# BOTNET DETECTION BASED ON DNS RECORDS AND ACTIVE PROBING

Iria Prieto, Eduardo Magaña, Daniel Morató and Mikel Izal
*Public University of Navarre, Campus Arrosadia, 31006 Pamplona, Spain*

Keywords: Botnet, Command and control, Domain name, DNS record, WHOIS.

Abstract: Computers connected to Internet are constantly threatened by different types of malware. One of the most important malware are botnets that convert infected computers into agents that follow actions instructed by a command-and-control server. A botmaster can control thousands of agents. This means a significant capacity to accomplish any kind of network attack (DoS), email spam or phishing. In this paper, communication peculiarities with the command-and-control server are used to provide an identification of computers infected by a botnet. This identification is based mainly in DNS records of registered domains where command-and-control servers are hosted. Therefore, processing overhead is reduced avoiding per packet or per flow network supervision.



## 1 INTRODUCTION

Computers, and specially computers connected to the Internet, are becoming an essential tool in working and entertainment environments. It is usual to send confidential information through e-mail, make an online bank transaction, online shopping, etc. Unfortunately, the popularity of Internet has been accompanied by the growth of network attacks which try to obtain benefit from this information. Some attacks proceed from personal computers that can be infected with unwanted software known as malware.

Malware covers a large range of software like viruses, worms, trojans, spyware, loggers and botnets. Recently, the type of malware growing at a fastest rate is botnets (Zhaosheng et al., 2008). A botnet is characterized by having a set of compromised computers called bots. These bots are controlled remotely by a command-and-control (C&C) server managed by the botmaster. They use a special protocol that is known as C&C channel. Through this channel, the botmaster can send instructions to bots to perform new attacks, infect other machines or update botnet software. This channel can use well-known protocols such as IRC, HTTP or P2P protocols in order to hide itself from any try of identification (John et al., 2009)(Zhaosheng et al., 2008).

Currently there is a large collection of active botnets in the Internet. Some of them are Rustock (Chiang and Lloyd, 2007)(John et al., 2009), Zeus (Binsalleeh et al., 2010), Conficker (Porras et al., 2009), Kraken (Jae-Seo et al., 2008)(Stone-Gross et al., 2009), etc. Botnets can propagate attacks through networks quickly and, furthermore, those attacks can have high impact because of the high number of controlled agents. An example of this impact is shown in (Zeljka, 2009), that describes how in year 2009 89.5 billion unsolicited emails were sent every day by compromised computers participating in a botnet.

Early detection of botnets is very important as it can provide a certain grade of trust in network services. Even Internet Service Providers are interested in its identification because of the great percentage of unwanted traffic generated. Antivirus and antispyware programs try to identify botnet software in infected computers with traditional schemes based on code signatures. However, botnet software mutates quickly and therefore those schemes are not useful. Similar identification schemes can be performed by firewalls or intrusion detection systems, this time applying signature-based schemes over network traffic in the C&C channel. Again, these protocols change continuously or even they are encrypted so identification results are not good enough. Besides, overhead processing is significant in high-speed networks as signature checks have to be performed per packet.

Usually C&C servers are identified by one or several domain names that have to be known a priori by bots. This will allow bots to contact the C&C server and check for their availability. Therefore, previously

Prieto I., Magaña E., Morató D. and Izal M..
BOTNET DETECTION BASED ON DNS RECORDS AND ACTIVE PROBING.
DOI: 10.5220/0003522903070316
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2011), pages 307-316
ISBN: 978-989-8425-71-3

to any C&C communication, bots have to resolve IP addresses of already known domain names for C&C servers. These domain names have specific characteristics that can be used to identify suspect domain names to be part of a C&C server and therefore it can identify computers participating in a botnet.

In this paper, a new method to identify computers infected by botnets is proposed. This method will combine in-depth analysis of DNS records with extra information obtained from active probing in order to obtain an indicator of suspect for domain names. Detection capabilities will be demonstrated in a real scenario.

The rest of the paper is organized as follows. Section 2 presents the state of the art in botnet detection techniques using C&C network traffic. In section 3, the network scenario and the traffic traces used are presented. Section 4 introduces selected metrics to use in the identification, based on experimental analysis. In section 5, architecture of botnet identification system is presented. Evaluation and results of the proposal are presented in section 6. Finally, conclusions and future work are presented.

## 2 STATE OF THE ART IN BOTNET DETECTION BASED ON NETWORK TRAFFIC

Detection of botnets at an early stage of infection is a challenging task. The great majority of techniques in the state of the art take into account the main characteristic of botnets: the C&C channel which allows owners to update and control bots. Botnet detection techniques are usually oriented to discover those channels.

Botnet detection techniques can be classified based on how and which data is processed into signature-based, anomaly-based and DNS-based (Feily et al., 2009).

Signature-based detection methods look for certain patterns in network traffic like recognizable protocol headers, payloads, packet sizes or interarrival times. The C&C protocol of well-known botnets can be characterized and this characterization can be used to identify their traffic. This characterization can be implemented as rules in an intrusion detection system (IDS). One specific characteristic of botnets is the existence of concrete C&C servers whose IP addresses can be known apriori through this characterization. Therefore, any computer exchanging traffic with those IP addresses can be identified as a bot. In (Goebel and Holz, 2007), detection of IRC bots is made applying data mining techniques over nicknames, IRC servers and used ports.

Modifications of botnet software, protocol or C&C servers are usual. For example, the mutation speed of STORM botnet is estimated as once every 30 minutes (Grizzard et al., 2007). Therefore, signature-based detection methods have practical limitations.

Another approach to botnet detection consists on characterizing normal traffic and, later, identifying deviations with the presence of botnet infection. This approach is called anomaly-based detection. Botnets can use application protocols implemented over standard IRC or HTTP protocols, so it is not easy to identify C&C communication from a normal chat or web traffic. In (Binkley and Singh, 2006), bots connected to an IRC channel are identified by their specific activity: IRC messages used, communication profile, number of sent/received packets, number of shared channels, etc. In (Gu et al., 2008b), extensions are made to support IRC and HTTP-based protocols, this time using correlation of communications from multiple bots and level of network activity. Part of the identification is also based on signatures as packet payloads are analyzed.

Data mining techniques are used in Botminer (Gu et al., 2008a). It uses K-means algorithm to cluster data with metrics corresponding to normal traffic and metrics corresponding to malicious activity. Anomaly-based detection schemes have to be tuned up for specific scenarios and this is one of their main disadvantages. Also, the rate of false positives can be quite high depending on the percentage of total identification that we wanted to achieve. Some apriori knowledge about C&C protocols is also needed, but in less extension than with signature-based methods.

The last approaches for botnet detection methods in the state of the art are DNS-based methods. Botnets are controlled by one or more C&C servers whose IP addresses have to be known by bots. Usually, domain names are known instead of IP addresses. Those domain names can be hardcoded in the bot code or can be updated online via some configuration file. Therefore, before contacting C&C servers, a DNS resolution request has to be made to map the known domain name into the corresponding IP address. In (Dagon, 2005), bots are detected with the hypothesis that botnet domain name requests are concentrated in a time window for several infected computers. Therefore, correlations in DNS requests are supervised. Also, the time-to-live (TTL) of domain names is considered. The TTL indicates the number of seconds for which the mapping of domain name to IP address is valid. Botnets usually use TTL values around few seconds in order to be able to change the mapping

dynamically and avoid C&C servers to be discovered. However, short TTL values also appear in other domain names as those registered by Content Delivery Networks (CDN) so the results are not conclusive. Whitelists can be used, but again a priori extensive knowledge is required.

In order not to be easily identified, botnets usually register multitude of domain names for the same C&C server. Bots usually check for these domain names and most of those mappings are not existent in each moment. Even unreliable and temporal DNS servers are used. In those cases, a "NXDOMAIN (Non-Existent Domain)" answer is obtained for the domain name request. In (Villamarn-Salomon and Brustoloni, 2008), the rate of NXDOMAIN responses is used as an indicator of botnet presence. However, last versions of botnets do not show this behavior (Feily et al., 2009). In (Feily et al., 2009), time proximity in DNS requests between different bots is used to identify botnets. The traffic of a high number of computers has to be analyzed in order to be able to find correlations in DNS requests. This could be the case of an Internet trunk of an Internet Service Provider. However, the detection needs to propagate between several hosts before being detectable.

Domain names used in C&C servers sometimes follow a pattern because they are generated algorithmically. In (Yadav et al., 2010), a methodology to detect botnet domain names is presented looking for those patterns in domain names that are different to those generated by humans.

# 3 NETWORK SCENARIO

For the proposal and evaluation, real traffic traces have been obtained from Public University of Navarre (UPNA, Spain). Its Internet access link has been monitored specifically for DNS request/response packets. The main significant results have been obtained from a traffic trace dated on September 15-17th, 2010. In this trace, 4,807,719 DNS requests have been performed that correspond to 452,601 different domain names. DNS responses are 3,962,032, corresponding to 405,338 different domains and 67,671 domain names have returned NXDOMAIN at least once.

Also a testbed with Zeus, Conficker and Kraken botnets has been deployed. This testbed has been secured with a honeywall (Jones and Romney, 2004) and it has allowed to obtain direct information about domain names requested by infected hosts. A blacklist of domain names corresponding to C&C channels have been discovered this way. It contains 100,108 domains.

# 4 CHARACTERIZATION OF DNS METRICS TO BE USED IN BOTNETS IDENTIFICATION

An in-depth study of relation between botnets and domain names has been made in order to improve current proposals in the state of the art. Current metrics in the state of the art have been evaluated: DNS TTL, DNS NXDOMAIN and DNS pattern. New metrics with significant importance have been identified and evaluated: DNS record age, DNS e-mail record, authoritative DNS server, DNX MX record and web presence.

## 4.1 DNS TTL

These works (Dagon, 2005) (Holz et al., 2008) (Perdisci et al., 2009) have stated the relation between short TTLs in DNS names definition and the presence of botnets behind those names. However, in our revision, TTLs with zero value have been discovered in multitude of domains, most of them because of misconfiguration of authoritative DNS servers. Also short TTLs have been found in successful services such as Google (TTL=46, 68, 300, etc.), YouTube (TTL=66, 70, 89, etc.) or Facebook (TTL=1, 6, 7, 8, etc.). Those short TTL values are chosen in order to use DNS as load balancer (as for example in Content Delivery Networks) and allowing to adapt users better to performance and availability of end-servers.

In figure 1, the cumulative distribution functions of TTL values for normal and botnet domain names are plotted. As explained before, a large percentage of normal domains uses short TTLs. For example, 60% of normal domains use TTL values equal or less to 500. Therefore, using DNS TTLs in botnet identification is not significant nowadays.

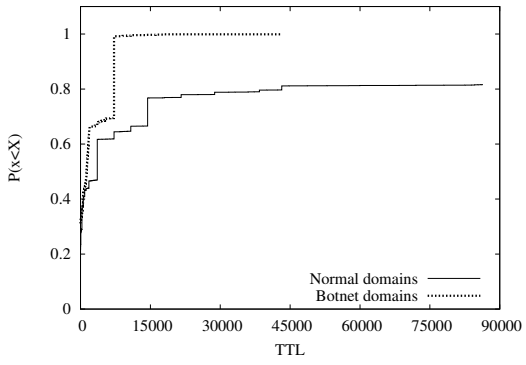


Figure 1: Cumulative distribution of TTL for normal and botnet domains.

## 4.2 NXDOMAIN

The NXDOMAIN technique proposed in (Villamarn-Salomon and Brustoloni, 2008) is based on rating the number of non-existent answers in domain name requests. This technique has been evaluated in our traces and only 172 domains answered NXDOMAIN responses with at least one correct answer. This correct answer is necessary to validate the existence of the domain. All those 172 domains correspond to normal domains so the NXDOMAIN rate is not significant in botnet evaluation.

The number of domains that always answered with NXDOMAIN responses were 67,427. Only 311 of them correspond to botnet domains and it is not possible to identify botnet and normal domains with this ratio. Therefore, results are not as good as expected.

## 4.3 Domain Name Pattern

Metrics based on domain name patterns (Yadav et al., 2010) have been discarded because they need a previous analysis of how DNS names registered for botnets are generated. Our goal is to provide a generic scheme of botnet identification without previous individual characterization because those characteristics can change easily in different versions of the software.

## 4.4 DNS Record Age

This metric is related with the creation date of the domain name under study. When a domain name is registered, information about domain owner, creation date and other characteristics are stored. This information is accessible via WHOIS service. Therefore, a WHOIS request is enough to obtain the creation date of certain domain. It has been observed that botnet domains are usually very young with one year or less age. It is reasonable because sooner or later those domains are blacklisted, old domains are not reused and new ones have to be registered continuously to allow normal botnet operation. Therefore, DNS record age can be used to identify domains suspect of being assigned to botnets. In (Passerini et al., 2008), record age is used to characterize suspicious domain names extracted from emails. In our proposal, domain names are extracted directly from all DNS requests in the network.

For weighting this metric, domains with less than one year of age have more importance. Our proposal is reflected in equation 1. This ratio will be bounded between 0 and 1. CaptureDay indicates the date when DNS request was intercepted, and CreationDate is the date when domain name register was created. Both numbers in Unix Epoch format can be subtracted to get the difference in days.

$$DNS\,record\,age = \frac{1}{1+\frac{(CaptureDate-CreationDate)-1}{365}} \quad (1)$$

Figure 2 shows the cumulative distribution function of the number of days since a domain name was registered. Botnet domains are concentrated in the first hundreds of days. Normal domain names are distributed linearly the first 5-6 years as new domain names are created continuously. Following, some older domain names are unregistered and linearity is lost. DNS record age is, therefore, a good indicator of how a domain name is suspected to be a botnet. However, it is not enough by itself because every day normal domain names are created and most of them do not correspond to botnets.

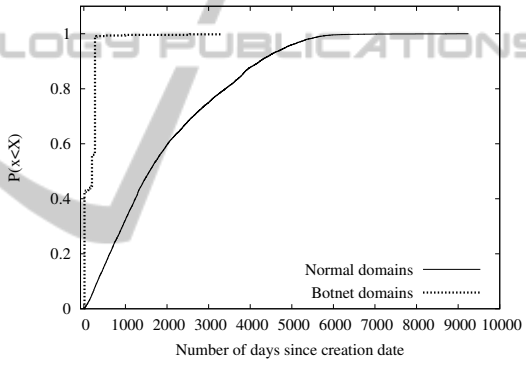


Figure 2: Cumulative distribution function of the number of days since a domain name was registered.

## 4.5 DNS e-Mail Record

Also from WHOIS information, details about domain owner can be obtained like names, addresses, telephones or emails of administrative and technical contacts. Some general behavior has been observed on the patterns used in hosting email for botnets domains. Those emails are mainly free-hosting based, like hotmail, yahoo, gmail, live, etc. Therefore, the presence of this type of free email hosting accounts can be used to identify botnets domains. This metric will be 1 for domains registered with those types of emails and 0 otherwise. Those emails can be present also in normal domains, so results are not conclusive.

For our data, figure 3 presents the percentage of free-hosting based emails for normal and botnet domains. For some of those free-hosting, the differences are significant and therefore usable in botnet identification. The spenglers service corresponds to spen-

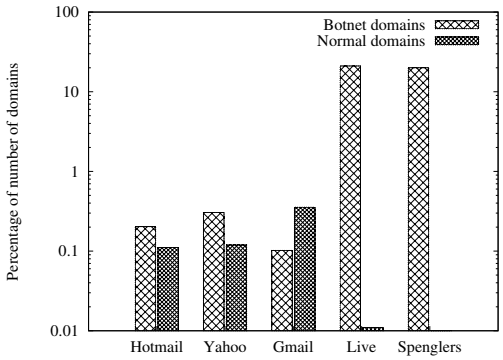


Figure 3: Percentage of free-hosting based emails for normal and botnet domains.

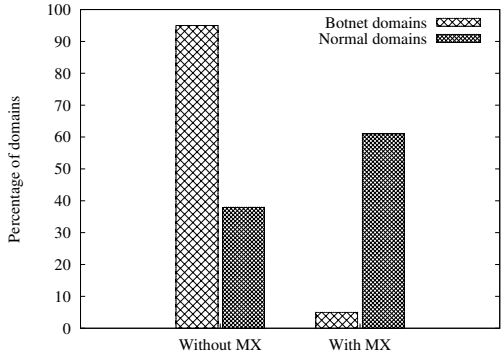


Figure 4: Percentage of domains with MX registers for normal and botnet domains.

glers.biz, a specific domain name registered to generate emails. In this case, 197 conficker domain names were registered using splengers emails.

## 4.6 Authoritative DNS server

A authoritative DNS server is the last DNS server responsible of resolving a domain name and its subdomains. Those servers can be any host running a DNS server software. However, botnets tend to use a concrete set of authoritative DNS servers. Therefore, a blacklist of suspected authoritative DNS servers, where previously registered botnet domains have been detected, can be created. Those authoritative DNS servers can also host normal domain names so the metric is not conclusive. This metric will be 1 for those in blacklist and 0 otherwise.

Analyzing the traffic trace under study, 477 botnet domain names and 3,108 normal domain names are hosted in DNS serves in blacklist. This means that 48.4% of botnet domains (477 of 985) and less than 1% of normal domains are hosted in those specific DNS servers. Therefore, this metric is also significant in botnet identification.

## 4.7 DNX MX Record

Domain names can register A address records or CNAME canonical name records, but also MX mail exchange records which map a domain name to a list of message transfer agents (email servers) for that domain. Botnet domain names usually do not register MX records because they are not used, but normal domain names (mainly for web services) usually have associated MX records. Therefore, the absence of MX records is another hint to find botnet domain names. Again, for this metric, the absence of MX records will be scored as 1 and 0 otherwise.

In figure 4 results are summarized. 95.96% of botnet domain names and 38.71% of normal domain names do not have MX record.

## 4.8 Web Presence

As stated before, the great majority of domain names are registered to be used in web hosting services, and at the same time, email service is provided or subdomains are defined for different tasks. As those web hosting services are accessible through standard 80 port, the presence of this web service can be checked actively. Therefore, the absence of response to a standard HTTP request directed to that domain will be another hint to locate domains suspect to be botnets. In figure 5 web presence results are presented for normal and botnet domains. Around 65% of botnet domains does not have web presence.

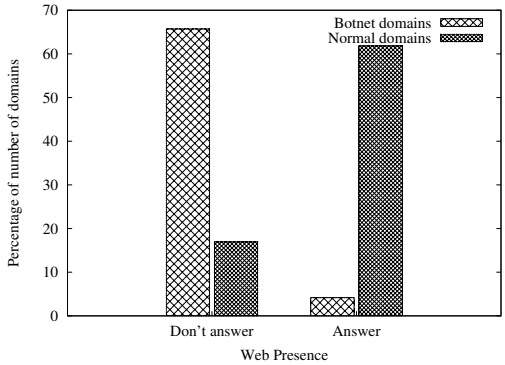


Figure 5: Web presence for normal and botnet domains.

# 5 BOTNET DETECTION SYSTEM (BDS): STRUCTURE AND OPERATIONS

A system called BDS (Botnet Detection System) has been implemented considering previous metrics.

Those metrics have advantages related with the low processing power needed as only DNS requests have to be analyzed. This DNS requests can be less than 1% of total traffic in a network, and therefore adaptable for high-speed networks. Moreover, botnet detection based on DNS requests allows to detect botnets in an early infection stage and to perform active countermeasures like instructing a firewall to block suspicious traffic.

BDS has been programmed in C language using WHOIS client tool (Net-Whois, 2010), wget (Wgettool, 2009), DNS lookup utility (DiG, 2009), and perl script (DNSDUMP, 2010). However, to make the evaluation easier, the input is fed with traffic traces captured previously.

Figure 6 shows modules used in BDS implementation: DNS filter, DNS processor, database, extra data collector and evaluator. The first two modules work in real time extracting DNS requests from network traffic and storing this information in the database. The last module works asynchronously, extending the information about new domains available in the database.

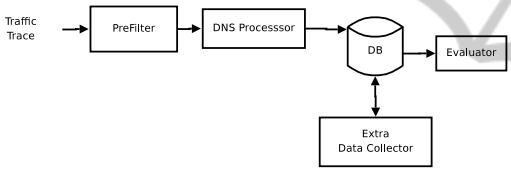


Figure 6: Botnet detection system (BDS) block diagram.

First, the input traffic trace with all packets captured in an Internet connection is filtered looking for DNS traffic. Basic filtering based on UDP/TCP port 53 is applied. This is the “DNS Filter” module. Second, DNS traffic is sourced to the “DNS Processor” module where which domain name is requested each time and what response information is answered is identified. This module parses DNS requests/responses obtaining information per seen domain name. In request packets, extracted information is full domain name, source IP address (host requesting DNS resolution), destination IP address (DNS server answering the request) and timestamp. In response packets, extracted information is resolved IP addresses and domain name aliases. All this information is stored in the “Database” module. Frequently used domain names can be requested several times, so all these requests will be stored in the database with different timestamps and specific values if they change.

Asynchronously, “Extra data collector” monitors the database detecting when a new domain name is added in order to obtain extra information about the domain needed in the proposed metrics. This extra information proceeds from the following queries:

- WHOIS query: each time a new domain is detected, a WHOIS query is performed in order to get extra information about the domain name record. This extra information is DNS record age, DNS e-mail record and authoritative DNS server that are stored in database.
- DNS MX request: a specific DNS request asking for MX servers is performed. This will allow to get the DNX MX record that is also stored in database
- Web checking at port 80: web presence metric is obtained by checking the presence of a web server listening at port 80 of IP address resolved in the original DNS request. This metric is stored in database.
- Blacklists update: when new botnet domain names are detected, their authoritative DNS servers are marked in a blacklist to be considered suspicious.

This extra information is costly to obtain (for example, several seconds are usually needed to perform a WHOIS query) but this process only happens when a new domain is detected or after a timer in the range of days to detect changes in these metrics over time. All information is stored in the database to speed up later repetitions of same domain names. First working hours, new domain names are added continuously, but after some hours or days very few domain names are added per hour. In a new deployment, pre-calculated data can be provided in the database with metrics about most common domain names, because this information can be shared between different network scenarios.

For certain domain, once extra data has been collected if it was not collected before, “Evaluator” module is in charge of conforming a suspicion rate that measures if a domain name has probability to be a botnet. Suspicion rate is calculated applying some weights to a combination of previous metrics. Details will be shown in following section. This rate will allow to determine if a domain name is suspect of being a botnet or not.

## 6 EVALUATION

In the measured scenario, only DNS requests with successful A record or NXDOMAIN responses are considered, discarding not answered requests. This means 198,357 domain names under consideration. A botnet domain name blacklist has been conformed

using information obtained from the botnet testbed with infections of Zeus, Conficker and Kraken botnets, and specific online databases as Zeus tracker (ZeusTracker, 2011). The conformed botnet domain name blacklist is over 1,500 domains, and 985 of them have been identified in traffic traces of the scenario.

Metrics proposed in this paper are evaluated in table 1. This table shows identification rates of botnet domain names based on each of proposed metrics applied independently. True positives indicate the percentage of botnet domain names correctly identified. The percentage of normal domain names misidentified as botnets is titled false positives. The percentage of botnet domain names not identified by the metric is shown as false negatives. In DNS record age, a threshold of 0.5 has been considered to identify botnets. Although metrics like DNS MX records have a high percentage of true positives, also the rate of false positives is high. Most interesting metrics will be those with a higher difference between true positives and (false negatives+false positives). Therefore, DNS record age, DNS e-mail record or authoritative DNS server are better metrics.

Suspect ratio is composed by a weighted sum of proposed metrics: DNS record age, DNS e-mail record, authoritative DNS server, DNX MX record and web presence. These weights are obtained from the differences in identification shown in table 1. Resulting suspect ratio has been normalized. Suspect ratio has been calculated for all domain names observed in the captured traffic trace under study. Then, domain names have been sorted based on this suspect ratio hoping to have botnet domain names in the first positions.

Figure 7 shows the percentage of botnet domain names identified correctly as more domain names are considered in the sorted domain name list based on suspect ratio. It can be observed that almost the first 400 domain names are classified correctly as botnets, and later the proportion of botnet decreases, getting around 55% botnets in the first 800 domain names. Besides botnet domain names, domain names associated to other types of malware are detected, but they are negligible, around 2% of the total number of botnet domain names.

Considering the total number of 674 botnet domain names and 198,357 normal domains, the percentage of total identification shown in figure 8 represents the percentage of domain names identified correctly/wrongly as more domain names are considered in the domain name list sorted by suspect ratio. As seen before, for the approximately first 400 domain names, 70% of total number of botnet domain names

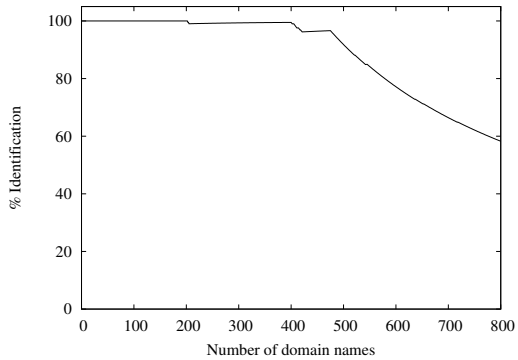


Figure 7: Botnet identification percentages in domain name list sorted by suspect ratio.

in trace are identified with only 3% of false positives.

The remaining 30% botnet domain names have inspection ratios indistinguishable from normal domains. This is due to the lack of extra information for those botnet domain names: there is no WHOIS response and the only available metrics are DNX MX record and web presence. Improvements in WHOIS querying could improve results, for example, balancing queries between several WHOIS servers.

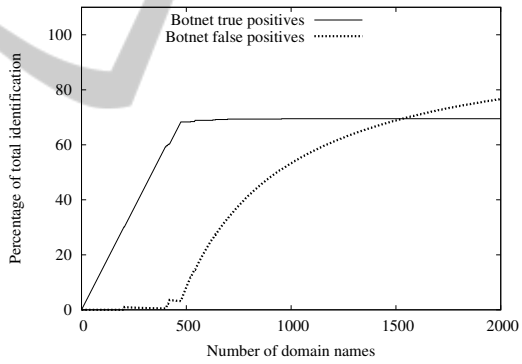


Figure 8: Percentages of total identification in domain name list sorted by suspect ratio.

A decision threshold has to be defined to differentiate suspect ratio that identifies botnet or not. If this threshold is too high, only a reduced percentage of botnet domain names will be identified. If this threshold is too low, a big percentage of false positives will result from the identification process. Figure 9 presents the cumulative distribution function of suspect ratio for botnet domain names and normal domain names. Botnet domain names are concentrated in higher suspect ratios. Most of them have a suspect ratio larger than 0.75. Normal domain names are concentrated in lower values of suspect ratio. Considering the full set of domain names, a suspect ratio around 0.75 can be considered to identify over 95% of botnet domain names and a low percentage of false

Table 1: Botnet domain names identification for each metric independently.

| Metric | True positives | False positives | False negatives |
|---|---|---|---|
| DNS record age | 69.2% (464) | 1.228% (3,070) | 30.1% (208) |
| DNS e-mail record | 30.6% (206) | 0.954% (1,892) | 69.1% (466) |
| Authoritative DNS server | 69.1% (466) | 1.57% (3,119) | 30.7% (207) |
| DNX MX record | 96.1% (648) | 25.4% (50,328) | 3.71% (25) |
| Web presence | 95.1% (641) | 27.03% (53,609) | 4.75% (32) |

positives (normal domain names considered as botnets).

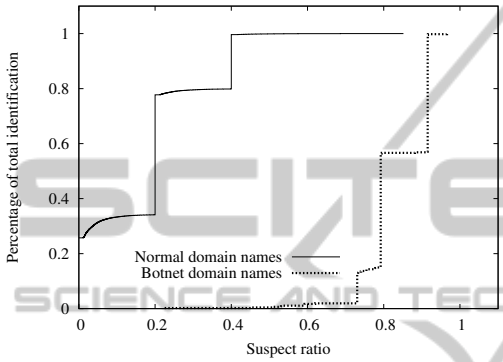


Figure 9: Cumulative distribution function for suspect ratio.

Proposed BDS is compared with results obtained by NXDOMAIN technique (Villamarn-Salomon and Brustoloni, 2008). In this case, a ratio is calculated for each domain that considers the number of NXDOMAIN responses between the total number of DNS requests. Two cases are distinguished. First, considering the ratio only if at least one correct answer is received. With this limitation, only active C&C channels are considered and non-active domain names are ignored. It will be called NXDOMAIN-1. Second, NXDOMAIN-2 will consider all domain names active or not.

Table 2 shows the number of domains that can be analyzed. In BDS, all domains can be processed, but with NXDOMAIN-1 and NXDOMAIN-2 only domains that return at least one NXDOMAIN response can be processed. In NXDOMAIN-1, only 172 normal domains answer at least with a correct DNS response for normal domains, and 0 for botnet domains. Therefore, NXDOMAIN-1 is not useful. NXDOMAIN-2 increases the number of processed botnet domains to 311 but it is a half of the total number of botnet domain names present at the traffic trace. NXDOMAIN-2 provides a very bad rate of false positives making this technique not usable. Our proposal, BDS, improves identification percentages significantly, achieving 68% of botnet identification with only 3.18% of false positives.

# 7 CONCLUSIONS

The expansion of botnets has increased over the last years. Therefore, their identification has become very important. A new technique for botnet identification has been presented in this paper. It is based on analyzing DNS requests and responses for domain names used in identification of Command&Control server. Extra information is obtained for each domain from WHOIS service, checking for MX servers availability and checking for web services presence.

As only DNS queries have to be processed, very high speed links can be monitored for the presence of botnets with low processing overhead. Most domain names are repeated by different users over time because they identify most common Internet services or web pages. Information about domain names is stored to be reused, meaning that in normal operation only new domains have to be checked for botnets. Detection is obtained in an early stage of infection because DNS queries are the first action that an infected computer performs.

A suspect ratio is defined based on a set of metrics: DNS record age, DNS e-mail record, authoritative DNS server, DNX MX record and web presence. Achieved results are promising. In an evaluation over an university Internet link, 65% of botnet domain names are identified with only 3% of false positives. This data outperforms results with techniques in the state of the art like NXDOMAIN-based.

Improvements are possible considering correlation of DNS request from the same IP addresses. Once a computer is identified as being part of a botnet, following DNS requests from the same computer have more probability to be related to the botnet. Even identifying only part of botnet domain names, all infected computers can be identified because each of them will request for dozens of domain names and at least one of them can be identified. Correlation of DNS queries between different computers can also be used to improve identification rate.

The number of queries for each domain name can be also used to improve the suspect ratio. However, we have not been able to use it because of the low number of infections in the network under study. This

Table 2: Number of analyzed domains and botnet identification rate for different techniques.

| Technique | Number of normal domains | Number of botnet domains | % true positives | % false positives |
|---|---|---|---|---|
| BDS | 250,062 | 674 | 68% | 3.18% |
| NXDOMAIN-1 | 172 | 0 | 0% | 0% |
| NXDOMAIN-2 | 67,427 | 311 | 20% | 99.87% |

would allow, for example, to ignore those misspelled domain names because they would be requested only once.

## ACKNOWLEDGEMENTS

This work was supported by S21sec labs through the research project SEGUR@, funded by the Spanish Ministry of Industry, Tourism and Trade, on the framework of CENIT programme with reference CENIT-2007 2004.

## REFERENCES

Binkley, R. and Singh, S. (2006). An Algorithm for Anomaly-based Botnet Detection. *Computer Science, PSU, USENIX SRUTI: '06 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet*.

Binsalleeh, H., Ormerod, T., Boukhtouta, A., Sinha, P., Youssef, A., Debbabi, M., and Wang, L. (2010). On the analysis of the zeus botnet crimeware toolkit. In *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, pages 31 –38.

Chiang, K. and Lloyd, L. (2007). A case study of the rustock rootkit and spam bot. In *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*, pages 10–10, Berkeley, CA, USA. USENIX Association.

Dagon, D. (2005). Botnet Detection and Response, The Network is the Infection. In *1st DNS-OARC Workshop*, Santa Clara, CA.

DiG (2009). Tool from the package dnsutils. http://www.ubuntuupdates.org/packages/show/105545.

DNSDUMP (2010). Perl script that captures and displays DNS messages. http://dns.measurement-factory.com/tools/dnsdump/.

Feily, M., Shahrestani, A., and Ramadass, S. (2009). A Survey of Botnet and Botnet Detection. In *Third International Conference on Emerging Security Information, Systems and Technologies*, Athens/Glyfada, Greece.

Goebel, J. and Holz, T. (2007). Rishi: Identify bot contaminated hosts by irc nickname evaluation. In *First USENIX Workshop on Hot Topics in Understanding Botnets (HotBots'07)*, Cambridge, MA.

Grizzard, J., Sharma, V., C. Nunnery, B. K., and Dagon, D. (2007). Peer-to-peer botnets: Overview and case study. In *First USENIX Workshop on Hot Topics in Understanding Botnets (HotBots'07)*, Cambridge, MA.

Gu, G., Perdisci, R., Zhang, J., and Lee, W. (2008a). BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection. In *17th USENIX Security Symposium (Security'08)*, San Jose, CA.

Gu, G., Zhang, J., and Lee, W. (2008b). Botsniffer: Detecting botnet command and control channels in network traffic. In *15th Annual Network and Distributed System Security Symposium (NDSS'08)*, San Diego, CA.

Holz, T., Gorecki, C., Rieck, K., and Freiling, F. C. (2008). Measuring and detecting fast-flux service networks. In *15th Annual Network and Distributed System Security Symposium (NDSS'08)*, San Diego, CA.

Jae-Seo, L., HyunCheol, J., Jun-Hyung, P., Minsoo, K., and Bong-Nam, N. (2008). The activity analysis of malicious http-based botnets using degree of periodic repeatability. In *Security Technology, 2008. SECTECH '08. International Conference on*, pages 83 –86.

John, J. P., Moshchuk, A., D.Gribble, S., and Krishnamurthy, A. (2009). Studying spamming botnets using botlab. In *Proceedings of the 6th USENIX symposium on Networked systems design and implementation*, pages 291–306, Berkeley, CA, USA. USENIX Association.

Jones, J. K. and Romney, G. W. (2004). Honeynets: an educational resource for it security. In *Proceedings of the 5th conference on Information technology education*, CITC5 '04, pages 24–28, New York, NY, USA. ACM.

Net-Whois (2010). Module for parsing WHOIS information. http://search.cpan.org/∼ivsokolov/Net-Whois-Parser-0.05/.

Passerini, E., Paleari, R., Martignoni, L., and Bruschi, D. (2008). Fluxor: Detecting and monitoring fast-flux service networks. In *Proceedings of the 5th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, DIMVA '08, pages 186–206, Berlin, Heidelberg. Springer-Verlag.

Perdisci, R., Corona, I., Dagon, D., and Lee, W. (2009). Detecting malicious flux service networks through passive analysis of recursive dns traces. In *Computer Security Applications Conference, 2009. ACSAC '09. Annual*, pages 311 –320.

Porras, P., Sadi, H., and Yegneswaran, V. (2009). A foray into confickers logic and rendezvous points. In *In USENIX Workshop on Large-Scale Exploits and Emergent Threats*.

Stone-Gross, B., Cova, M., Cavallaro, L., Gilbert, B., Szydlowski, M., Kemmerer, R., Kruegel, C., and Vigna, G. (2009). Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference*

*on Computer and communications security*, CCS '09, pages 635–647, New York, NY, USA. ACM.

Villamarn-Salomon, R. and Brustoloni, J. (2008). Identifying Botnets Using Anomaly Detection Techniques Applied to DNS Traffic. In *5th Annual Consumer IEEE Communications and Networking Conference (CCNC2008)*.

Wget-tool (2009). GNU Wget package for retrieving files using HTTP, HTTPS and FTP. http://www.gnu.org/software/wget/.

Yadav, S., Reddy, A. K. K., Reddy, A. N., and Ranjan, S. (2010). Detecting algorithmically generated malicious domain names. In *Proceedings of the 10th annual conference on Internet measurement (IMC2010)*, IMC '10, pages 48–61, New York, NY, USA. ACM.

Zeljka, Z. (2009). Top 10 botnets and their impact. http://www.net-security.org/secworld.php?id=8599.

ZeusTracker (2011). The ZeuS Tracker tracks ZeuS Command and Control servers. https://zeustracker.abuse.ch/.

Zhaosheng, Z., Guohan, L., Yan, C., Fu, Z., Roberts, P., and Keesook, H. (2008). Botnet research survey. In *Computer Software and Applications, 2008. COMPSAC '08. 32nd Annual IEEE International*, pages 967–972.