UCL

Marked for Disruption: Tracing the Evolution of Malware Delivery Operations Targeted for Takedown

Colin C. Ife   Yun Shen   Steven J. Murdoch   Gianluca Stringhini

EPSRC   snyk   Symantec   THE ROYAL SOCIETY   NSF   BOSTON UNIVERSITY

UCL

Steven J. Murdoch

UCL

THE ROYAL SOCIETY

Yun Shen

Symantec

Gianluca Stringhini

BOSTON UNIVERSITY

NSF

Colin C. Ife

UCL

EPSRC snyk

Appreciation to

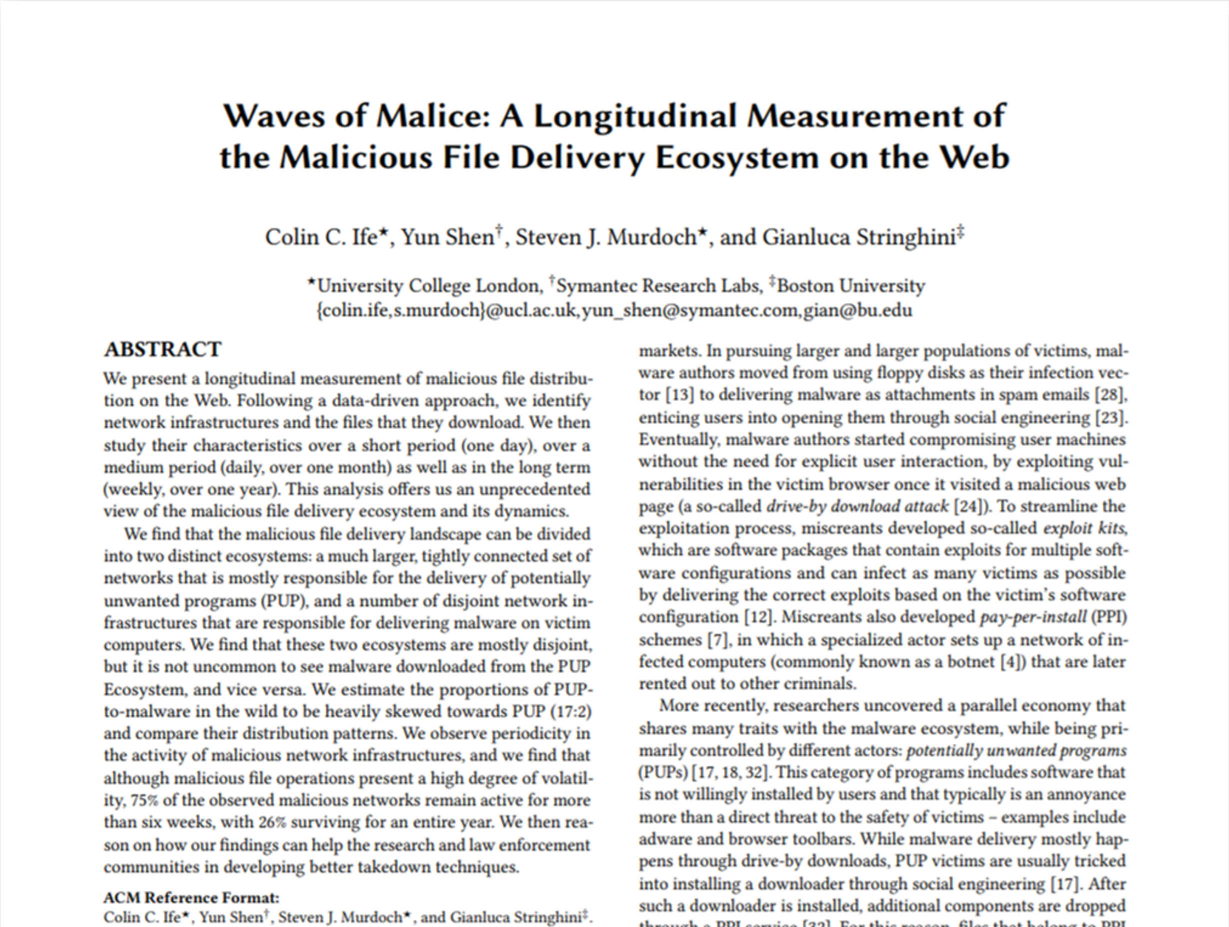
snyk

for making my attendance to
RAID ‘21 possible!

Agenda

- ❑ Motivation
- ❑ Methodology
- ❑ Key Findings
- ❑ Summary

Motivation



Waves of Malice: A Longitudinal Measurement of the Malicious File Delivery Ecosystem on the Web

Colin C. Ife*, Yun Shen†, Steven J. Murdoch*, and Gianluca Stringhini‡

*University College London, †Symantec Research Labs, ‡Boston University
{colin.ife,s.murdoch}@ucl.ac.uk,yun_shen@symantec.com,gian@bu.edu

ABSTRACT

We present a longitudinal measurement of malicious file distribution on the Web. Following a data-driven approach, we identify network infrastructures and the files that they download. We then study their characteristics over a short period (one day), over a medium period (daily, over one month) as well as in the long term (weekly, over one year). This analysis offers us an unprecedented view of the malicious file delivery ecosystem and its dynamics.

We find that the malicious file delivery landscape can be divided into two distinct ecosystems: a much larger, tightly connected set of networks that is mostly responsible for the delivery of potentially unwanted programs (PUP), and a number of disjoint network infrastructures that are responsible for delivering malware on victim computers. We find that these two ecosystems are mostly disjoint, but it is not uncommon to see malware downloaded from the PUP Ecosystem, and vice versa. We estimate the proportions of PUP-to-malware in the wild to be heavily skewed towards PUP (17:2) and compare their distribution patterns. We observe periodicity in the activity of malicious network infrastructures, and we find that although malicious file operations present a high degree of volatility, 75% of the observed malicious networks remain active for more than six weeks, with 26% surviving for an entire year. We then reason on how our findings can help the research and law enforcement communities in developing better takedown techniques.

ACM Reference Format:
Colin C. Ife*, Yun Shen†, Steven J. Murdoch*, and Gianluca Stringhini‡.

markets. In pursuing larger and larger populations of victims, malware authors moved from using floppy disks as their infection vector [13] to delivering malware as attachments in spam emails [28], enticing users into opening them through social engineering [23]. Eventually, malware authors started compromising user machines without the need for explicit user interaction, by exploiting vulnerabilities in the victim browser once it visited a malicious web page (a so-called *drive-by download attack* [24]). To streamline the exploitation process, miscreants developed so-called *exploit kits*, which are software packages that contain exploits for multiple software configurations and can infect as many victims as possible by delivering the correct exploits based on the victim's software configuration [12]. Miscreants also developed *pay-per-install* (PPI) schemes [7], in which a specialized actor sets up a network of infected computers (commonly known as a botnet [4]) that are later rented out to other criminals.

More recently, researchers uncovered a parallel economy that shares many traits with the malware ecosystem, while being primarily controlled by different actors: *potentially unwanted programs* (PUPs) [17, 18, 32]. This category of programs includes software that is not willingly installed by users and that typically is an annoyance more than a direct threat to the safety of victims – examples include adware and browser toolbars. While malware delivery mostly happens through drive-by downloads, PUP victims are usually tricked into installing a downloader through social engineering [17]. After such a downloader is installed, additional components are dropped through a PPI service [32]. For this reason, files that belong to PPI

Extending Prior Work

- **We already used download metadata to characterise malware delivery networks (MDNs) on the Web ("Waves of Malice," 2019).**
- ***What else could this data tell us?***
- ***"How effective are botnet takedowns on malware delivery?"*** 🤔
    - Botnet: a network of malware-infected devices controlled by an actor.
    - Takedown: an offensive technique used to disrupt a botnet.

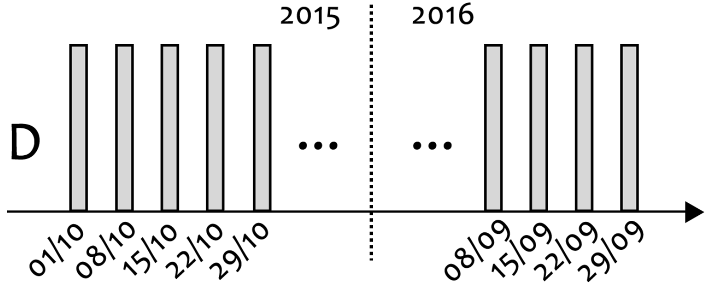
Research Aims

- ❑ **Analyse the evolution of malware delivery operations that were targeted for takedown.**
- ❑ **Answer important questions, such as:**
  1. After a takedown operation, what happens next? How do the operators react?
  2. For the targeted malware operations, are there additional or better intervention points?

Research Methodology

Dataset

- ❑ **Symantec download telemetric data**
- ❑ **81.5M download events (from 12M users)**



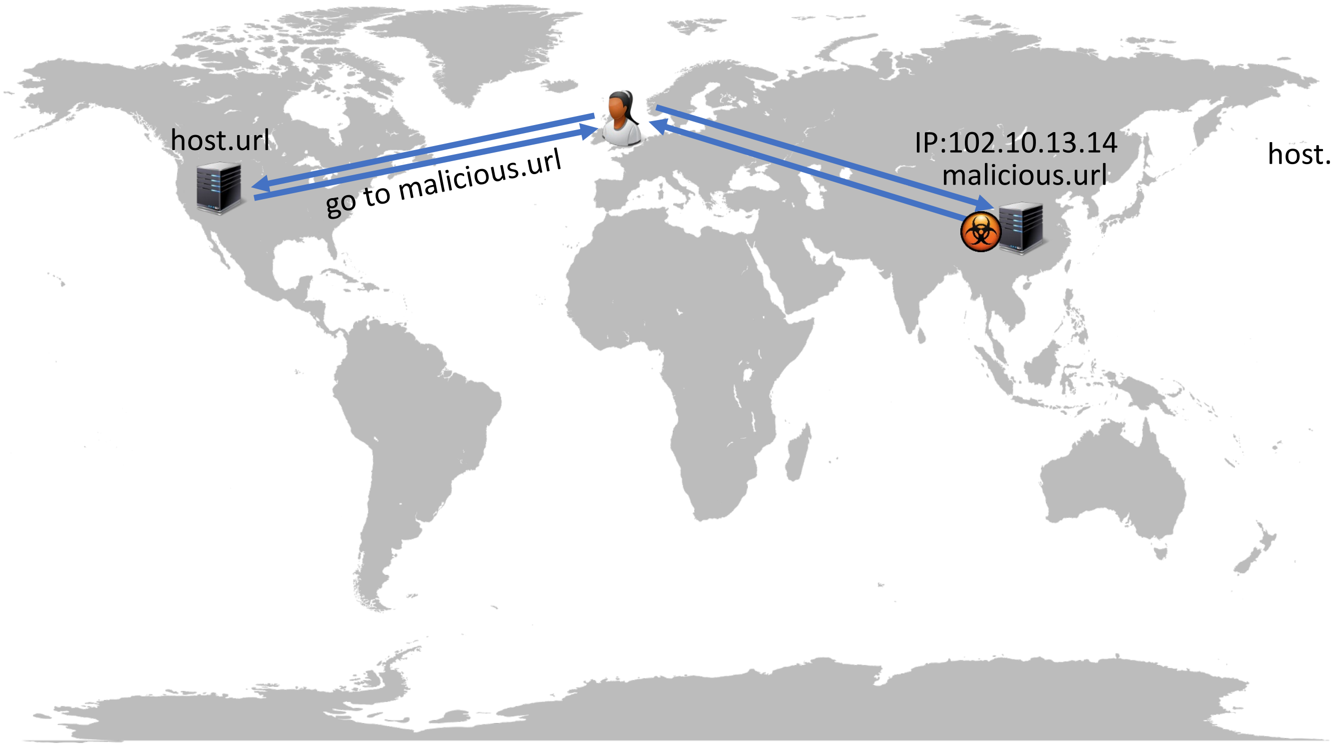
- ❑ **Focus on malicious files → Low reputation score**

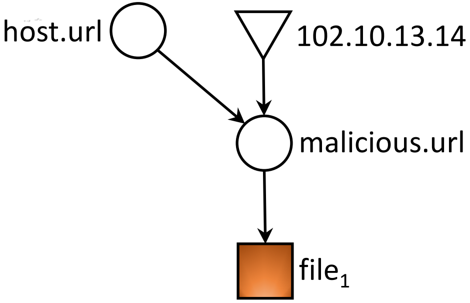
Dataset

A download event includes:

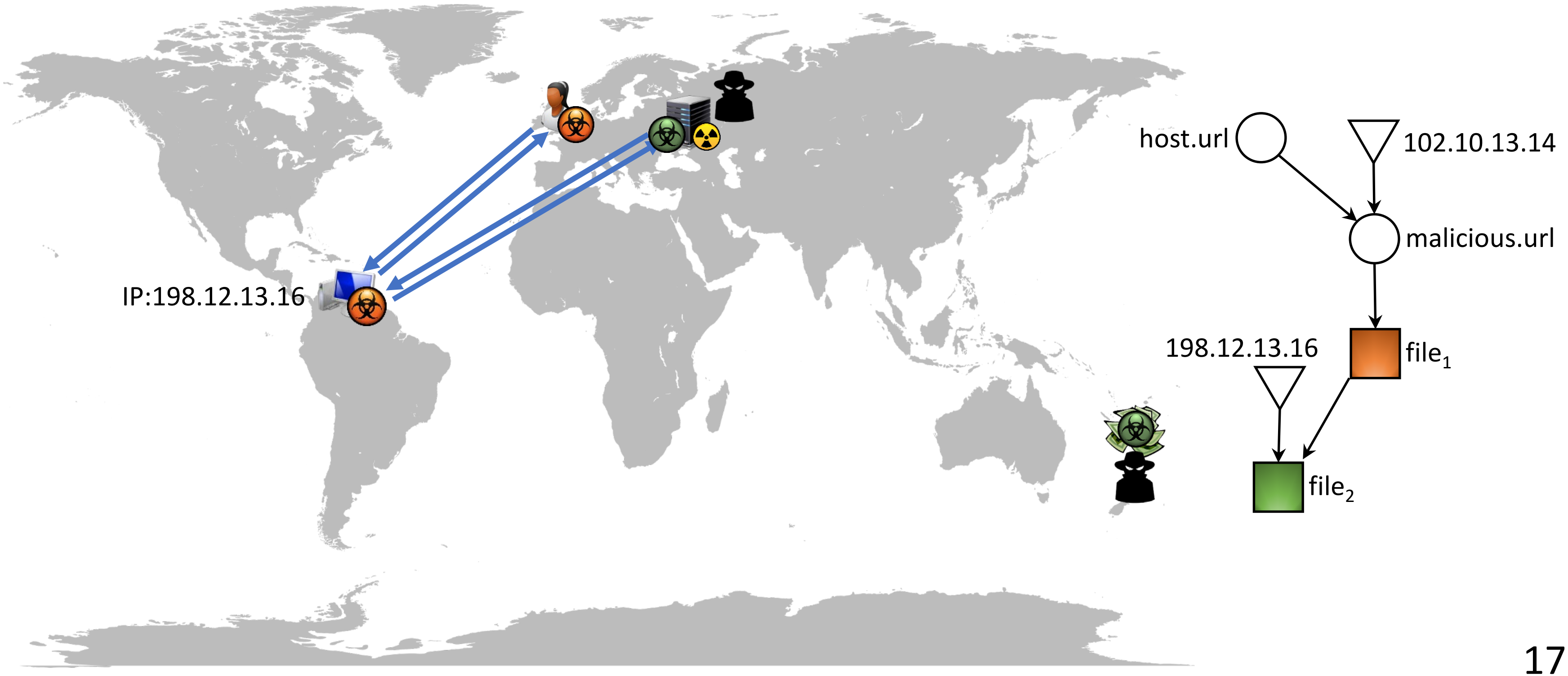
- Timestamp
- **SHA-2 of file (256 bits)**
- File name
- Size of file in bytes
- **Host URL**
- **Landing page URL** (redirects to Host URL)
- **IP address** of server hosting file
- **Parent file SHA-2**
- **Landing page URL of parent file**

An example of a malicious file delivery event





An example of a malicious file delivery event

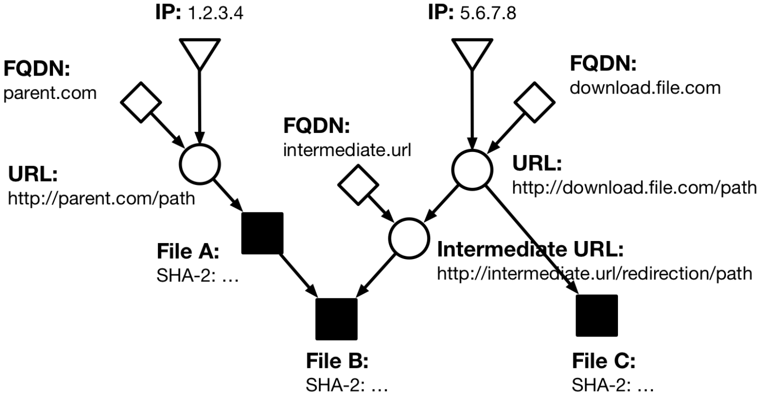


Graph Abstraction

- ❑ Graph-building technique based on prior work:

  (Ife et al., 2019)

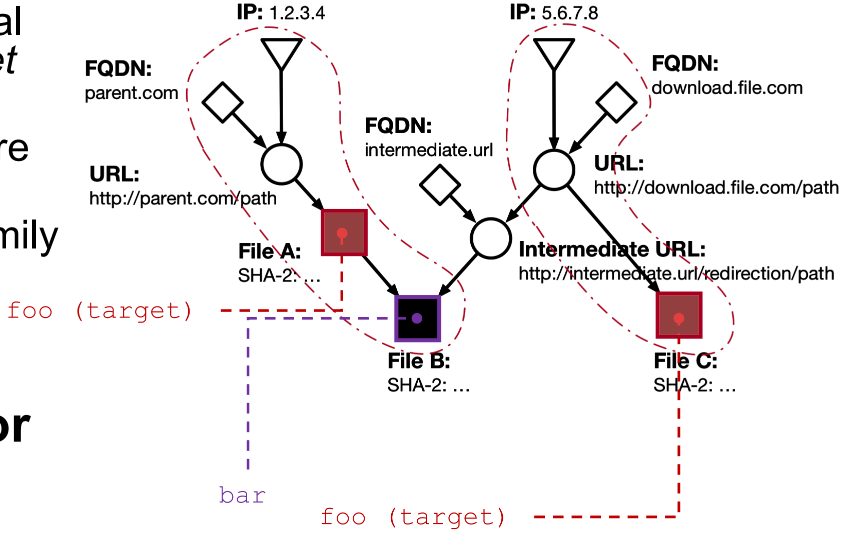
  - Each unique file (SHA-2), host, or IP address are represented as **nodes**.
  - Downloads and network-level associations are represented as **directed edges**.



IP: 1.2.3.4
IP: 5.6.7.8
FQDN: parent.com
FQDN: download.file.com
FQDN: intermediate.url
URL: http://parent.com/path
URL: http://download.file.com/path
File A: SHA-2: …
Intermediate URL: http://intermediate.url/redirection/path
File B: SHA-2: …
File C: SHA-2: …

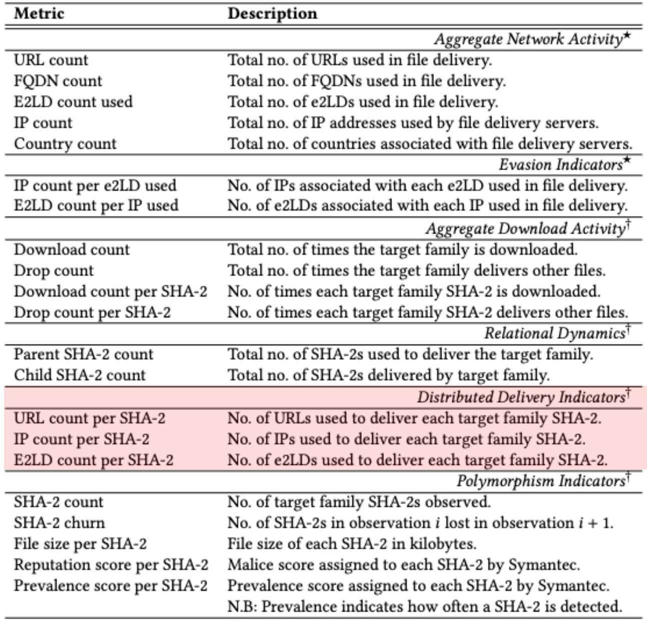
Tracking and Analysing Operations

- ❑ **For each 24-hour graph snapshot:**
  - Assign labels to file hashes using VirusTotal data and the AVClass labeller *(Sebastián et al., 2016)*.
  - Aggregate all nodes pertaining to a malware family (the **"target family"**).
  - Aggregate all nodes linked to the target family nodes.
  - All nodes connected to a target family represent its **global delivery operation**.
- ❑ **Compute time-series metrics for each target family's global delivery operation.**



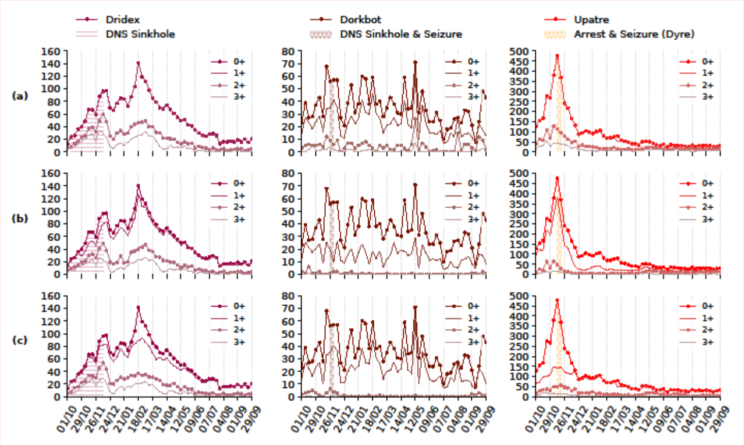
IP: 1.2.3.4
IP: 5.6.7.8
FQDN: parent.com
FQDN: download.file.com
FQDN: intermediate.url
URL: http://parent.com/path
URL: http://download.file.com/path
File A: SHA-2: ...
Intermediate URL: http://intermediate.url/redirection/path
foo (target)
File B: SHA-2: ...
File C: SHA-2: ...
bar
foo (target)

Tracking and Analysing Operations



Metric	Description
	Aggregate Network Activity★
URL count	Total no. of URLs used in file delivery.
FQDN count	Total no. of FQDNs used in file delivery.
E2LD count used	Total no. of e2LDs used in file delivery.
IP count	Total no. of IP addresses used by file delivery servers.
Country count	Total no. of countries associated with file delivery servers.
	Evasion Indicators★
IP count per e2LD used	No. of IPs associated with each e2LD used in file delivery.
E2LD count per IP used	No. of e2LDs associated with each IP used in file delivery.
	Aggregate Download Activity†
Download count	Total no. of times the target family is downloaded.
Drop count	Total no. of times the target family delivers other files.
Download count per SHA-2	No. of times each target family SHA-2 is downloaded.
Drop count per SHA-2	No. of times each target family SHA-2 delivers other files.
	Relational Dynamics†
Parent SHA-2 count	Total no. of SHA-2s used to deliver the target family.
Child SHA-2 count	Total no. of SHA-2s delivered by target family.
	Distributed Delivery Indicators†
URL count per SHA-2	No. of URLs used to deliver each target family SHA-2.
IP count per SHA-2	No. of IPs used to deliver each target family SHA-2.
E2LD count per SHA-2	No. of e2LDs used to deliver each target family SHA-2.
	Polymorphism Indicators†
SHA-2 count	No. of target family SHA-2s observed.
SHA-2 churn	No. of SHA-2s in observation i lost in observation $i + 1$.
File size per SHA-2	File size of each SHA-2 in kilobytes.
Reputation score per SHA-2	Malice score assigned to each SHA-2 by Symantec.
Prevalence score per SHA-2	Prevalence score assigned to each SHA-2 by Symantec. N.B: Prevalence indicates how often a SHA-2 is detected.

Table 1: The network★ and downloader† metrics used to analyse each malware delivery operation.



Malware Operations Studied

❑ Dridex

- A trojan that steals banking credentials. Operates as a payload only.
- Spreads through malicious emails, adjacent networks, and exploit kits hosted on compromised webpages.
- Takedown: 60-day DNS Sinkhole and Disinfection from early Oct 2015 led by the FBI. Two other takedowns occurred between Aug–Sep 2015.

❑ Dorkbot

- A family of worms known to steal data from compromised systems, disable security apps, and distribute other malware.
- Takedown: DNS Sinkhole and Seizure in Dec 2015 by a collaboration of security companies and law enforcement.

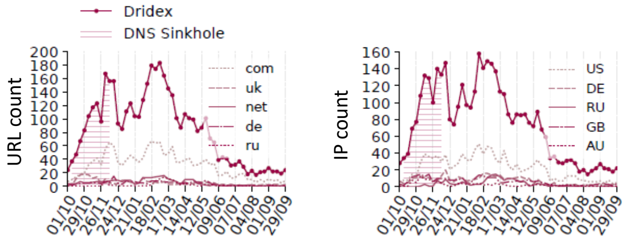
❑ Upatre

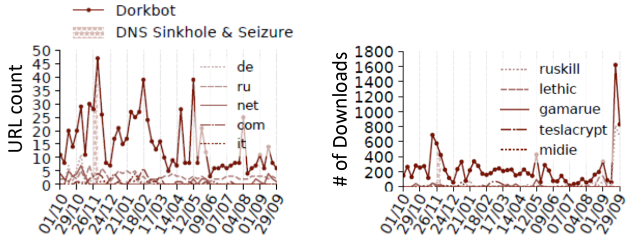
- A dedicated dropper malware known most for delivering *Dyre*.
- We identified a correlation between the Dyre takedown and significant drops in Upatre activity. *(Ife et al., 2019)*
- Takedown: Arrest and Seizure against Dyre in Nov 2015 led by Russian law enforcement.

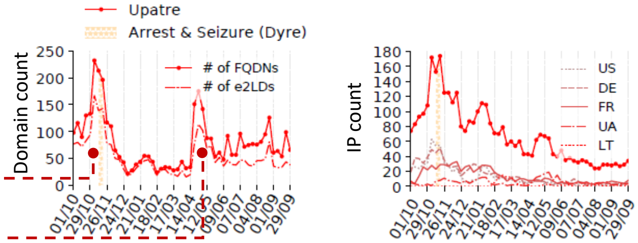
Key Results

Takedown Resilience and Predictability

- ❑ **Each malware operation responded differently, but all showed resilience to takedown.**
  - Dridex ramped up server usage during the 60-day sinkhole, increasing its concentration of servers in the US and UK.
  - Dorkbot showed no major changes to its operation after the takedown.
  - Upatre activity dropped in the short-term **BUT** shifted to a centralised infrastructure several months after the Dyre takedown.







Change in domains used: `*.ymail.com, *.afx.ms`

`*.alfafile.net, slingto.*.ru (DGA)`

Takedown Resilience and Predictability

❑ **Criminology recognises a number of common offender reactions to anticrime interventions:**

- *Displacement*: a change in an offender's behaviour to circumvent an intervention. ↗️
- *Defiance*: an increase in offender activity in retaliation to an intervention. 😡

❑ **The malware operators' reactions were characteristic of these behavioural models.**

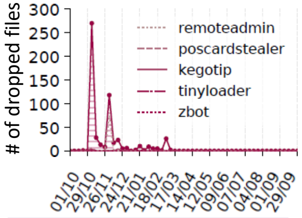
- Dridex ramped up server usage during the 60-day sinkhole, increasing server concentration in the US and UK. ↗️ 😡
- Upatre shifted to a more centralised infrastructure several months after the Dyre takedown. ↗️

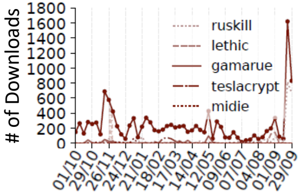
→ **Factor in side-effects for more effective takedown strategies.**
→ **Are takedowns the only way forward?**

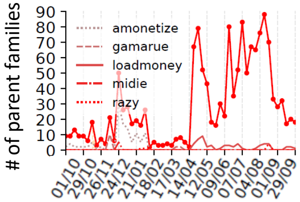
…and the Unpredictable

❑ **We observed anomalous and previously undocumented behaviours:**

- Dridex – a secret malware distributor! Bursts of ransomware, backdoors, and competing brands of banking trojans were delivered.
- Dorkbot exhibited a massive spike in downloads through Ruskill in late 2016 - an emerging business relationship?
- Upatre also exhibited many deliveries through multiple upstream malware in 2016 – what led to this significant change in delivery model?





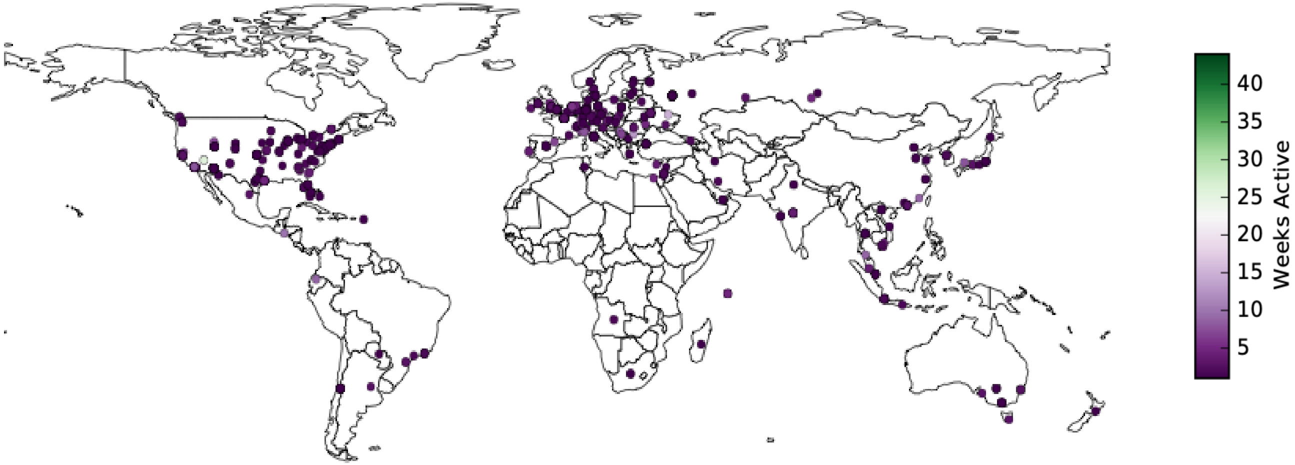


→ **Need for better monitoring techniques, particularly using <u>multiple intelligence sources</u>.**

Distributed Delivery Architectures

- ❑ **All three malware operations made significant use of distributed delivery methods…**
  - Dridex used shared-hosting services and CDNs in up to 35 different countries.

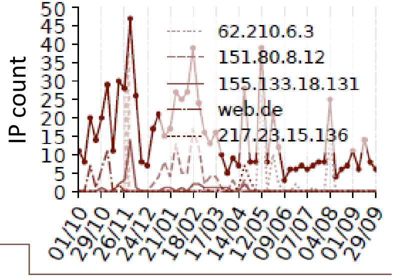
Distributed Delivery Architectures



Dridex Geographic Server Activity

Distributed Delivery Architectures

- ❑ **All three malware operations made significant use of distributed delivery methods…**
  - Dridex used shared-hosting services and CDNs in up to 35 different countries.
  - Dorkbot constantly rotated delivery through different international servers (“fast” and “slow” flux).
  - Upatre heavily used multi-region CDNs and cloud-based services (e.g., `ymail.com`, `alfafile.net`).
- ❑ **…making server-based takedowns more difficult (redundancy, detection).**



→ **More coordination required within the international security community.**
→ **Need for better security hygiene among the abused services.**

Polymorphism and “Super Binaries”

- ❑ **Polymorphic malware change their identifiable features to avoid detection.**
- ❑ **Polymorphism rigorously employed by all three malware:**
  - 10s to 100s of SHA-2s used per day for each operation.
  - Dorkbot was the most elusive: throughout the year, ~50% of hashes were assigned very low malice scores by Symantec systems.
- ❑ **We also observed Pareto’s Principle (80-20 rule): a minority of files were responsible for most download activity.**
  - E.g., for Dridex, less than 1% of binaries were responsible for all dropping activity over the year.

> → **Hash-based tracking? Good luck!**
>
> → **Focus on flagging these “super binaries” to disrupt malware delivery most effectively.**

…and much more in the paper!

Limitations

- ❑ **Inherited limitations from the previous study:**
  - Limited view of only one stage of the malware supply chain (delivery).
  - VirusTotal’s limited file coverage.
  - False positive malware labels.
- ❑ **Lack of ground truth on specifics of takedown operations.**
- ❑ **Survivorship bias and generalisability.**
- ❑ **Correlation does not imply causation! → A future research direction?**

Summary

- ❑ **Using download metadata, we devised a novel technique to model, track, and dissect malware delivery operations on the Web.**
  - → Graph-building code available at https://github.com/colinife/mdn
- ❑ **We applied this technique to study three different malware operations, making a number of key findings:**
  - The tendency of malware operators to move their operations elsewhere after a takedown

    → **should be factored into takedown strategies to manage these side-effects**
  - The common use of distributed delivery architectures (particularly through CDNs), making coordinated takedowns harder

    → **need for greater coordination; better security practices among service providers**
  - The presence of “super binaries” which carry out most delivery activity in an operation

    → **detecting and disrupting these would yield the most impact**
  - We discovered some previously undocumented malware behaviours

    → **need for better monitoring techniques for malware operations**

Thank you for listening!

colin.ife@snyk.io

🐦@ColinIfe

colinife.com