DGALAB

DGA Clustering and Analysis: Mastering Modern, Evolving Threats

Alexander Chailytko
Team Leader
alexanderc@checkpoint.com

Aliaksandr Trafimchuk
Malware Researcher
aliaksandrt@checkpoint.com

Check Point®
SOFTWARE TECHNOLOGIES LTD.

WHAT IS DGA?

- DGA stands for Domain Generation Algorithm
- Periodically generates large number of domain names, however only part of them will be actually registered by adversary
- A small part of these domains will be used by malware to communicate with malicious C&C servers
- Hard to preemptively block such communications

DEFINING THE PROBLEM

- We don’t know which domains will be generated and registered
- Too many domains need to be blocked every day
- Only a small number of them will actually be contacted by the malware
- Sinkholing is not effective in its current form
- Creating protections against such threats cannot be easily automated

DGALAB

SUPPORTED DGA TYPES

1. Static DGA – generates the same domains every time
2. Date-based DGA – current date is used as input for the algorithm to generate domains
3. Seed-based DGA – utilizes hardcoded seed as input for the algorithm. This seed cannot be easily extracted
4. Combination of the last two

MAIN FEATURES

- Modular system
- Generates the full list of domains
- Combines similar DGAs automatically into one group
- Effective and easy to use whitelisting system
- Resource saving emulation

MAIN FEATURES

- A lot of statistics are gathered and used for intelligence needs
- Domain level malware detection and categorization
- Pre-generation of domains for the future

UNDER THE HOOD

- Highly modified Cuckoo Sandbox and CuckooMon
- VMWare Workstation as hypervisor
- A lot of fixes to the virtual environment itself (NetBIOS, TCP/IP stack, WinSock)
- Own kernel mode driver for increased successful emulation rate

CATEGORIZATION

Preassigned family

Sample #1

Sample #2

Sample #3

Sample #4

Emulate

Hypervisor

VMs

Categorize

Category #1

Sample #1

Sample #3

Category #2

Sample #2

Sample #4

CATEGORIZATION

Preassigned Real family [Symmi]

Category #1

Sample #1

Sample #3

Category #2

Sample #2

Sample #4

Real family [Tinba]

OVERALL STATISTICS

- More than 28 families supported
- 100000+ samples processed
- 500+ unique categories generated
- 725000+ unique malicious domains collected

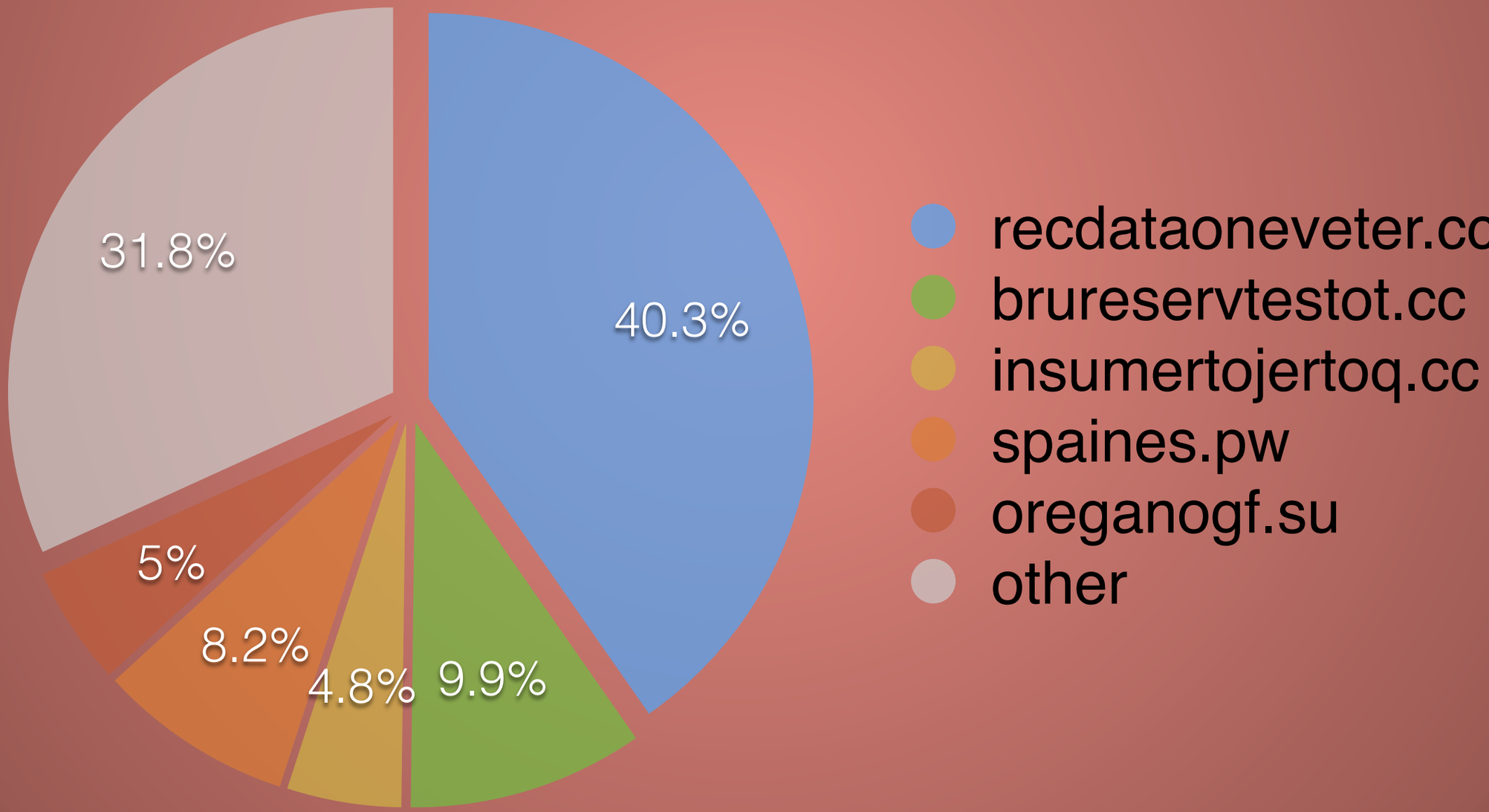
CASE STUDY: TINBA

Family statistics

- Total samples: 68356
- Number of categories: 120
- Unique domains collected: 101311

CASE STUDY: TINBA

Statistics by category



SUMMARY

- We gather domains even from samples that fail to execute on other popular platforms
- DGALab is able to automatically feed gathered data to threat intelligence databases
- Almost zero false positive rate
- Lightning fast categorization even of the unknown malware using the generated domains
- Emulation of one sample takes approximately 3 minutes

DEMO TIME!