What’s Lurking in New Core Domains?

Yuriy Yuzifovich, Nominum Data Science and Security Research

What My Team Does

Process real-time DNS data streamed from ISPs around the world

Create highly validated, continuously updated, threat feeds to:

- Protect DNS servers – “purpose built” amplification domains, randomized subdomains

- Protect networks – bot activity

- Protect subscribers – phishing, malware, bot C&C, adware, browser hijacks, etc

We’ve been tracking new core domains for 5+ years, they’re fundamental to all of our work

But it became clear we needed to understand more about them, more quickly

What's A Core Domain?

Sometimes called an "effective" 2nd level domain

Usually captures domain ownership

www.***example.com***

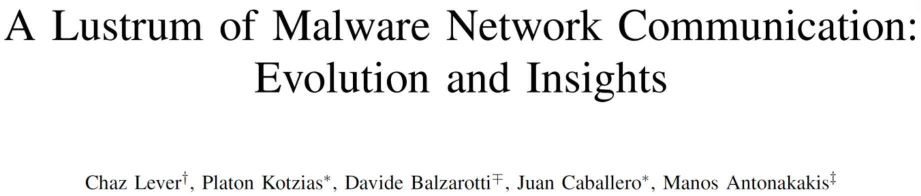
www.***example.com.uk***

Why Look at New Core Domains?

DNS is a facilitator of most malicious activity – it makes threats more agile and stealthy

New core domains are a good demarcation point to detect malicious activity

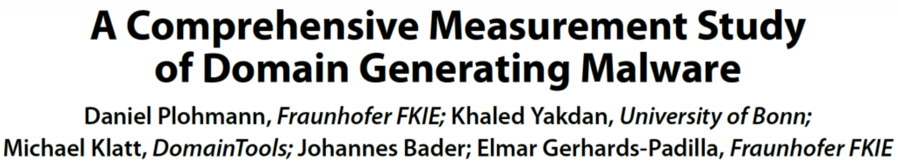
- Often the first signal observable on the Internet



A Lustrum of Malware Network Communication: Evolution and Insights

Chaz Lever[†], Platon Kotzias[*], Davide Balzarotti[∓], Juan Caballero[*], Manos Antonakakis[‡]

"We find that a significant percentage of malware domains can be seen in passive DNS several weeks, in many cases even months, before the actual malware sample was dynamically analyzed by the security community."



A Comprehensive Measurement Study of Domain Generating Malware

Daniel Plohmann, *Fraunhofer FKIE;* Khaled Yakdan, *University of Bonn;*
Michael Klatt, *DomainTools;* Johannes Bader; Elmar Gerhards-Padilla, *Fraunhofer FKIE*

"…DGAs have become very relevant to malware authors, especially over the last 2 years, as 25 out of the 43 considered DGAs surfaced 2013 and later."

The world seems to be waking up to the power of DNS data!

Resolver Data

Live streamed client side *and* recursive requests/responses

- ~1 – 1.5M QPS and growing
- Every incoming request* (PCs, phones/tablets, IoT devices, well configured and not well configured home gateways, etc)
- Recursive requests/authoritative responses
- Worldwide data stream normalizes the diurnal flow

Unique perspective observing both flows

- Richness of authoritative data

*** Source IPs anonymized – no PII**

Objectives

Improve coverage *and* precision

Reduce noise in the data set – throw more power at data of interest to improve coverage

Reduce false positives in resulting threat feeds to improve precision:

> Pre-infection – click through option for users to avoid phishing etc
>
> Post-infection – "silent" blocking of bot C&C and other traffic, no user feedback loop…

Evaluate relevance of queries

Batch Processing

Stream Processing

Our Initial Thought

We can just use Big Data tools…

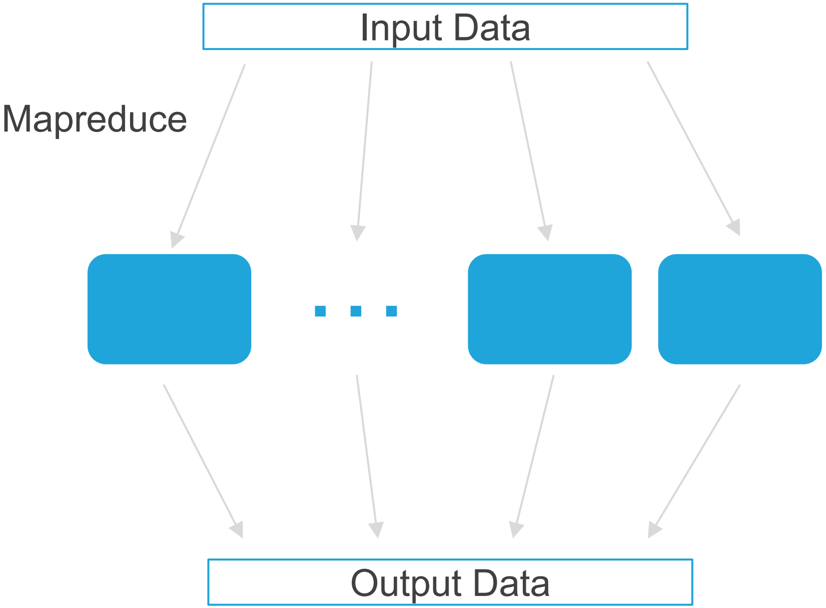
This will be easy!

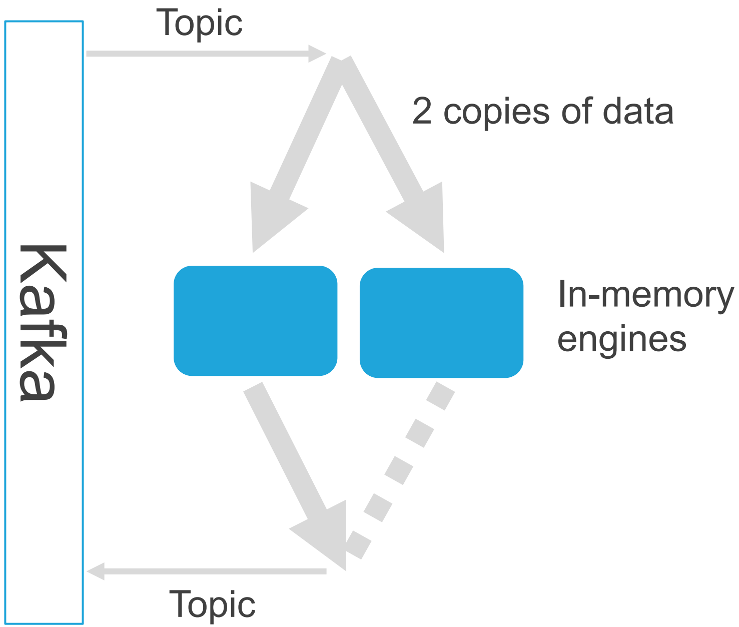
Where We Ended Up

Streamlined

Efficient

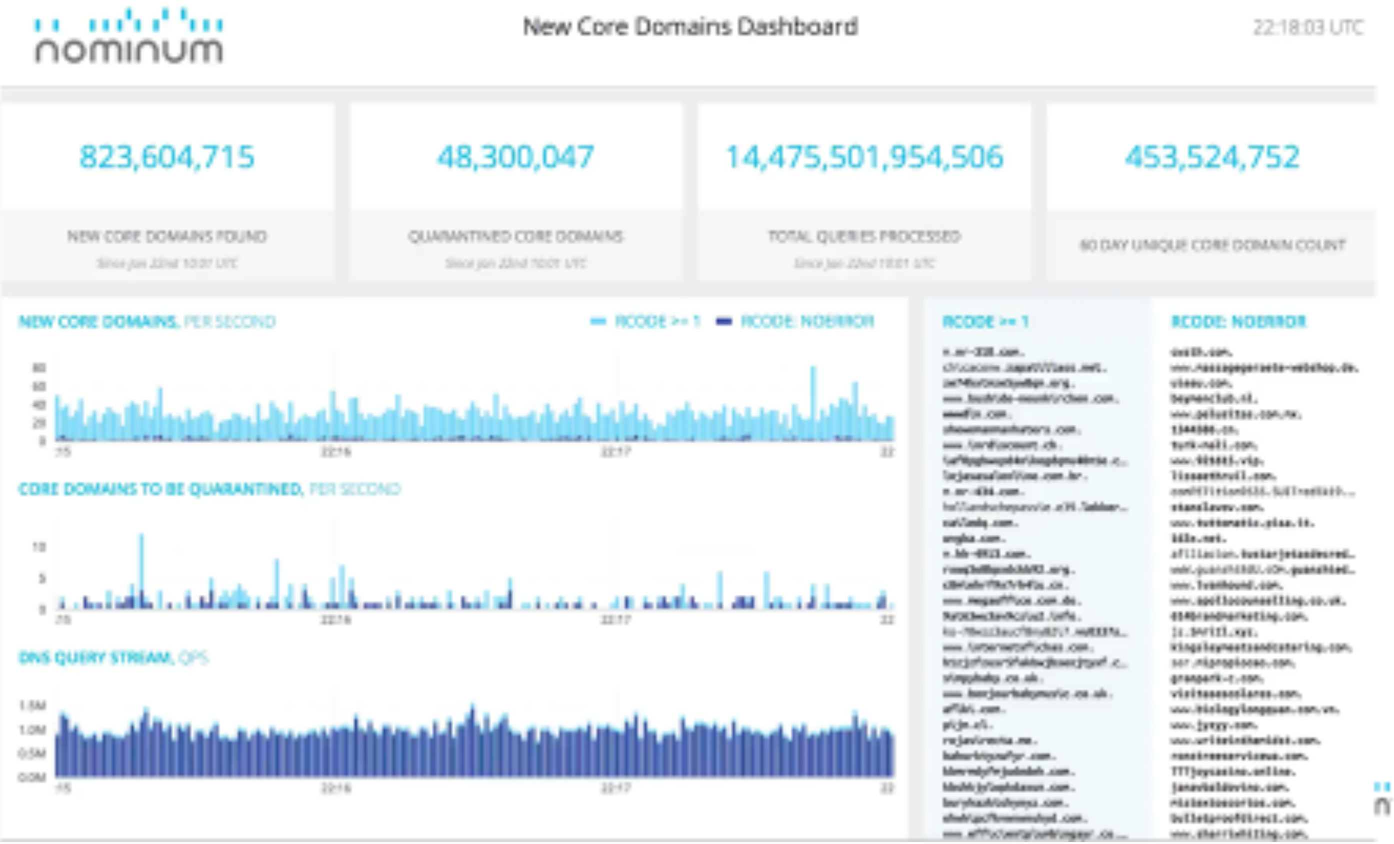
Real Time Stream Processing





High Level View of the Output

This will connect to a live portal



New Core Domains Dashboard

22:18:03 UTC

823,604,715	48,300,047	14,475,501,954,506	453,524,752
NEW CORE DOMAINS FOUND Since Jan 22nd 10:01 UTC	QUARANTINED CORE DOMAINS Since Jan 22nd 10:01 UTC	TOTAL QUERIES PROCESSED Since Jan 22nd 10:01 UTC	60 DAY UNIQUE CORE DOMAIN COUNT

RCODE >= 1

[illegible]

RCODE: NOERROR

[illegible]

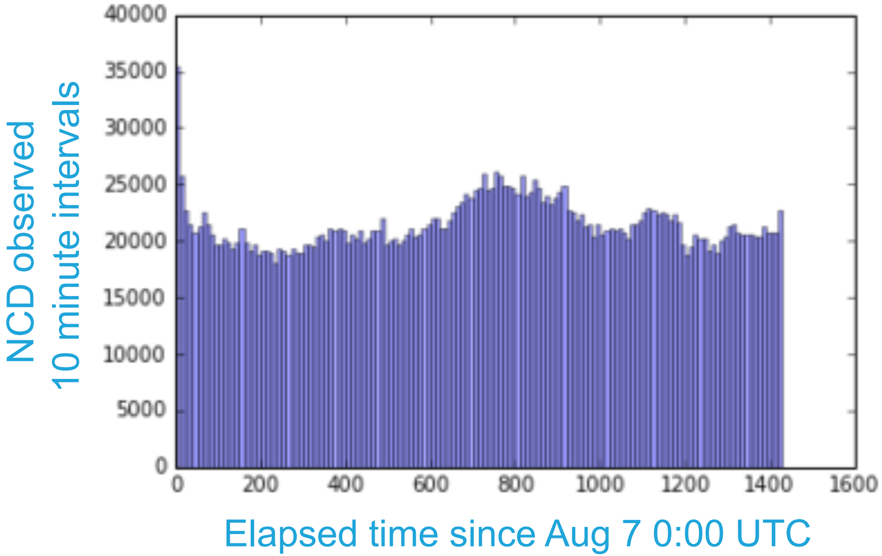
Tracking A Single Day's Data

Aug 7th total New Core Domains: 3,094,508

Observed queries for these domains: Aug 7 - 12

Distribution of RCODE:

- Resolved (Rcode-0): 269,341 (8.7%)
- Non-existent (Rcode-3): 2,798,079
- Other unresolved codes: 280,088



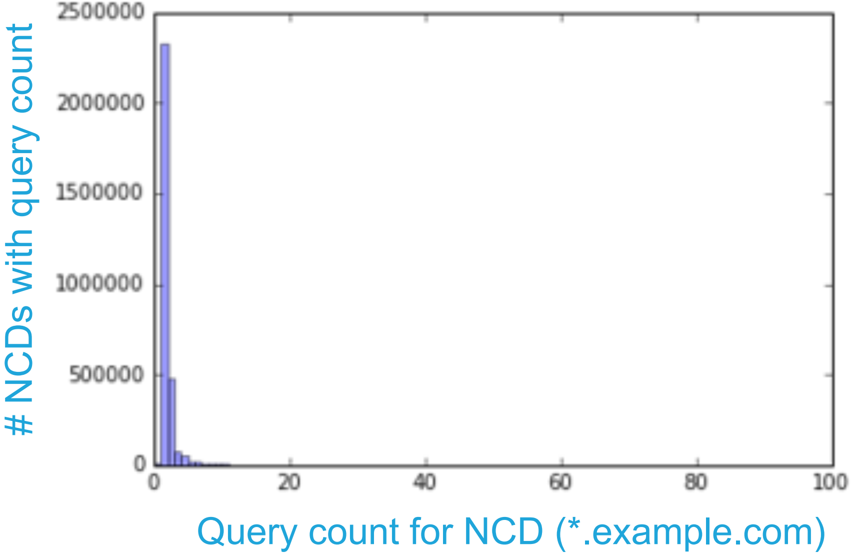
nominum™

Tracking A Single Day’s Data

Core domains with >1,000 queries over 6 days: 3,772

- Resolved: 602 (16.0%)
- Unresolved: 3,170

Total Query Count for domains with <1,000 queries over 6-days: 13,146,899
Total Query Count for domains with >1,000 queries over 6-days : 46,075,459

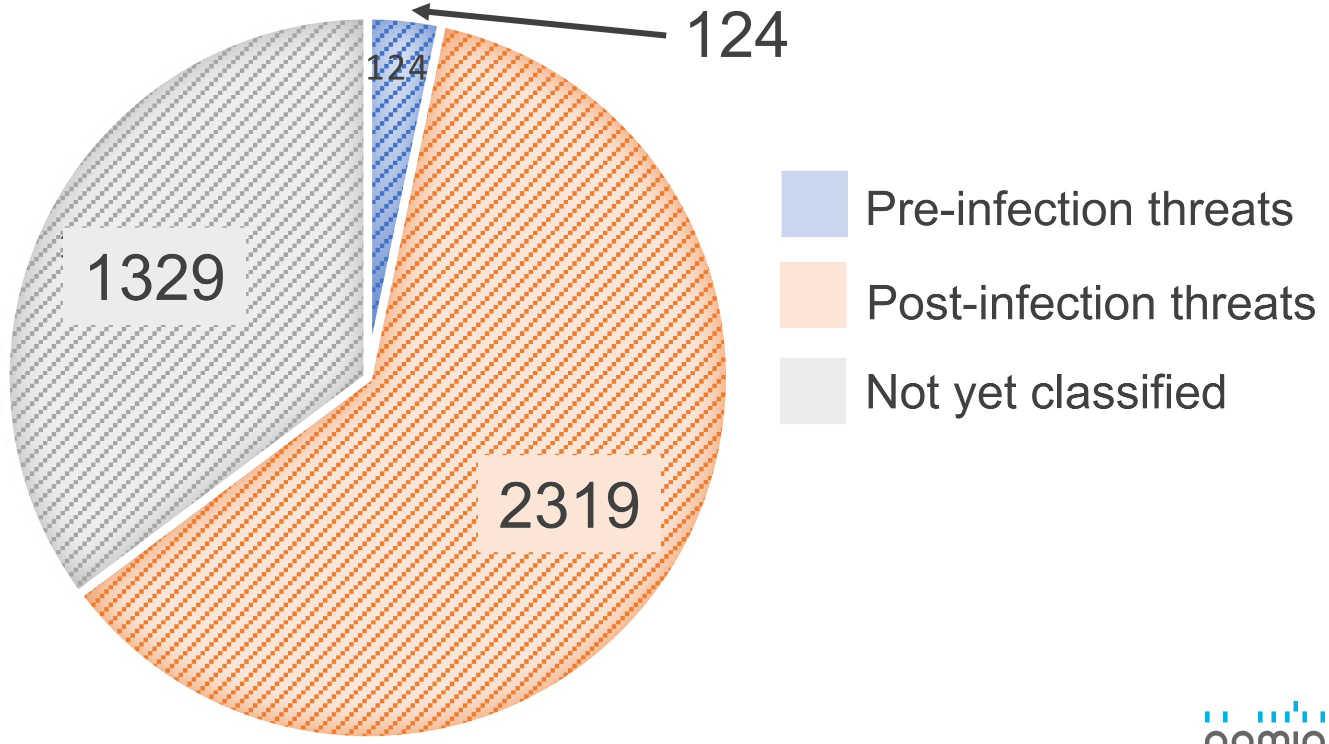


Categorization

- Anomaly detection
- Domain Reputation System
- Behavioral analysis (AKA ”Shapes”)
- Domain2Vec
- Malware DGA reverse engineering

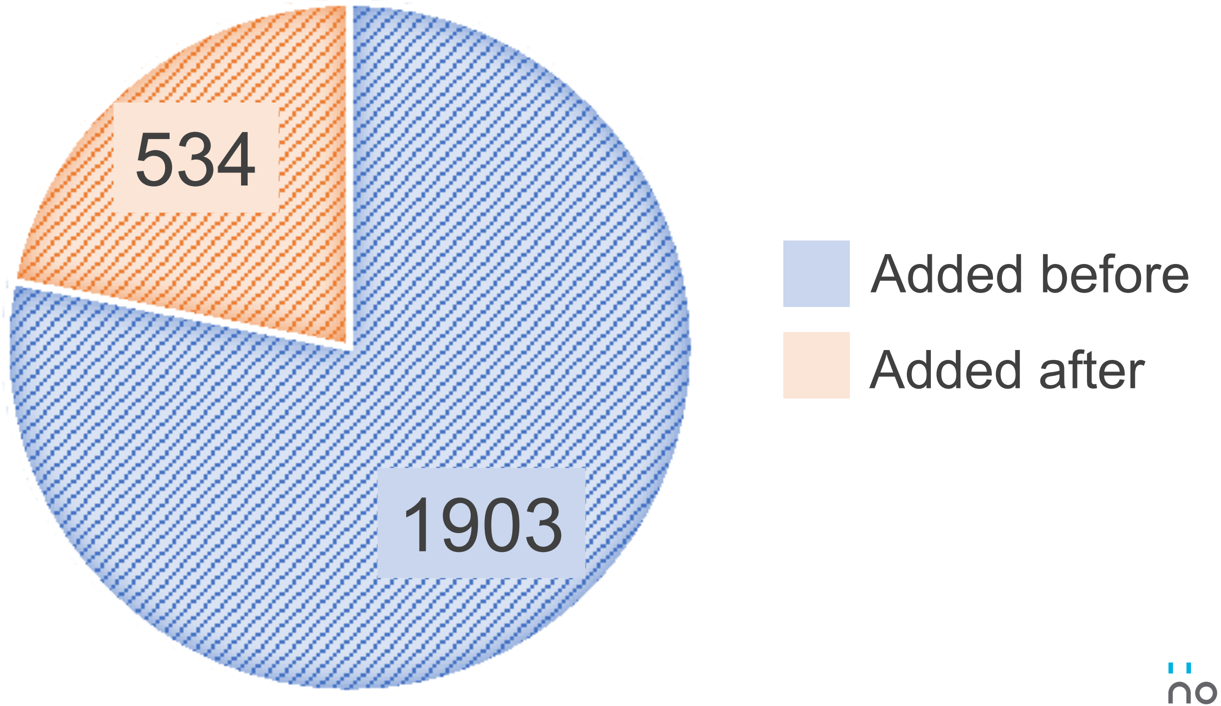
Trends for NCDs with > 1000 Queries

~2/3 could be classified…

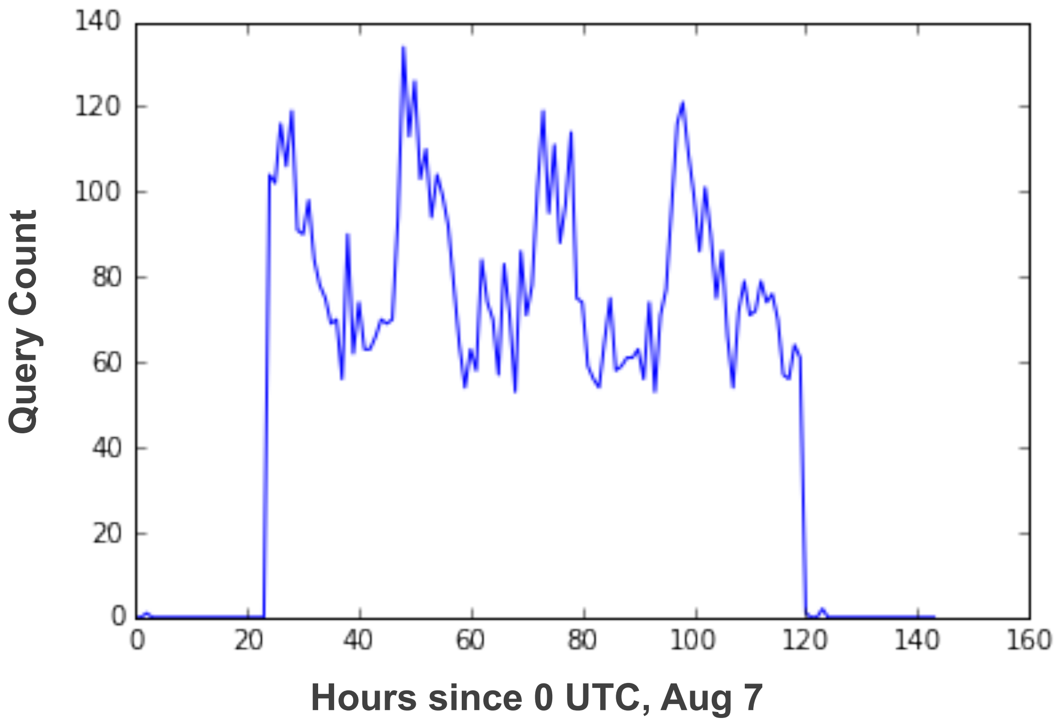


Trends for NCDs with > 1000 Queries

78% classified based on DGA



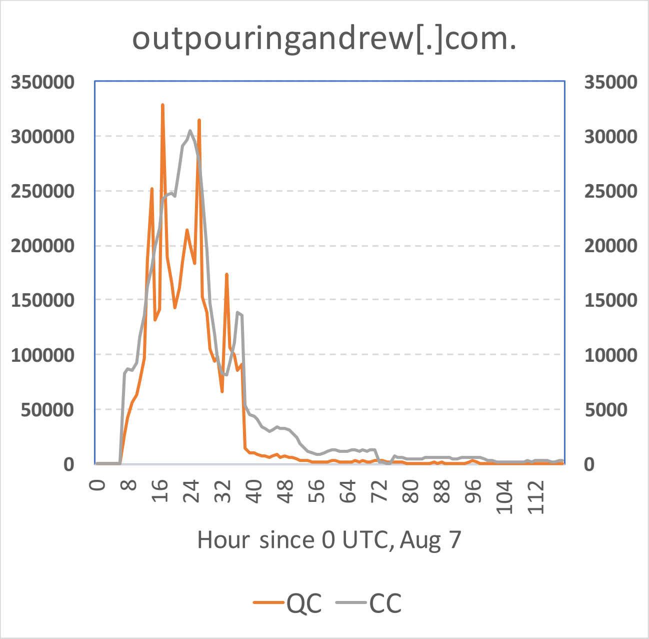
Botnet: 96 hours of Necurs activity

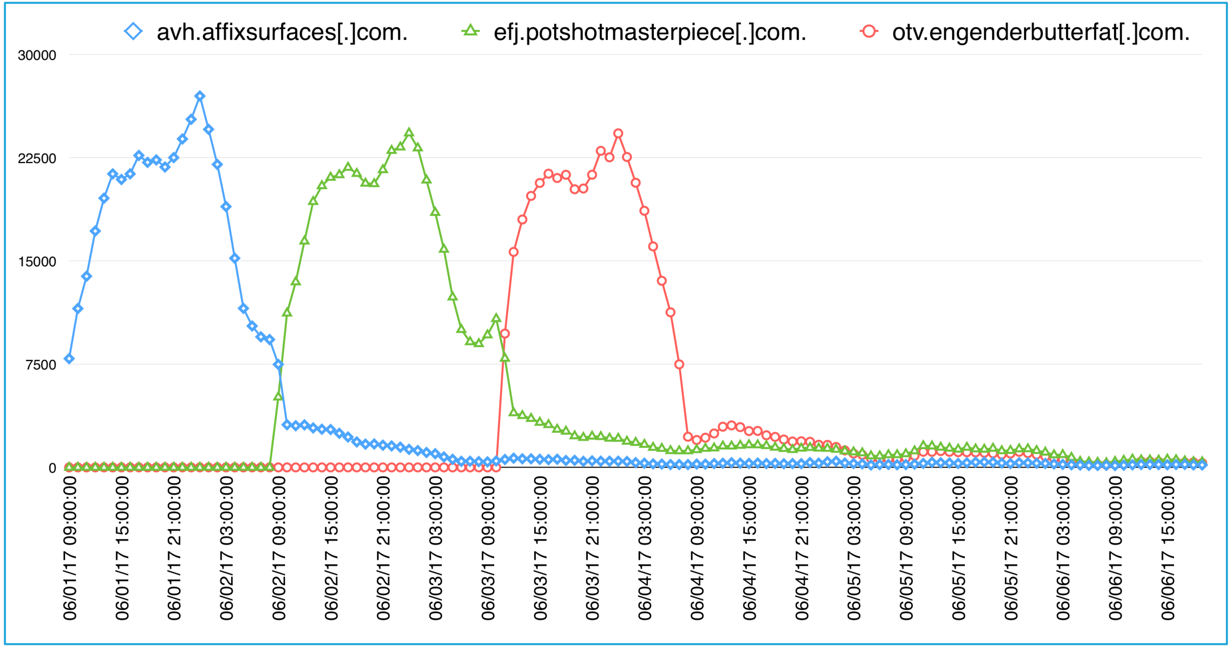


Sample Necurs Domains:

Ifpwfosanoqjy[.]tw.
cssihavkm.pro.
mxmovvf.eu.
fpenuqjeswvabxslj.tj.
hjvxygklnudicrijsj.me.
bmwbkbnilvdhcru.de.
txiutcqkmgdhncywvshxk.xxx.
ulkdvadt.xxx.
blglpststrwlvcudwdkx.nu.
yomaxuactrtsp.biz.
xejnatsypccahv.so.
vairdnhdwgr.com.

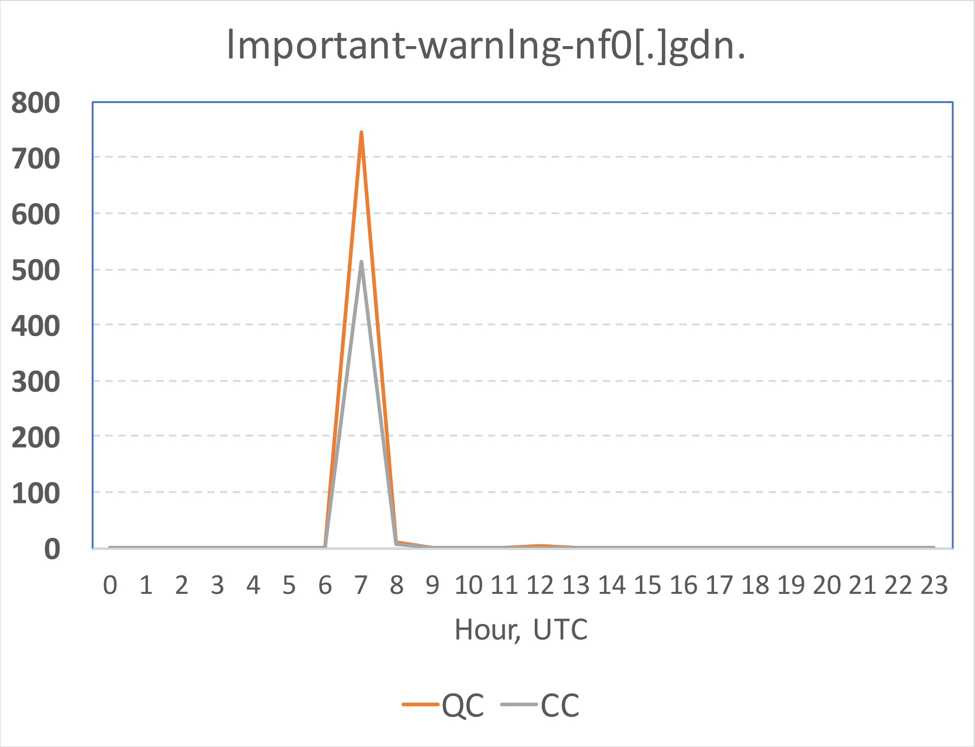
Browser hijacking: Waves





https://www.nominum.com/tech-blog/reclaiming-hijacked-browser/

Pre-infection: Sharing the Load


Important-warnIng-nf0[.]gdn.

Sample Domains and Peak Hour:

```
downioad-requlred-sf0.gdn.: Hour = 0
 lmportant-notlce-8a0.gdn.: Hour = 0
lmportant-warnlng-qf0.gdn.: Hour = 1
downioad-requlred-zf0.gdn.: Hour = 2
downioad-requlred-3g0.gdn.: Hour = 2
 lmportant-notlce-y90.gdn.: Hour = 2
   warnlng-n0tice-hg0.gdn.: Hour = 3
 lmportant-notlce-ga0.gdn.: Hour = 3
lmportant-warnlng-lf0.gdn.: Hour = 4
   warnlng-n0tice-cg0.gdn.: Hour = 5
   warnlng-n0tice-gg0.gdn.: Hour = 6
downioad-requlred-jf0.gdn.: Hour = 6
 lmportant-notlce-v90.gdn.: Hour = 6
```

Summary

- **Multi-stage processing is essential**
  - Maximize use of resources
  - Improve coverage and precision
- **NCDs are an essential component**
  - Real time classification engine
- **NCDs reveal valuable insights**
  - But they’re a ***starting point*** for further analysis
  - Combine with other intelligence (data and tools) to extract value