Multidimensional Investigation of Source Port 0 Probing

By

Elias Bou-Harb, Nour-Eddine Lakhdari, Hamad Binsalleeh and Mourad Debbabi

Presented At

The Digital Forensic Research Conference

DFRWS 2014 USA Denver, CO (Aug 3rd - 6th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

UNIVERSITÉ Concordia UNIVERSITY

NCFTA National Cyber-Forensics and Training Alliance CANADA

MULTIDIMENSIONAL INVESTIGATION OF SOURCE PORT 0 PROBING

Elias Bou-Harb, Nour-Eddine Lakhdari, Hamad Binsalleeh and Mourad Debbabi

Motivation

Deep cyberattacks cost U.S. banks millions

Google Warns Thousands Of Users About Potential State-Sponsored Cyber Attacks

Syria, aided by Iran, could strike back at US in cyberspace

Cyber attacks hit Twitter and New York Times

Chinese authorities say massive DDoS attack took down .cn domain

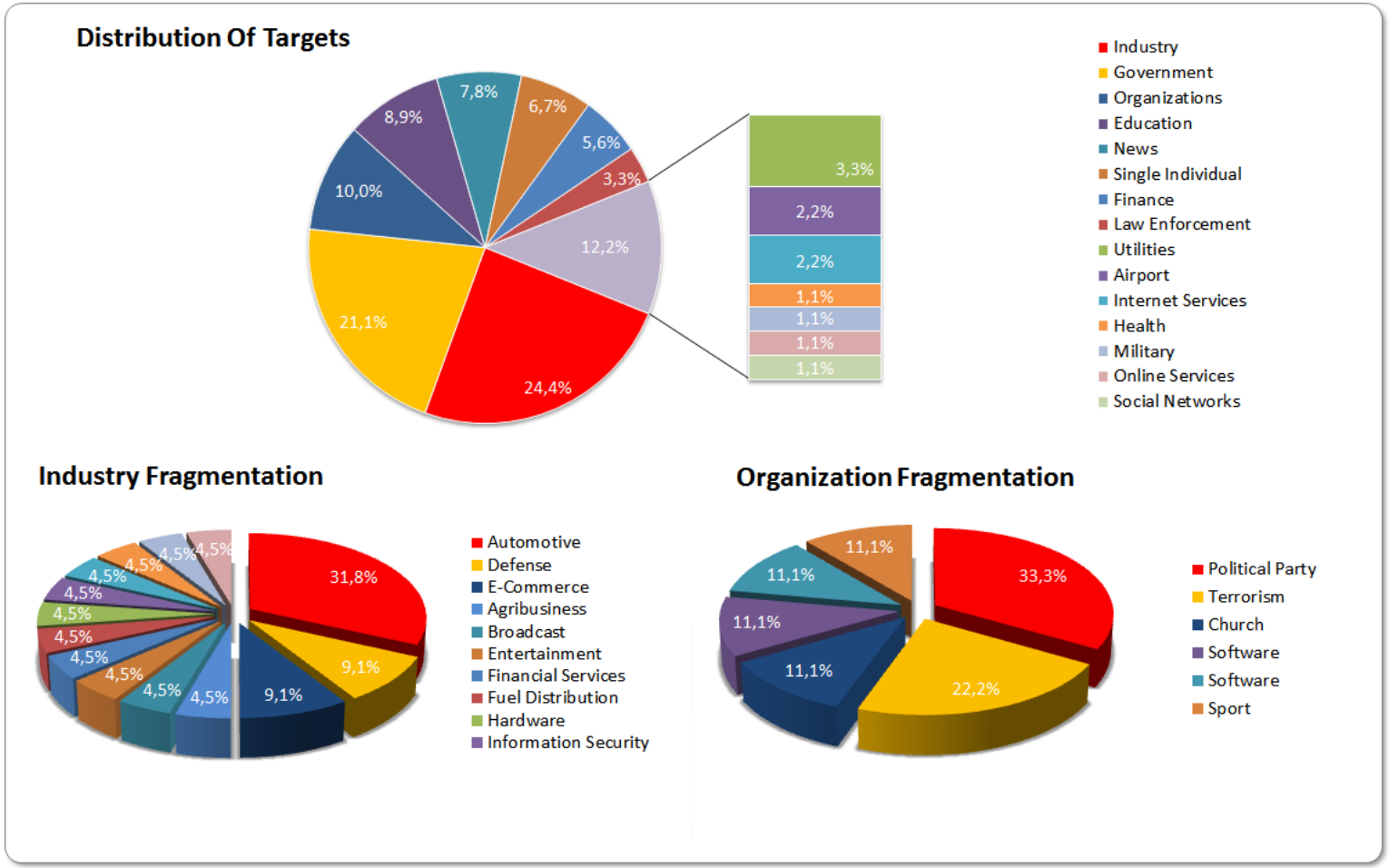
Senate website, CIA.gov reportedly hacked. LulzSec claims responsibility.

Facebook computers compromised by zero-day Java exploit

Hack Attack: Sony Confirms PlayStation Network Outage Caused By 'External Intrusion'

NCFTA
National Cyber-Forensics and Training Alliance CANADA

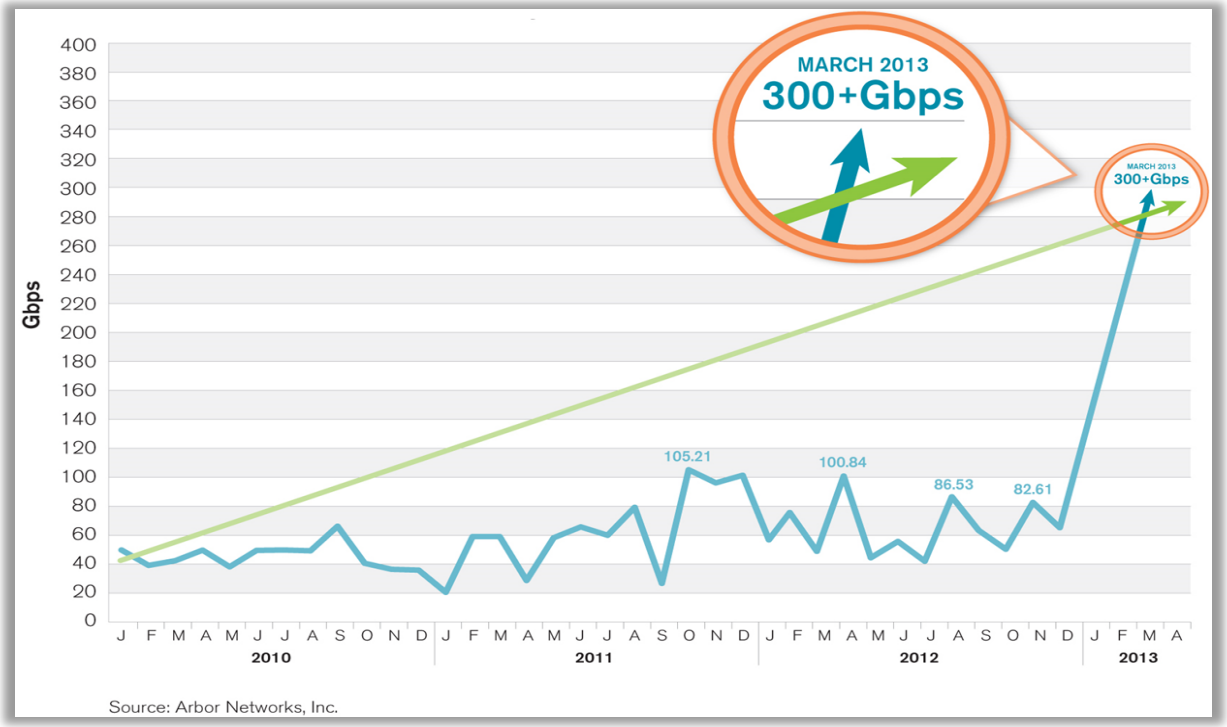
Motivation



NCFTA
National Cyber-Forensics and Training Alliance CANADA

Motivation

- Attacks on cyber physical systems rose 52% in 2013
- The surge of severe DDoS attacks



NCFTA
National Cyber-Forensics and Training Alliance CANADA

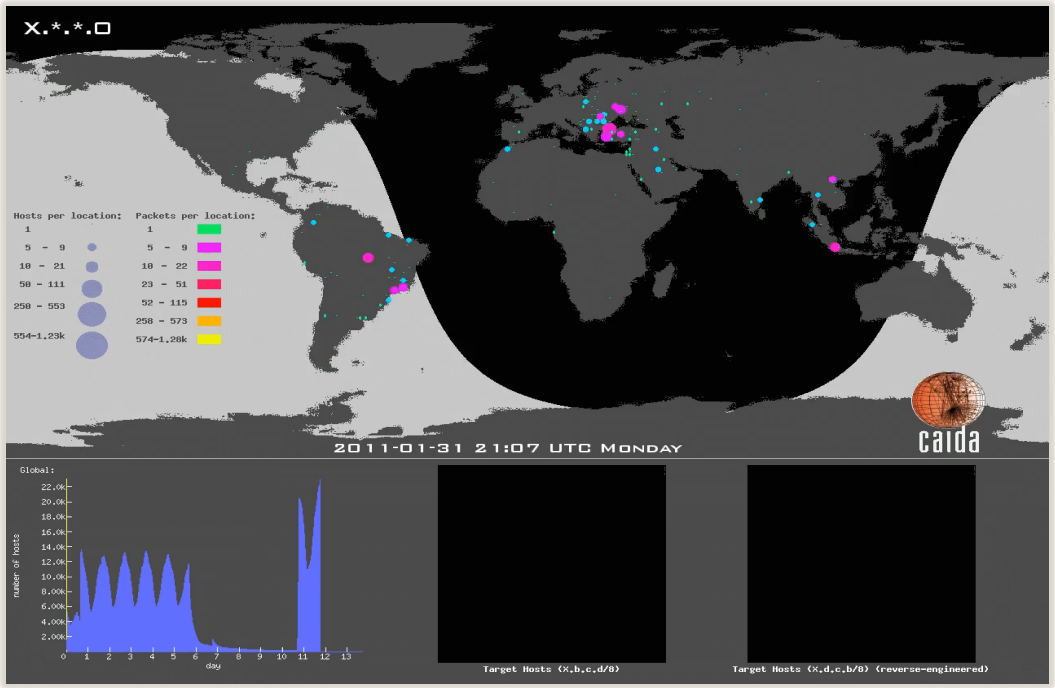
Motivation

Daily tens of thousands of threats (NCFTA Canada 2014)

NCFTA
National Cyber-Forensics and Training Alliance CANADA

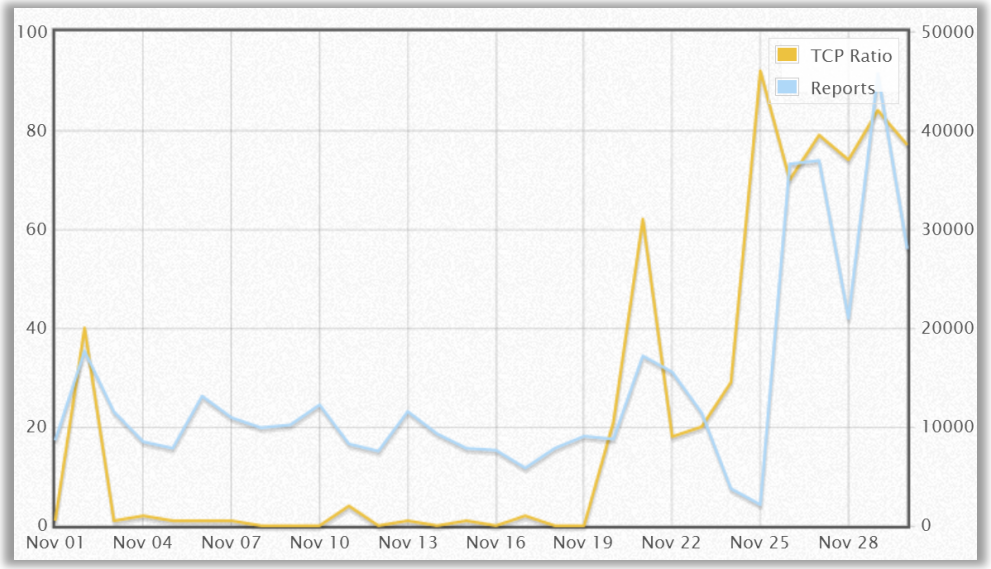
Motivation

- Probing:
  - Scan organizational networks or Internet wide services
  - Primary stage of an intrusion attempt
  - Facilitates more than 50% of cyber attacks
- Large-scale orchestrated probing campaigns



The event

- Traffic originating from TCP source port 0
- Investigated on November 24th and 25th, 2014



Inferred & Validated by Internet Storm Center (ISC)

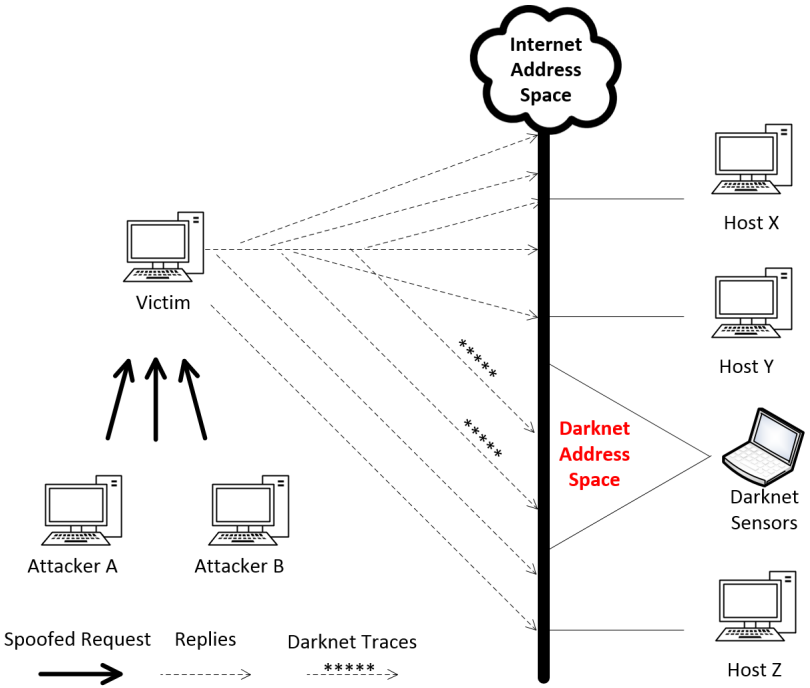
Objectives

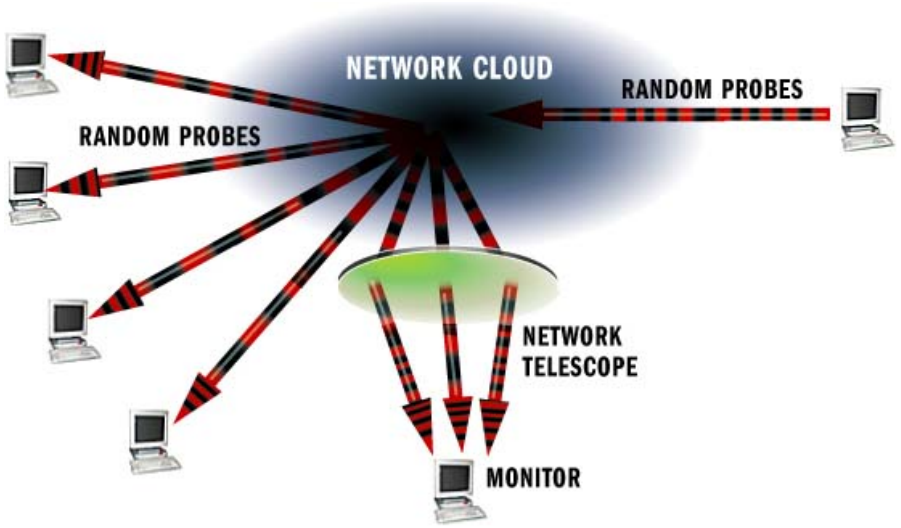
- Investigate the nature of such unprecedented increase of traffic originating from source port 0
- Infer the maliciousness of the sources
- Attribute the event to a certain malware infection

Approach

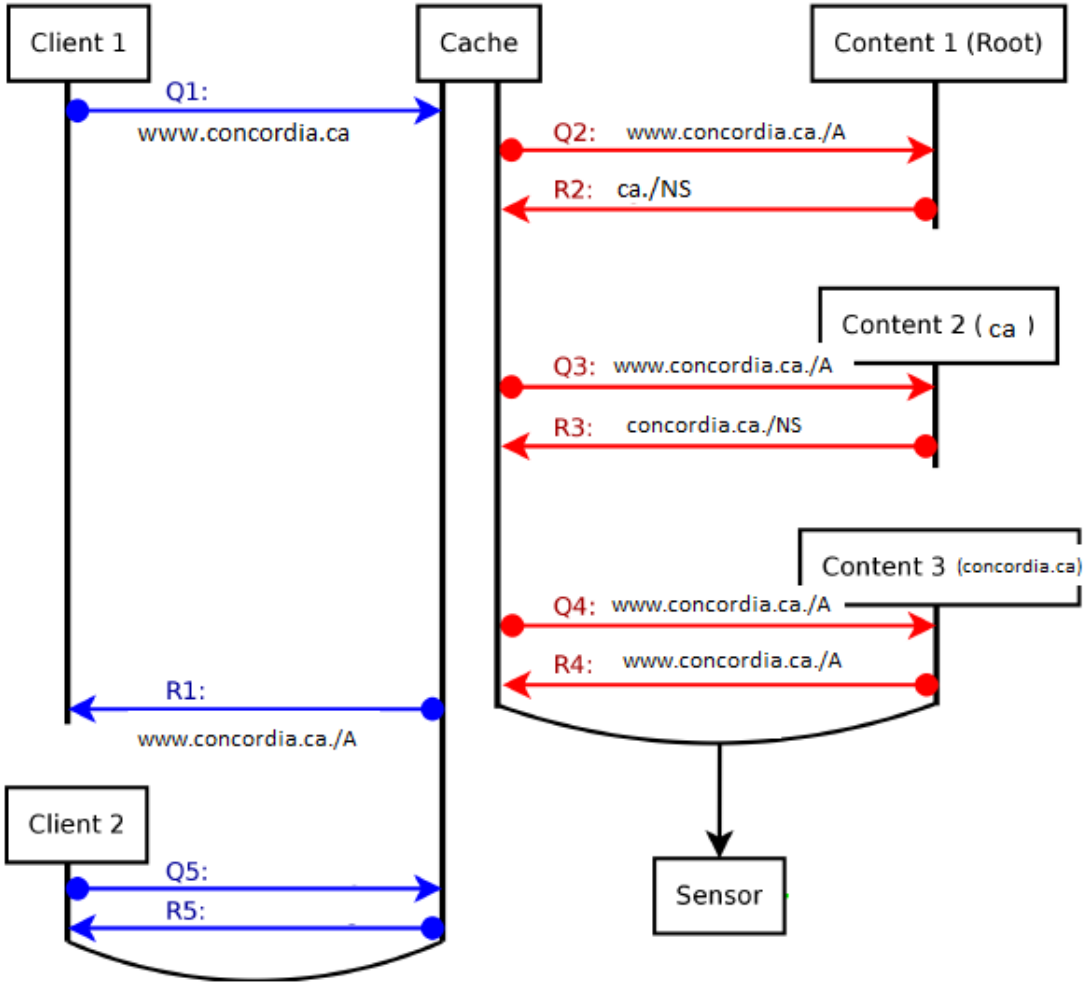

Darknet Analysis
Passive DNS Correlation
Malware Data Correlation

Darknet





Passive DNS

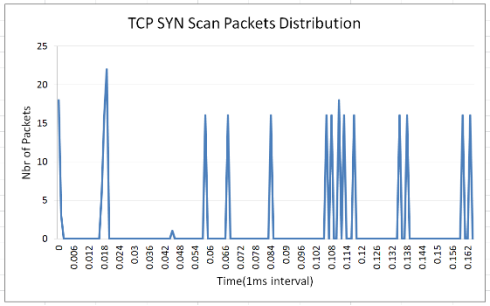


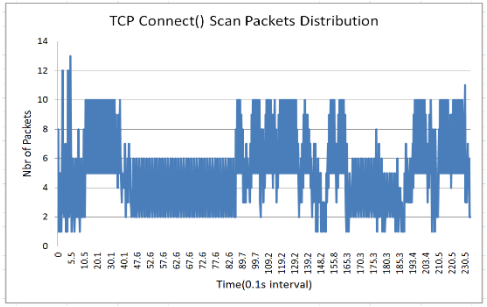
Approach

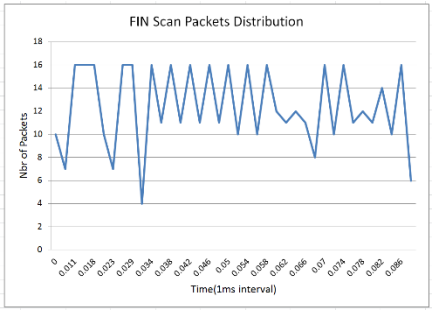
- ❑ **Traffic Extraction** by filtering traffic that employs TCP source port 0
- ❑ **Misconfiguration Filtering** to exclude darknet noise
- ❑ **Traffic Fingerprinting** to distinguish between probing and backscattered traffic
- ❑ **Traffic Clustering** to reveal sources possessing similar traffic patterns
- ❑ **Behavioral Analytics** to infer the strategies, natures and mechanisms of the sources

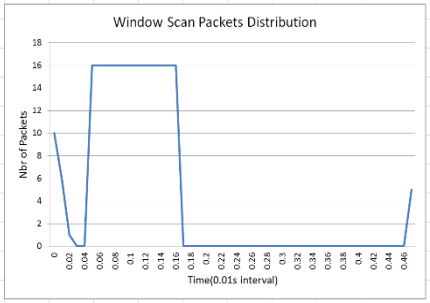
Traffic Fingerprinting Technique

- Observation:
  - Probing techniques (TCP SYN, UDP, ACK, etc.) demonstrate a similar temporal correlation and similarity when generating their corresponding probing traffic







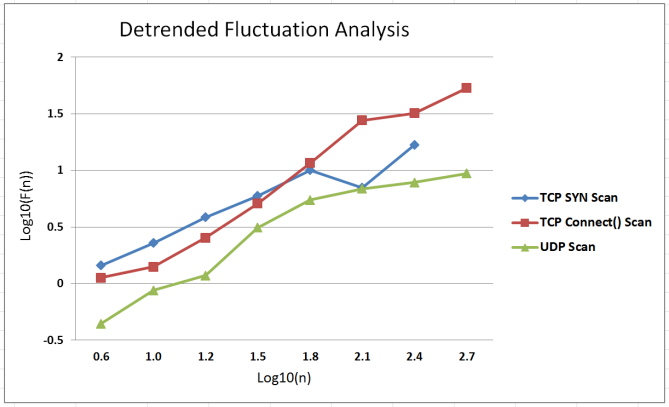


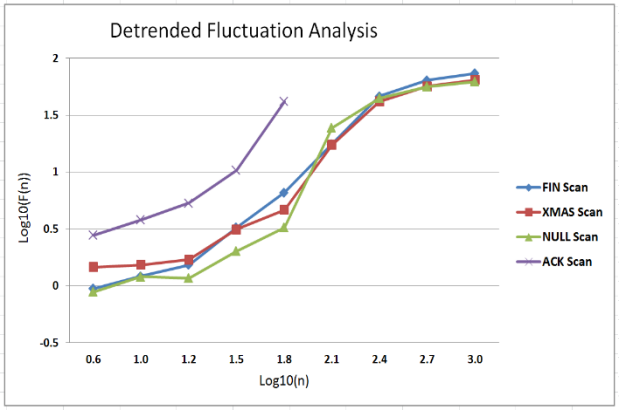
Traffic Fingerprinting Technique

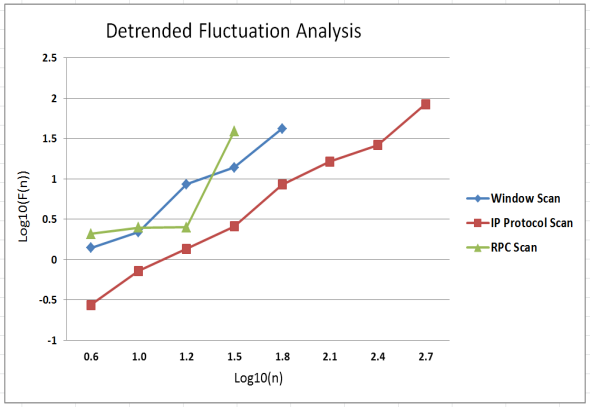
- Detrended Fluctuation Analysis (DFA)
  - To determine the statistical self-affinity of a signal
  - Excessively used in medicine, economy and geology
  - 2 works related to cyber security, none related to probing

- The fluctuations are characterized by a scaling exponent α:
  - $\alpha < 0.5$: anti-correlated
  - $\alpha \approx 0.5$: uncorrelated or white noise
  - $\alpha > 0.5$: correlated
  - $\alpha \approx 1$: 1/f-noise or pink noise
  - $\alpha > 1$: non-stationary, random walk like, unbounded
  - $\alpha \approx 1.5$: Brownian noise

Traffic Fingerprinting Technique



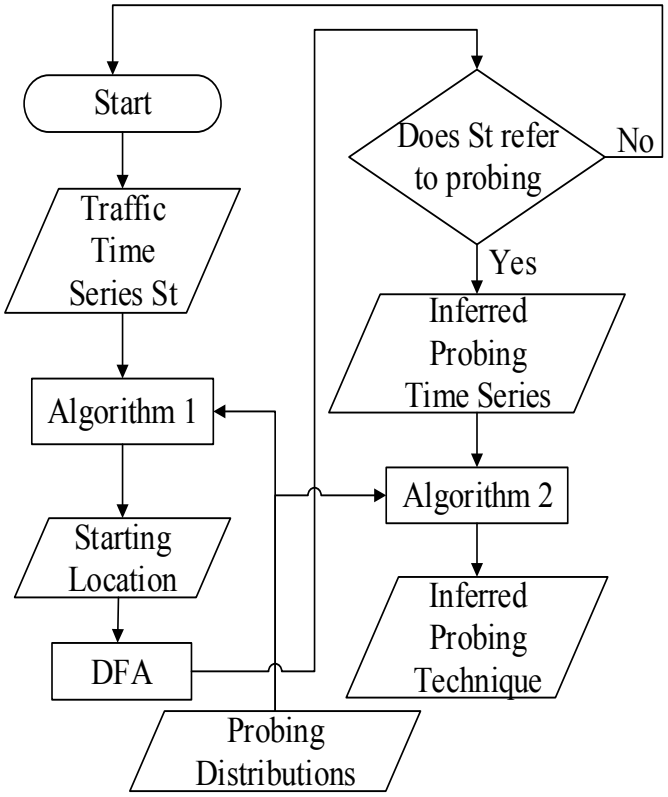




Probing Technique	Exp1: α	Exp2: α
TCP SYN	0.57	0.74
TCP Connect()	0.87	0.69
FIN	0.31	0.24
Xmas	0.30	0.27
Null	0.37	0.41
UDP	0.66	0.58
IP Protocol	1.13	1.22
ACK	0.44	0.29
Window	1.24	1.18
RPC	1.31	1.29
ICMP Echo	0.25	0.29

Correlation Status	Probing Techniques
Anti-Correlated	FIN Probing Xmas Probing Null Probing ACK Probing ICMP Echo Probing
Correlated	TCP SYN Probing TCP Connect() Probing UDP Probing
Non-Stationary	IP Protocol Probing Window Probing RPC Probing

Traffic Fingerprinting Technique

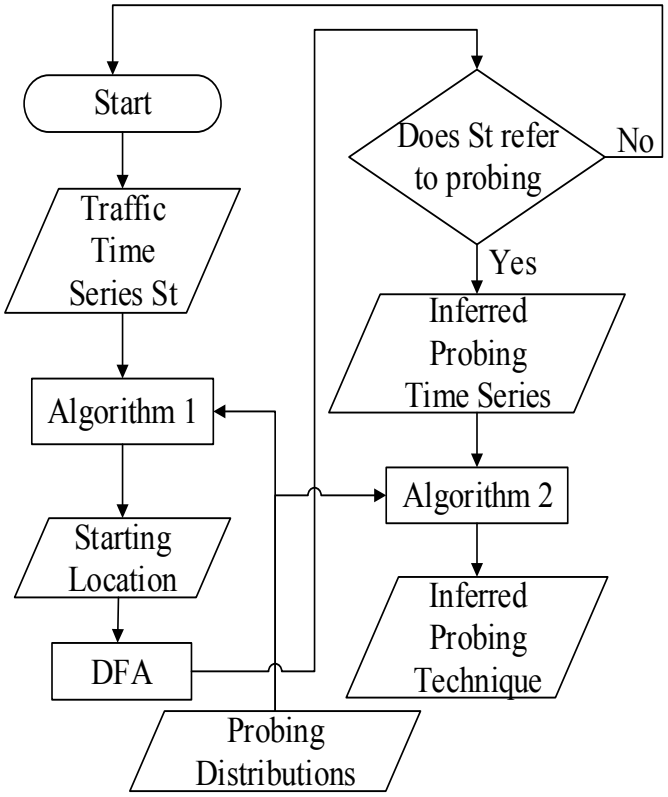


input : A time series S_t of the distribution under testing; the set of time series Sc_p of the distributions of the scanning techniques.
output: X, reflecting the starting location on where to apply DFA in S_t

m=length(S_t);
for every Sc_p **do**
n=length(Sc_p);
for i=$1 \rightarrow (m-n)$ **do**
  s[i]=compare[$S_t(1+i, \cdots, n+i)$, $Sc_p(1, \cdots, n)$];
end
S[p]=min(s[]);
end
X=min(S[]);
return (X);

compare(A, B)
for i=$1 \rightarrow n$ **do**
  $K[i]$= $\|\mathbf{E}\| = \mathrm{d}(A(i), B(i))$;
  sum+=$K[i]$;
  return (sum);
end

Traffic Fingerprinting Technique



input : A time series S_b of the probing distribution that DFA was previously applied on; a cluster of time series Sc_b of the distributions of the scanning techniques related to the correlation status.

output: Sc_{bi}, reflecting one scanning technique that is estimated to be generating the probing activity found in S_b.

for every Sc_{bi} **do**

$Bha_{bi} = \|\mathbf{Bha}\| = \mathrm{d}(\mathbf{Sc_{bi}}, \mathbf{S_b})$;

end

$d_i = Min(Bha_{bi})$;

return $(Sc_{bi}|i \text{ of } d_i)$;

Traffic Clustering

Features		
Data Link Features	1	Delta time with previous capture packet
	2	Packet Length
	3	Frame Length
	4	Capture Length
	5	The flag 'frame' is marked
Network Layer Features	6	IP Header length
	7	IP Flags.
	8	IP Flags: reversed bit
	9	IP Flags: do not fragment bit
	10	IP Flags: more fragments bit
	11	IP Fragment offset
	12	IP Time to live
	13	IP Protocol
Transport Layer Features	14	TCP Segment length
	15	TCP Sequence number
	16	TCP Next sequence number
	17	TCP Acknowledgement number
	18	TCP Header length
	19	TCP Flags
	20	TCP Flags: congestion window reduced
	21	TCP Flags: ECN-Echo
	22	TCP Flags: Urgent
	23	TCP Flags: Acknowledgement
	24	TCP Flags: Push
	25	TCP Flags: Reset
	26	TCP Flags: Syn
	27	TCP Flags: Fin
	28	TCP Window size
	29	UDP Length

Behavioral Analytics

Statistical, heuristical, entropy, fuzzy hashing techniques

- Is the probing traffic random or does it follow a certain pattern?
- How are the targets being probed?
- What is the nature of the probing source?
- Is the probing targeted or dispersed?

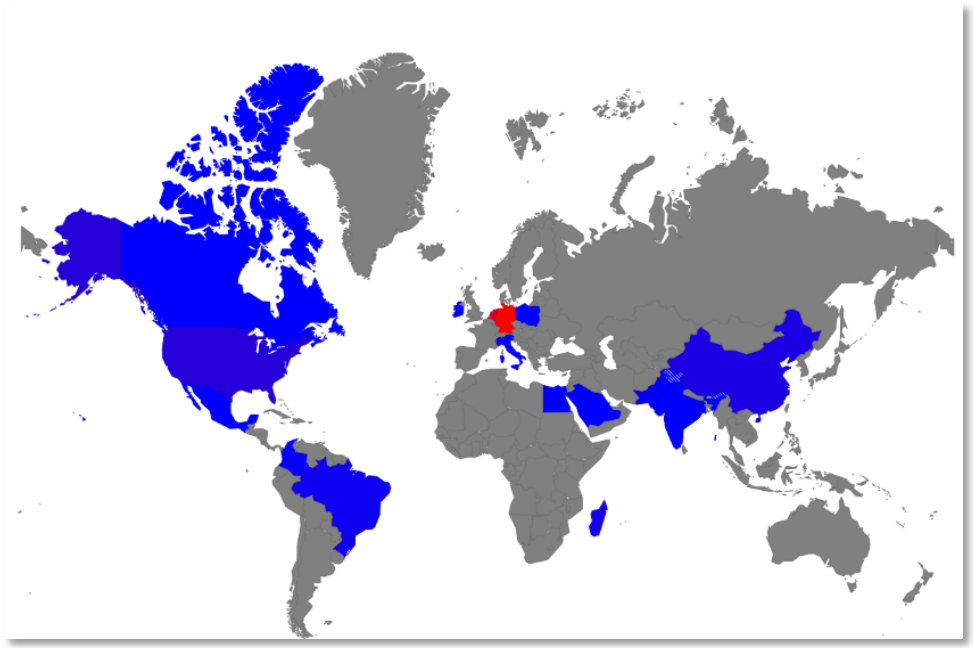
Darknet Inferences

- Peak:
  - Generating more than **1 million packets**
  - Similar traffic from other days is **less than 1000 packets**
  - 97% of traffic refer to **probing activities**

- Overlapping observations by ISC and NCFTA:
  - The TTL of the packets changes with source IP address
  - Packets with a TCP header length of 0 or packets with odd flag combinations (i.e., URG, PSH, Reserved)
  - The packets arrival rate is slow, way far from DDoS levels
  - Xmas Tree-like probing in most TCP flags

Darknet Inferences: Probing Sources

- Originating from
  - 27 unique sources
  - 17 countries
  - 24 ISPs
  - 25 organizations



NCFTA
National Cyber-Forensics and Training Alliance CANADA

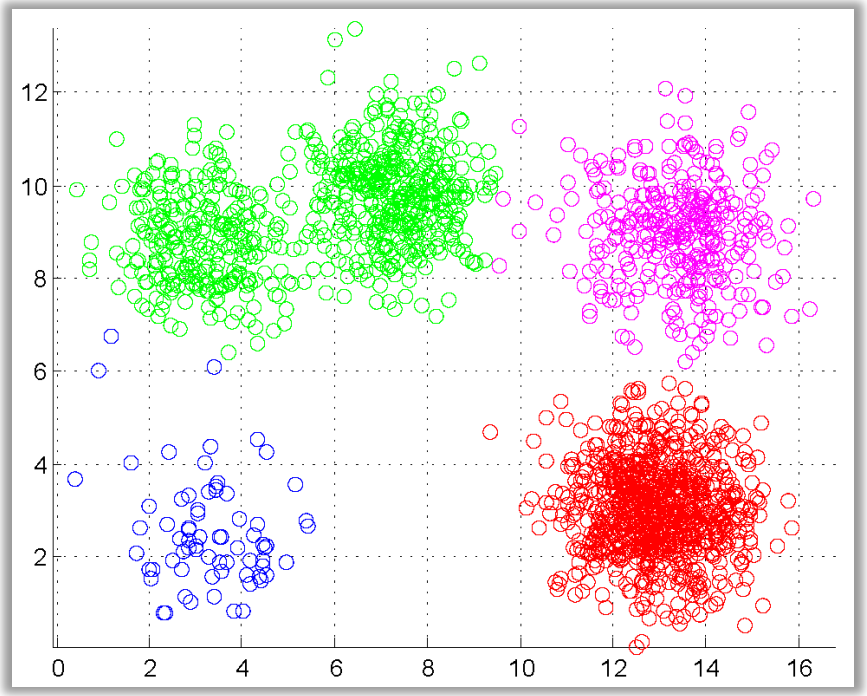
Darknet Inferences: Probing Sources

IP concentration

Darknet Inferences: Targeted Ports

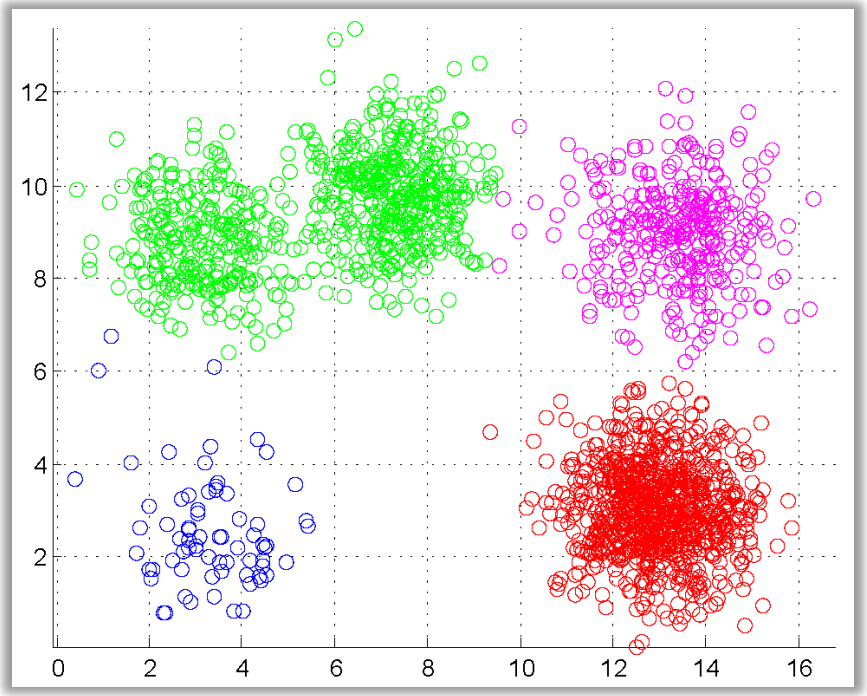
- Port 0 from X.X.X.163, the sole German IP:
  813 thousand packets/destinations
- Port 445: 8952 packets
- Port 22: 7080 packets
- Port 3389: 5606 packets
- > 60,000 ports from X.X.X.201, the sole Dutch IP

Darknet Inferences: Traffic Similarity



- One cluster for the horizontal scan on port 0 from the German IP
- Another cluster for the horizontal scan on > 60 thousand ports from the Dutch IP
- Another cluster for probing on other ports such as 445, 22 and 3389
- Last cluster represents misconfigurations/malformed packets

Darknet Inferences: Behavioral Analytics

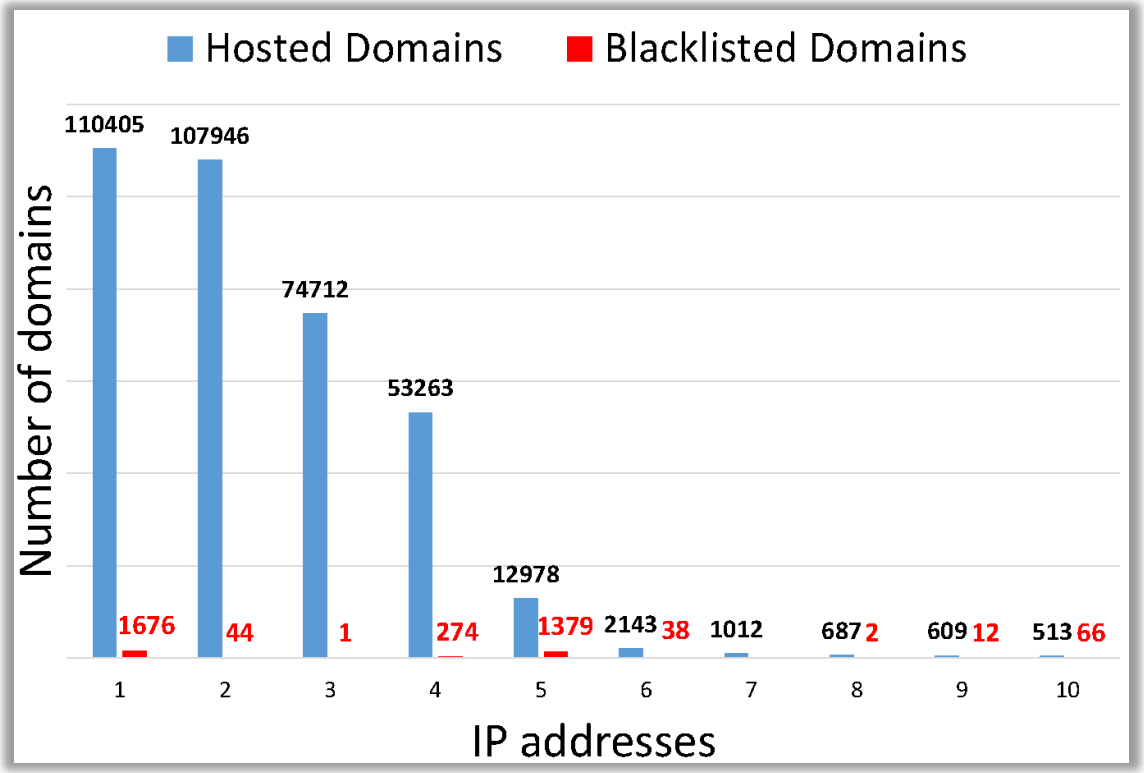


- Probing clusters 1 and 2:
  - random probing traffic as opposed to using patterns
  - sequential strategy
  - probing tool
- Probing cluster 3:
  - pattern usage
  - permutation strategy
  - bots

Approach

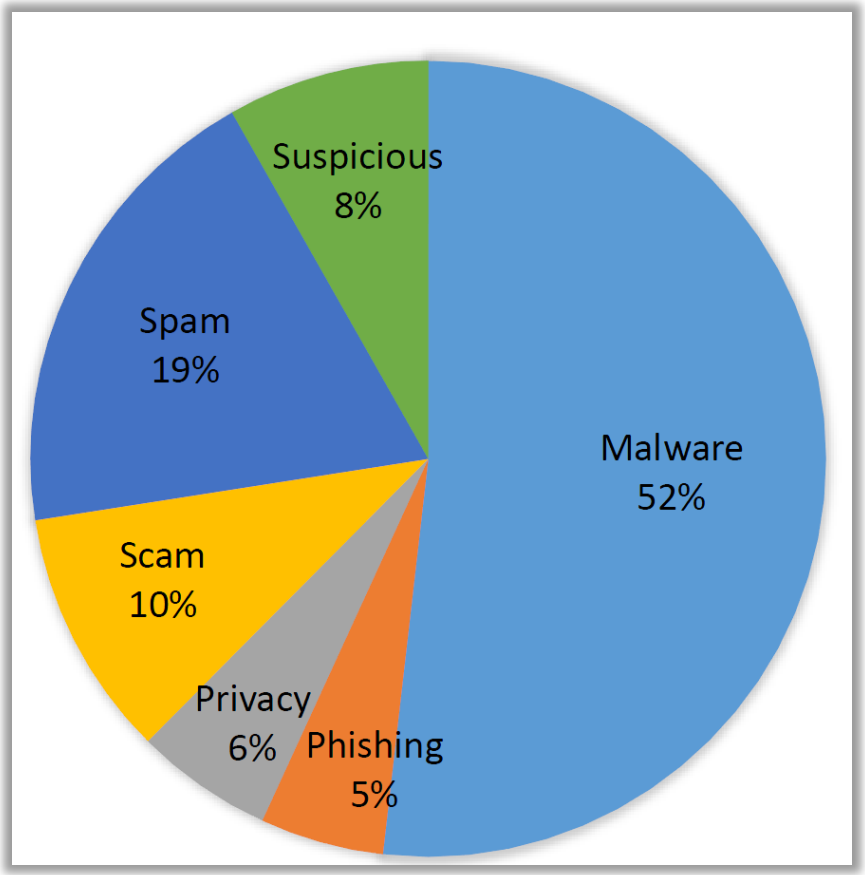
- **Hosting capability** to infer malicious domains affiliated with the sources
- **Intensity** to detect accessibility and involvement
- **Aliveness** to verify the effectiveness of the malicious domains

Passive DNS Inferences



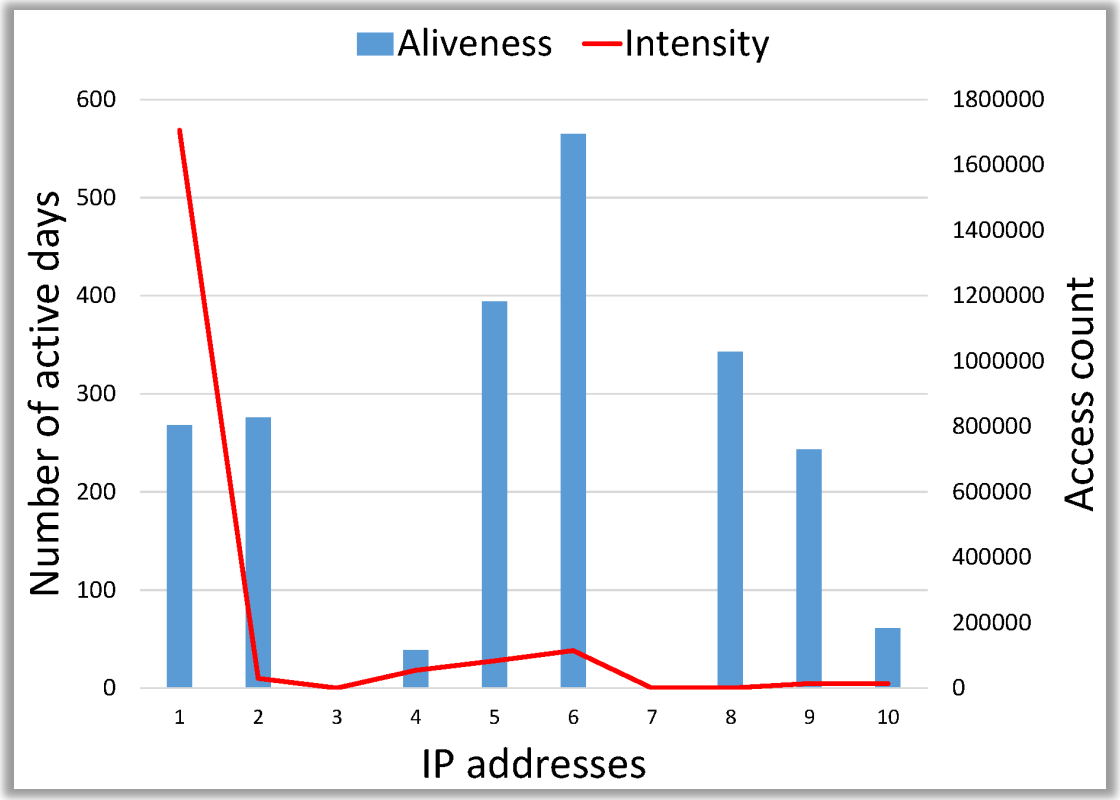
- 513 to 110 thousand hosted domains per the probing IPs
- 28% of the probing IPs are hosting blacklisted/malicious domains

Passive DNS Inferences



Half of the malicious domains could be attributed to malware activities

Passive DNS Inferences



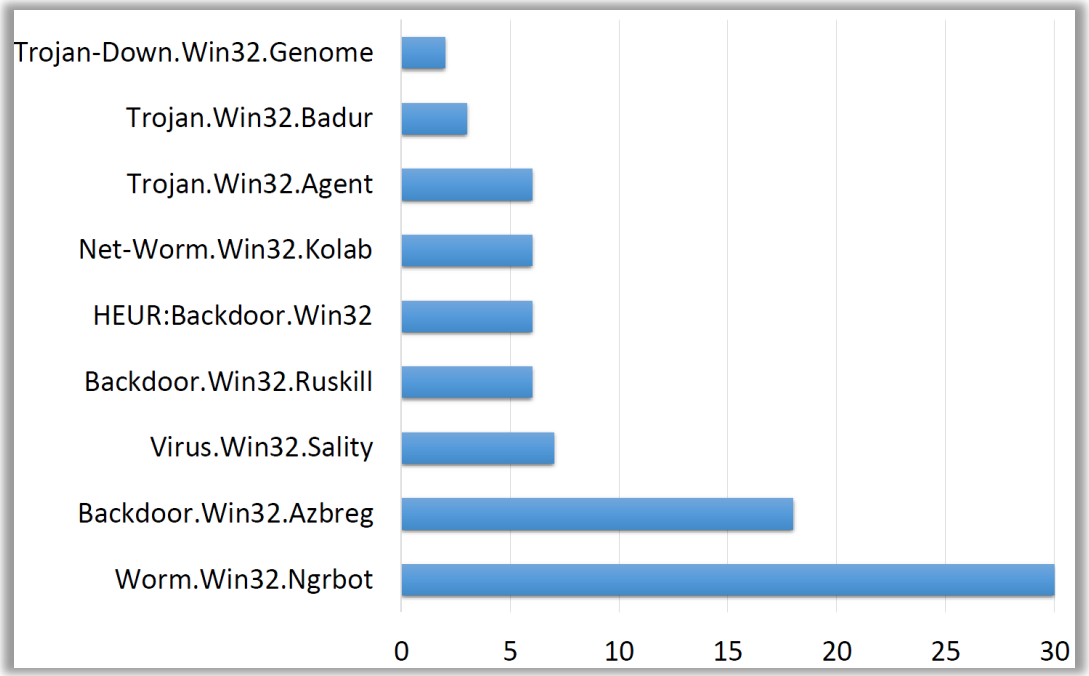
- Some malicious domains are very effective; high access count and low active days
- Domains with prolonged active days provide back-end services to others malicious activities

NCFTA
National Cyber-Forensics and Training Alliance CANADA

Approach

- **Machine infection** to deduce malware samples that infect the sources
- **Malware Attribution** to infer the malware samples that generate such traffic

Malware Inferences



Email-Worm.Win32.Mydoom
Worm.Win32.AutoRun
Virus.Win32.Sality
Virus.Win32.Expiro
Backdoor.Win32.Xtoober
Trojan-Downloader.Win32.Agent
Trojan-Dropper.Win32.Small
Trojan.Win32.Pincav
Trojan.Win32.Jorik
Trojan-Downloader.Win32.Delf
Trojan-Downloader.Win32.Genome
Backdoor.Win32.Gbot
Backdoor.Win32.Popwin
Email-Worm.Win32.Rays
Email-Worm.Win32.Runouce
Packed.JS.Agent
Trojan-Banker.Win32.Banker
Trojan-Downloader.Win32.FlyStudio
Backdoor.Win32.Banito
Backdoor.Win32.VB
HackTool.Win32.Injecter

Virus.Win32.Sality is the common factor

Summary

- Traffic is indeed reconnaissance/probing activities originating from three different horizontal scans
- 28% of the scanning sources host malicious/blacklisted domains as they are often used for spamming, phishing and other fraud activities
- Bot probing sources (i.e., probing cluster 3) are infected by Virus.Win32.Sality

Outcome

- Devised a generic approach that could be applied to analyze other cyber events with similar nature
- Demonstrated the value added of employing
  - Databases of cyber security data
    (darknet, passive dns and malware)
  - Tools and APIs that can effectively utilize the data
- Permitted prompt data analysis to support investigation of cyber events

Acknowledgement

- Research members of the NCFTA Canada lab headed by Prof. Mourad Debbabi

- Natural Sciences and Engineering Research Council of Canada (NSERC)

NCFTA
National Cyber-Forensics and Training Alliance CANADA

Questions

Elias Bou-Harb
Ph.D. Candidate
Cyber Security R&D

e_bouh@encs.concordia.ca

NCFTA
National Cyber-Forensics and Training Alliance CANADA