and use only the Digits. In automatic text generation using Markov Chain property, generate with 37 features. The patterns in the shared data to be extracted. Botnets are using sub directories for their communication, the botmaster not run a web server. The botnets using Third Level Domain (3LD) instead of sub directories, it is easily detect by ratio based method. The botmaster using Second Level Domain (2LD) with 3LD, that is not to detected. Evaluating DNS based communication system use, packet inspection and statistical analysis. In a Packet Inspection, detect the packet with their header. The UDP packets are not to be handled ,using TCP and analyzing with data not with header. In a statistical detection takes clear storage and take computation cost.

3.4. Attacks using Command and Control Scheme

The botnet command and control channel as a protocol used by the bots and botmaster for their communication. The botmaster can send a attack commands to bot for steal the information. Botnet operator are using IRC and HTTP server as a command and control channel .Email and the Bluetooth provides better infrastructure for their communication. DNS queries as a stealthy command and control channel provides the distributed infrastructure for storing data.DNS as a highly traffic channel and data can be hide easily.DNS tunneling is a technique for transmitting data for bypass firewalls. In the DNS traffic requires to translate the domain names to IP address and back and make more traffic. The techniques for hiding query activities are piggybacking query strategy and exponentially distributed query strategy. In the automatic domain flux method, domain names can be changed frequently and evaluate it by Markov chain analysis. A statistical method helps to detect the anomalies in the content of DNS data whether it contains normal request or anomaly request.

3.5. Problem Statement

DNS provides the distributed infrastructure for storing data, stealthy command and control channel. The following drawbacks are identified in the existing system.

- DNS sensitive data access is not controlled
- Data leaks are not accurately detected
- User intention is required for anomaly detection
- Detection latency is high

4. Bayes Classifier for Attack Detection

The domain name query based attacks for service providers are detected using the bayes classifier algorithm. In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausibleefficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests. An advantage of Naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

5. RSA Algorithm

The domain name service sensitive attributes are secured using the RSA algorithm. The Rivert, Shamir, Adelman (RSA) scheme is a block cipher in which the Plaintext and cipher text are integers between 0 and n-1 for some n. A typical size for n is 1024 bits or 309 decimal digits.

Key Generation

Select p,q
p and q both prime , $p \neq q$
Calculate $n = p \times q$
Calculate $\phi(n)=(p-1)(q-$
$1)$ Select integer e
$\gcd(\phi(n),e) = 1; 1 < e <$
$\phi(n)$ Calculate d
$d = e^{-1} \bmod$
$\phi(n)$ Public key
$KU = \{e, n\}$
Private key
$KR = \{d, n\}$

Encryption

Plaintext
$M < n$
Cipher text
$C = M^{e} \pmod{n}$

Decryption

Cipher text C
Plaintext
$M = C^{d} \pmod{n}$

6. Preventing DNS Based Botnet Communication

Botmaster can use DNS as a command and control channel (Fig. No: 6.1) for sending command to the Bots. Botmaster can send DNS data to the Bots or clients. Bot can send a service request to the service provider. The service provider can use statistical analysis to detect user intension based request. The machine intension based requests are to be detecting by the classification process. In classification using patterns to differentiate the normal request and anomaly request. The service provider sends back the response to the client.

The Domain Name Service based attack detection system is designed to handle command and control message process over DNS query values. The DNS query analysis is performed with statistical analysis. The system analyzes the DNS query values to identify the insertion of DDoS attack detection information. Machine learning approach is used to detect the DDoS attacks. The system is divided into five major modules. They are Request Observer, Statistical analysis, Bayesian analysis, Command and control request handler and Security and Privacy process.

The request observer module is designed to collect service requests from the clients. Statistical analysis module is designed to detect DDoS attacks using statistical methods. Bayesian approach based attack detection mechanism uses the classification process. The command and control handler is designed to manage the DNS query based botnet communication activities. Security and Privacy module are designed to provide security and privacy for DNS parameters.

6.1. Request Observer

Stream requests are collected from various clients. The server assigns session instances to new stream requests. The stream requests are processed by the web server. The responses are redirected to the clients.

6.2. Statistical Analysis

The statistical analysis model is used to detect the Service attacks. The request flow and its similarity are analyzed with frequency and interval values. The user session patterns are learned from the request flow analysis Attack decisions are made with the support of DNS query format. The requests are assigned with normal or attack labels in the analysis.

6.3. Bayesian Analysis

The Bayesian analysis model is used to detect DNS query based attacks using machine learning approaches. The classification model is used for the request category identification process.

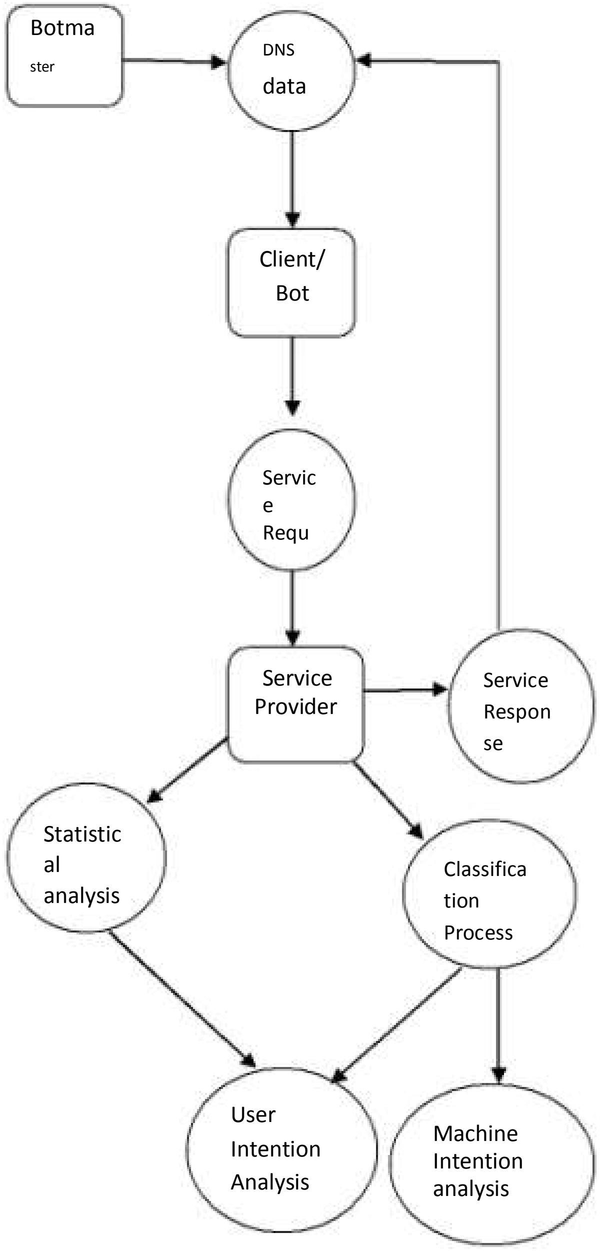

Fig. No: 6.1. Preventing DNS Based Botnet Communication

The Bayesian classification algorithm is used for the attack detection process. The attacker requests are rejected by the service provider.

6.4. Command and Control Request Handler

The command and control requests are exchanged between the botmaster and bots. The DNS query values are used to exchange the command and control requests. The DNS query responses are verified with the domain name server details. Command and control instructions are dropped with reference to their category.

6.5. Security and Privacy

The DNS query values are submitted from the clients in the networks. The DNS responses are prepared by the domain name server. The user can add additional parameters to the DNS responses. The RSA algorithm is used to protect the DNS parameter values. Sensitive attribute values are protected with cryptography methods.

7. Conclusion

The DNS query based attack detection scheme is enhanced to provide privacy preserved data traffic analysis. Automated anomaly detection is adapted to the system. Naiva Bayesian classification technique is integrated to the system. Small query analysis mechanism is integrated with the system. The system performs botnet communication detection and control operations. DDoS attack detection mechanism is included in the system. The system improves the detection accuracy with minimum latency. DNS parameter security is also provided by the system.

REFERENCES

[1] B. Stone-Gross, M. Cova,"Your Botnet Is My Botnet: Analysis of a Botnet Takeover," Proc. ACM 16th Conf. Computer and Comm. Security (CCS), Nov. 2009.

[2] R. Villamarı´n-Salomo´n “Identifying Botnets Using Anomaly Detection Techniques Applied to DNS Traffic,” Proc. IEEE Fifth Consumer Comm. and Networking Conf, 2008.

[3] Dagon, “Botnet Detection and Response, the Network Is the Infection,” Proc. Domain Name System Operations Analysis and Research Center Workshop, 2005.

[4] Newman and A. Serjantov, “Metrics for Traffic Analysis Prevention,” Proc. Privacy Enhancing Technologies Workshop, 2003.

[5] Shang and Wills, “Piggybacking Related Domain Names to Improve DNS Performance,” Computer Networks, 2006.

[6] B. Jang and H. chul Kim, “DNS Resolution with Renewal Using Piggyback,” J. Comm. and Networks, Aug. 2009.

[7] A. Karasaridis, D.A. Hoeflin, “Detection of DNS Anomalies Using Flow Data Analysis,” Proc. IEEE GlobeCom, 2006.

[8] I. Moskowitz and A.R. Miller, “Covert Channels and Anonymizing Networks,” Proc. ACM, 2003.

[9] X. Hu, M. Knysz, and K.G. Shin, “Measurement and Analysis of Global IP-Usage Patterns of Fast-Flux Botnets,” Proc. 30th Ann. Int’l Conf. Computer Comm. (INFOCOM), 2011.