Big Data and Computer Security

DNS & botnets

Prof. Jean-Noël Colin, Prof. Bruno Dumas
Prof. Benoît Frenay, Prof. Renaud Lambiotte

University of Namur

UNIVERSITÉ
DE NAMUR

Introduction

- Objectives
    - define and structure a global approach solve the problem
    - identify useful features
    - apply visualization techniques
    - analyse data
- The important thing to demonstrate is your ability to analyse a problem in depth, and identify the elements and methods that can help you solving it

Domain Name Service

- Originally defined in RFC1034 & RFC1035, further developed and extended in many other RFCs
- Usually transported over UDP, but also over TCP, on port 53
- Client/server protocol
- Actors

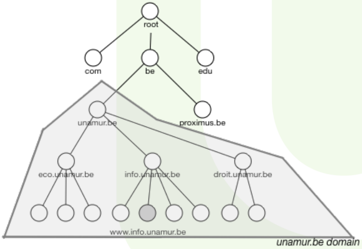
  Client needs to resolve a name

  Resolver builds and sends the queries to the nameserver and processes the response

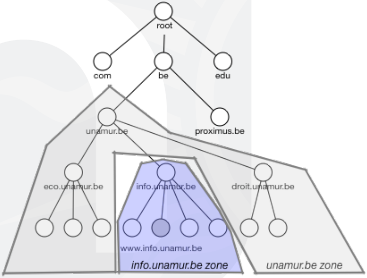
  NameServer handles queries and produces responses

DNS Protocol

Domains vs zones



DNS domain



DNS zone

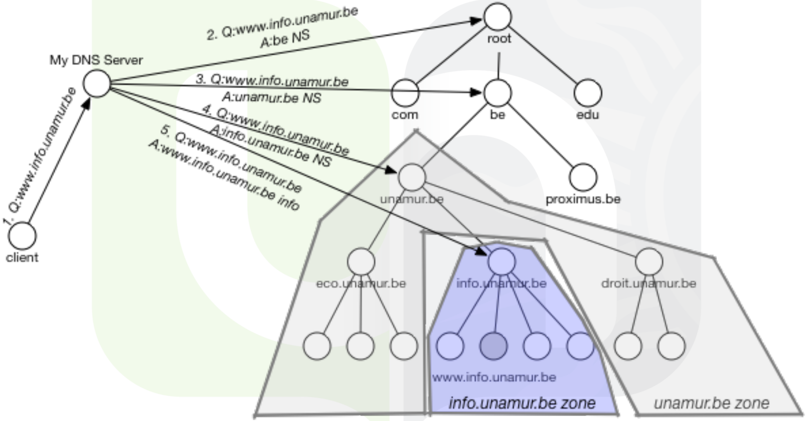
DNS Protocol

Recursive vs non-recursive (iterative) query

recursive queried nameserver has to reply with the requested information if it exists, with an error otherwise. It can't refer to another authority

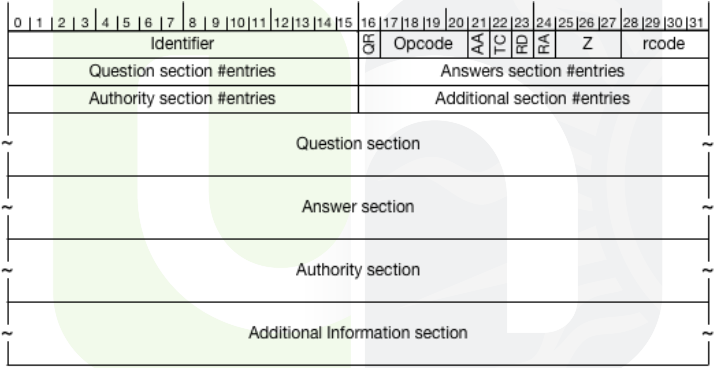
iterative queried nameserver replies with the best answer *it already knows*; if it doesn't know the answer to the query, it returns data about the nameservers closest to the domain name in the query *in its local data*

DNS Protocol



DNS Query

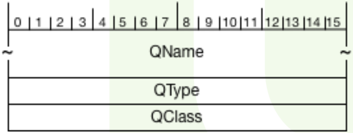
DNS data structures



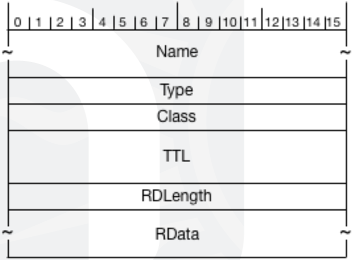
<table>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="16">Identifier</td><td>QR</td><td colspan="4">Opcode</td><td>AA</td><td>TC</td><td>RD</td><td>RA</td><td colspan="3">Z</td><td colspan="4">rcode</td></tr>
<tr><td colspan="16">Question section #entries</td><td colspan="16">Answers section #entries</td></tr>
<tr><td colspan="16">Authority section #entries</td><td colspan="16">Additional section #entries</td></tr>
<tr><td colspan="32">Question section</td></tr>
<tr><td colspan="32">Answer section</td></tr>
<tr><td colspan="32">Authority section</td></tr>
<tr><td colspan="32">Additional Information section</td></tr>
</table>

DNS Message (query **and** response)

DNS data structures



DNS Question



DNS Resource Record

DNS RR types

- IP addresses and names: A, AAAA, PTR, CNAME
- NS authority: NS, SOA
- mail information: MX
- DNSSEC Info and keys
- additional information: TXT, HINFO, MINFO
- see http://www.iana.org/assignments/dns-parameters/dns-parameters.xhtml

DNS uses and abuses

- Some for good

  RFC5782 management of domains blacklists

  RFCs 4034,4025,2535,2930,2230,4255. . . management of cryptographic materials (RRSIG, DNSKEY, NSEC RRs. . .)

  RFC4408,6376 email validation (SPF – Sender Policy Framework, DKIM – Domain Keys Identified Mail signatures)

- Some for bad

  DNS signaling use DNS queries to exfiltrate stolen data; use DNS responses to send commands from C&C to bots

  DNS tunneling use DNS message to tunnel protocols like http, ftp. . .

Introduction

What is a botnet?

- a system of computers that are under the control of a criminal (the botmaster), after having been infected by a piece of malware that took advantage of a vulnerability or exploit to provide remote access and control of the machine

What is it used for?

- sending spam
- carry out DDoS attacks
- data and identity theft
- illegal file storage and sharing
- spread of malicious pieces of payload
- proxy malicious communications (to hide mothership)
- manipulation of polls, cracking of passwords. . .

Botnet topologies

- star topology
    - not very resilient or resistant to detection
- multi-server topology, with interconnected servers
    - a bit better
- tree (hierarchical) topology
    - not the full tree is visible to everyone
    - resilient to the take down of a subtree
    - possible to rent part of the botnet
- random topology
    - no central C&C
    - P2P communications

Botnet communications

- Why?
  - send commands (C&C → bots)
  - send code updates (C&C → bots)
  - send collected data (bots → C&C)
- Channels
  - IRC
  - HTTP
  - P2P
  - DNS
  - pop3, sms. . .
- all require to know the IP address of the C&C node

Botnet communications

How does a bot reach C&C?

- knowing its IP address (easy to trace)
- knowing its domain name and using DNS to lookup IP (single IP: easy to trace)
- knowing its domain name and using DNS to lookup IP (multiple rotating IPs : difficult to trace)
- knowing its domain nameS and using DNS to lookup IP (multiple names, rotating IPs: difficult to trace)
- use multiple NameServers
- use proxy bots to proxy between bots and C&C and hide actual C&C identity

Botnet & DNS

Fast Flux services – domain flux

- a single IP address is mapped to multiple domains, either known by bots or dynamically generated (DGA)
- detection more difficult due to domain diversity (incl. TLD)
- wildcarding subdomains: define a RR *.mydomain.com with 1 IP, and have bots query for <anything>.mydomain.com; botmaster knows which bots are active
- Domain Generation Algorithms (DGA): botmaster regularly generates a set of domain names and registers/unregisters them quickly; only one is used as C&C; bots have to look up names until the find the right one

  Ex. Conficker-C generated 50.000 domains per day, distributed across 110 TLDs

Botnet & DNS

Fast Flux services – IP Flux

- a single domain is mapped to multiple IP addresses; using multiple addresses decreases the probability of detection

  single flux relies on ‘honest’ DNS servers; bots send DNS queries and NS responds with rotating addresses for C&C

  double flux introduces an extra level of indirection; relies on ‘honest’ and ‘fake’ DNS servers; bots send queries to honest NS, which refers to ‘fake’ or compromised NS with rotating addresses; here it is the addresses of NS that rotate

Botnet & DNS

What could reveal the presence of a bot on your network?

- query name
  - dictionary words?
  - # different characters? # digits? special characters?
  - name length? # domain components?
  - suspicious names (ex. yahoo or microsoft as 3LD)
- multiple machines looking for similar names may reveal infected machines using DGA

Botnet & DNS

What could reveal the presence of a bot on your network?

- returned IP:
  - IP belongs to DSL lines
  - number of returned IP outside of common range (1-3)
  - multiple IP in different AS may reveal fastflux
- use silent IP for communication between C&C and agents
  - use private IP (127.0.0.1, 192.168/16. . .) in response when C&C does not need to be contacted, to hide real C&C address

Botnet & DNS

What could reveal the presence of a bot on your network?

- TTL: a short TTL may indicate fastflux
- packet length: long packets may indicate tunneling
- #IP/domain over time: changin IPs for a domain may indicate fastflux
- window (first seen-last seen): legitimate domains should resolve to consistent set, while malicious names should appear during limited windows of time

The project

Objective

- identify DNS queries that are likely to originate from a bot trying to reach its C&C center

Practicalities

- report file to be submitted on WebCampus (check that you have access) or via email
- Deadline: 29/11/2017 18:00CET

The project

Some hints

- define your goal
- describe what you want to achieve, and how
- implement several approaches, measure and compare them
- re. visualization, scatter plots (Lecture 1. slide 37) and internal/external monitoring (Lecture 2. slides 12-16) may help

Resources

- https://documents.unamur.be/index.php/s/wCIgXKMwqT8SFeB a set of pcap files captured on the university network, containing DNS packets only, unfiltered (in folder ‘données projet 2’)
- https://github.com/jncolin/dns_pcap_extractor a python script that parses a pcap file and extracts the DNS information; adapt to create a csv file with the information you need from header fields, questions and answers
- https://www.alexa.com/topsites probably contains valid URL
- http://www.malwaredomainlist.com/, http://www.malwaredomains.com/ or https://isc.sans.edu/suspicious_domains.html hopefully contain illegal domain names
- http://www.secrepo.com/misc/zeus_dga_domains.txt.zip contains 31000 DGA domains from Dec 2014
- http://www.secrepo.com/ contains pointers to numerous datasets of various kind

References I

- A. L. Buczak and E. Guven, *A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection*, IEEE Communications Surveys Tutorials **18** (2016), no. 2, 1153–1176.

- Leyla Bilge, Sevil Sen, Davide Balzarotti, Engin Kirda, and Christopher Kruegel, *Exposure: A Passive DNS Analysis Service to Detect and Report Malicious Domains*, ACM Trans. Inf. Syst. Secur. **16** (2014), no. 4, 14:1–14:28.

- Carlos A. Catania and Carlos GarcíA Garino, *Automatic Network Intrusion Detection: Current Techniques and Open Issues*, Comput. Electr. Eng. **38** (2012), no. 5, 1062–1072.

- Greg Farnham, *Detecting DNS Tunneling*, Tech. report, SANS Institute, February 2013.

References II

- Nicole M. Hands, Baijian Yang, and Raymond A. Hansen, *A Study on Botnets Utilizing DNS*, Proceedings of the 4th Annual ACM Conference on Research in Information Technology (New York, NY, USA), RIIT '15, ACM, 2015, pp. 23–28.

- C. Livadas, R. Walsh, D. Lapsley, and W. T. Strayer, *Using Machine Learning Techniques to Identify Botnet Traffic*, Proceedings. 2006 31st IEEE Conference on Local Computer Networks, November 2006, pp. 967–974.

- Pedro Marques da Luz, *Botnet detection using passive DNS*, Master Thesis, Radboud University Nijmegen, Nijmegen, NL, 2014.

- Muhammad Mahmoud, Manjinder Nir, and Ashraf Matrawy, *A Survey on Botnet Architectures, Detection and Defences*, I. J. Network Security **17** (2015), 264–281.

References III

- Gunter Ollmann, *Botnet Communication Topologies*, White Paper, Damballa, 2009.
- K. I. Santosa, C. Lim, and A. Erwin, *Analysis of educational institution DNS network traffic for insider threats*, 2016 International Conference on Computer, Control, Informatics and its Applications (IC3INA), October 2016, pp. 147–152.
- Matija Stevanovic and Jens Myrup Pedersen, *On the Use of Machine Learning for Identifying Botnet Network Traffic*, Journal of Cyber Security and Mobility **4** (2016), no. 2 & 3.
- Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin, *Intrusion detection by machine learning: A review*, Expert Systems with Applications **36** (2009), no. 10, 11994 – 12000.

References IV

- D. Zhao and I. Traore, *P2p Botnet Detection through Malicious Fast Flux Network Identification*, 2012 Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, November 2012, pp. 170–175.

- G. Zhao, K. Xu, L. Xu, and B. Wu, *Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis*, IEEE Access **3** (2015), 1132–1142.