Insert here your thesis' task.

Czech Technical University in Prague

Faculty of Information Technology

Department of theoretical computer science

Master's thesis

Detection of DNS Anomalies via Data Mining Analysis of Network Traffic

Bc. Michal Pohořelý

Supervisor: Mgr. Rudolf Blažek, Ph.D.

26th February 2014

Acknowledgements

I would like to thank my supervisor Mgr. Rudolf Blažek for support during writing this thesis.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In Prague on 26th February 2014 . . . .

Czech Technical University in Prague
Faculty of Information Technology
© 2014 Michal Pohořelý. All rights reserved.
This thesis is a school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).

Citation of this thesis

Pohořelý, Michal. *Detection of DNS Anomalies via Data Mining Analysis of Network Traffic.* Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2014.

Abstract

This thesis is concerned with anomalies in DNS traffic and their identification and clasification. Part of this theses is also ananalysis of methods acceptable for classification. The main part is selection and implementation of selected method for classification of DNS anomalies with their results.

Keywords DNS, anomalies, datamining, neural nets, CZ.NIC.

Abstrakt

Tato práce se zabývá anomáliemi v DNS provozu a jejich identifikací a klasifikací. Součástí práce je také rozbor metod pro použití při klasifikace. Hlavní částí práce je výběr a implementace vybrané metody pro klasifikaci DNS anomálií s příslušnými výsledky.

Klíčová slova DNS, anomálie, datamining, neuronové sítě, CZ.NIC.

Contents

List of Figures

Introduction

Chapter 1

DNS - Domain Name System

Domain Name System is a hierarchical distributed naming system for any resource connected to the Internet [17]. You can imagine its function like a phone book what is translating human friendly domain names (e.g. www.nic.cz) to IP address. IP address is the identification of every computer or other device connected to the Internet and must be unique (except of NAT). IPv4 is 32 bits long and is composed of four octets like numbers in range of values 0 to 255 delimited by dots (e.g. 217.31.205.50). So you can see that form in IP address is much more difficult to memorize than www.nic.cz. And more difficult is situation in IPv6 which is 128 bits long and is composed of eight 16-bits hexadecimal values delimited by colon (e.g. 2001:1488:800:400::130). This is the main reason why the DNS is being used.

1.1 Normal DNS Usage

1.1.1 Components of DNS

There are three major parts of DNS [13]

- The domain name space and resource records which are specifications for a tree tree structured name space and data associated with the names. Every node of a tree contains a subset of information. Query operation are trying to extract information from a particular subset. Query contains the domain name and DNS type.

- Names servers are server programs which hold information about the domain tree's structure and set information. They can have information about any part of domain tree but practicaly they serve only a part of domain space. They also have a pointer to other name servers to serve any part of domain space. If the server knows everything about the subset of some part of domain space we call it authoritative name server.

- Resolvers are client applications that are extracting information from name servers in responce to client requests. They need to have access at least to one name server. Resolver asks the name server and name server give him an answer or link to next name server where the information should be found.

1.1.2 DNS record

Structure of DNS record [12]

- NAME - owner name, the name of the node to which this resource record pertains
- TYPE - 16bit type code
- CLASS - 16bit class code
- TTL - 32bit time interval of record lifetime
- RDLENGTH - 16bit length of RDATA
- RDATA - variable length string that describes the resource

1.1.3 DNS resolution

DNS resolution is proccess of finding an IP address of requesting host. The usual scenario is on Figure:1.1

At first client asks his local DNS server (usually ISP DNS server). If the local DNS server doesn't know the answer he will then ask root DNS server. He will give him answer where to find the TLD DNS server for .cz domain. Local DNS server will ask this TLD DNS server and he will get an answer where to find nic.cz DNS server. Local DNS server will continue there and he will get an exact answer with IP address of labs.nic.cz and he will give to client.

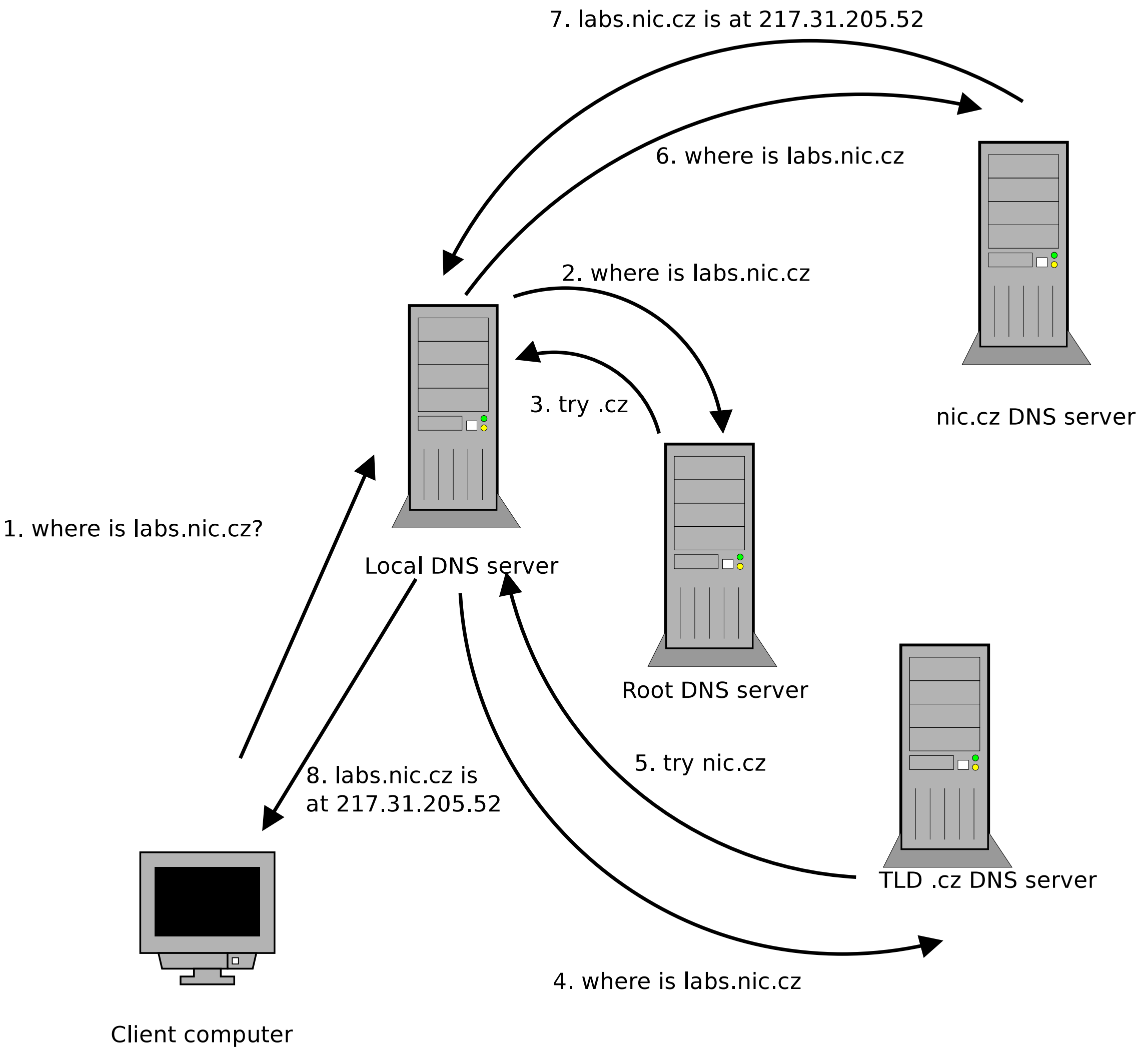

Figure 1.1: Resolution of requsted IP address

1.2 Anomalous DNS Traffic

Anomalous traffic is when DNS protocol is being used for any other reason than for translating domain names to IP address or reverse. We can then split this traffic to following groups.

Chapter 2

Analysis of DNS Attacks

2.1 DoS Attacks

DoS attacks can be classified into two major categories [9]

- Try to exploit vulnerabilities in the implemented software (ping of death). This type of vulnerability is mostly fixed in new devices and nowadays is the flooding attack more common.

- Try to overhelm system resources i.e. memory, CPU, network bandwith. In this type of attack is usually used the amplification DNS attack that lies in the fact that DNS respond messages may be substantially larger than DNS query messages. Attacker sends DNS request with victims IP address and the respond will deliver to victims network. This is because open recursive DNS servers are not checking request validity.

2.2 MiM Attacks

This type of attack is used to redirect user to fake website to gain private information of the client. When the client attempts to connect the website he sends a DNS request. The attacker captures this request and sends a fake answer with IP address with his webserver. Victim will start to communicate with attackers web server instead of the real one. [2]

2.3 DNS Tunelling

This is the technic when the attacker wants to circumvent a security policies. Typical example is to illegally browse the web when access fee is requested. DNS requests are usually not filtered on firewall so this communication is open. Attacker encapsulates his data into DNS packets. [5]

2.4 Botnets

Malicious botnets are distributed computing platforms predominantly used for illegal actvities such as launching Distributed Denial of Service (DDoS) attacks, sending spam, trojan and phishing emails, illegally ditributing pirated media and software, force distribution, stealing information and computing resource, e-business extortion, performing click fraud, and identity theft. Victim's computer beacame a part of botnet usually by launching some infected application.[6]

2.5 Other DNS Related Attacks

Chapter 3

Methods of DNS Attack Detection and Classification

3.1 Ad-hoc Methods

3.2 Signatures

3.3 Data-Mining Methods

3.3.1 Random Forest

3.3.2 Neural Nets

3.4 Statistical Methods

Chapter 4

Description of Selected Method

Chapter 5

Application Design

Chapter 6

Implementation

Chapter 7

Measurements and Testing

Conclusion

Bibliography

[1] Antonakakis, M.; Perdisci, R.; Lee, W.; etc.: Detecting Malware Domains at the Upper DNS Hierarchy. In *USENIX Security Symposium*, 2011. Available at WWW: <http://www.usenix.org/event/sec11/tech/full_papers/Antonakakis.pdf>

[2] Bai, X.; Hu, L.; Song, Z.; etc.: Defense against DNS Man-In-The-Middle Spoofing. In *Web Information Systems and Mining*, *Lecture Notes in Computer Science*, volume 6987, edited by Z. Gong; X. Luo; J. Chen; J. Lei; F. Wang, Springer Berlin Heidelberg, Jan. 2011, ISBN 978-3-642-23970-0, pp. 312–319. Available at WWW: <http://dx.doi.org/10.1007/978-3-642-23971-7_39>

[3] Bernhard M Hammerli: *Critical information infrastructures security: second international workshop, CRITIS 2007, Malaga, Spain, October 3 - 5, 2007 ; revised papers*. Berlin [u.a.: Springer, 2008, ISBN 3540890955 9783540890959 9783540891734 3540891730. Available at WWW: <http://dx.doi.org/10.1007/978-3-540-89173-4>

[4] Dewaele, G.; Fukuda, K.; Borgnat, P.; etc.: Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures. In *Proceedings of the 2007 workshop on Large scale attack defense*, 2007, p. 145–152. Available at WWW: <http://dl.acm.org/citation.cfm?id=1352675>

[5] Ellens, W.; Żuraniewski, P.; Sperotto, A.; etc.: Flow-Based Detection of DNS Tunnels. In *Emerging Management Mechanisms for the Future Internet*, *Lecture Notes in Computer Science*, volume 7943, edited by G. Doyen; M. Waldburger; P. Čeleda; A. Sperotto; B. Stiller, Springer Berlin Heidelberg, Jan. 2013, ISBN 978-3-642-38997-9, pp.

124–135. Available at WWW: <http://dx.doi.org/10.1007/978-3-642-38998-6_16>

[6] Feily, M.; Shahrestani, A.; Ramadass, S.: A Survey of Botnet and Botnet Detection. In *Third International Conference on Emerging Security Information, Systems and Technologies, 2009. SECURWARE '09*, 2009, pp. 268–273, doi:10.1109/SECURWARE.2009.48.

[7] van der Heide, H.; Barendregt, N.: DNS anomaly detection. In *WWW-Dokument, staff. science. uva. nl/ delaat/sne-2010-2011/p17/report. pdf*, 2011. Available at WWW: <http://ext.delaat.net/rp/2010-2011/p17/report.pdf>

[8] Jiang, N.; Cao, J.; Jin, Y.; etc.: Identifying suspicious activities through DNS failure graph analysis. In *2010 18th IEEE International Conference on Network Protocols (ICNP)*, 2010, pp. 144–153, doi: 10.1109/ICNP.2010.5762763.

[9] Kambourakis, G.; Moschos, T.; Geneiatakis, D.; etc.: Detecting DNS Amplification Attacks. In *Critical Information Infrastructures Security*, *Lecture Notes in Computer Science*, volume 5141, edited by J. Lopez; B. Hämmerli, Springer Berlin Heidelberg, Jan. 2008, ISBN 978-3-540-89095-9, pp. 185–196. Available at WWW: <http://dx.doi.org/10.1007/978-3-540-89173-4_16>

[10] Karasaridis, A.; Meier-Hellstern, K.; Hoeflin, D.: NIS04-2: Detection of DNS Anomalies using Flow Data Analysis. In *IEEE Global Telecommunications Conference, 2006. GLOBECOM '06*, 2006, pp. 1–6, doi: 10.1109/GLOCOM.2006.280.

[11] Mikle, O.; Slaný, K.; Veselý, V.; etc.: Detecting_Hidden_Anomalies_in_DNS_Communication-2011.pdf - Hledat Googlem. In *CZ.NIC*, 2011. Available at WWW: <https://www.google.cz/search?q=Detecting_Hidden_Anomalies_in_DNS_Communication-2011.pdf&ie=utf-8&oe=utf-8&rls=org.mozilla:cs-CZ:official&client=firefox-a&gws_rd=cr&ei=Vxl8UtDsO--O7QbIwoCIAw>

[12] Mockapetris, P.: Rfc 1034: Domain names-concepts and facilities, 1987. 1987.

[13] Mockapetris, P.: RFC 1035: Domain names: implementation and specification (November 1987). 1987.

[14] Nazario, J.; Holz, T.: As the net churns: Fast-flux botnet observations. In *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*, 2008, p. 24–31. Available at WWW: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4690854>

[15] Nelson, M. M.: *A practical guide to neural nets /*. Addison-Wesley Publishing Company: Reading, 1994, ISBN 0-201-63378-7.

[16] Vraštiak, P.: *Hledání anomálií v provozu DNS*. Diplomová práce, VUT Brno, Brno, 2011.

[17] wikipedia: Domain Name System. Jan. 2014, page Version ID: 587867901. Available at WWW: <http://en.wikipedia.org/w/index.php?title=Domain_Name_System&oldid=587867901>

Appendix A

Contents of CD

Visualise the contents of enclosed media. Use of `dirtree` is recommended. Note that directories src and text with appropriate contents are mandatory.

```
│ readme.txt ..................... the file with CD contents description
├─ data ......................................... the data files directory
│  └─ graphs ..................... the directory of graphs of experiments
│     ├─ *.eps ........................................ the B/W graphs
│     ├─ *.png ........................................ the color graphs
│     └─ *.dat .................................. the graphs data files
├─ exe ................. the directory with executable WBDCM program
│  ├─ wbdcm .................. the WBDCM program executable (UNIX)
│  └─ wbdcm.exe ........... the WBDCM program executable (Windows)
├─ src ................................... the directory of source codes
│  ├─ wbdcm ........................ the directory of WBDCM program
│  │  └─ Makefile ........... the makefile of WBDCM program (UNIX)
│  └─ thesis ........... the directory of LaTeX source codes of the thesis
│     ├─ figures ........................... the thesis figures directory
│     └─ *.tex .................. the LaTeX source code files of the thesis
└─ text ........................................ the thesis text directory
   ├─ thesis.pdf .................... the Diploma thesis in PDF format
   └─ thesis.ps ....................... the Diploma thesis in PS format
```