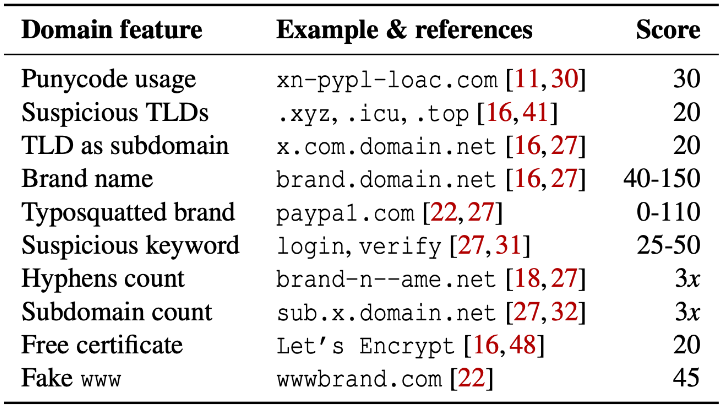
Catching Phishers by Their Bait

Motivation

- Phishing *kits* are commonly used by attackers
  - New avenue for detection
- Attacks are localized for a region
  - Dutch banking sector good target
- Detection trails attack, need more proactive approach for study
  - … and defense?

Methodology



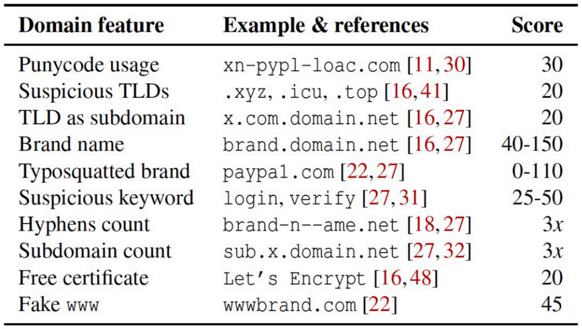
Domain feature	Example & references	Score
Punycode usage	`xn-pypl-loac.com` [11,30]	30
Suspicious TLDs	`.xyz`, `.icu`, `.top` [16,41]	20
TLD as subdomain	`x.com.domain.net` [16,27]	20
Brand name	`brand.domain.net` [16,27]	40-150
Typosquatted brand	`paypa1.com` [22,27]	0-110
Suspicious keyword	`login`, `verify` [27,31]	25-50
Hyphens count	`brand-n--ame.net` [18,27]	3*x*
Subdomain count	`sub.x.domain.net` [27,32]	3*x*
Free certificate	`Let's Encrypt` [16,48]	20
Fake `www`	`wwwbrand.com` [22]	45

- Gather phishing kits from fraud channels
- Create *fingerprint* for phishing kit
- Gather suspicious domains from certificate transparency logs
- Crawl suspicion domains, look for fingerprinting

Kit Collection and Fingerprints

- 50 Telegram channels using “snowball sampling”
- 70 kits downloaded
    - Free samples, leaked source!
- Fingerprints based on filenames, common text

Domain Collection

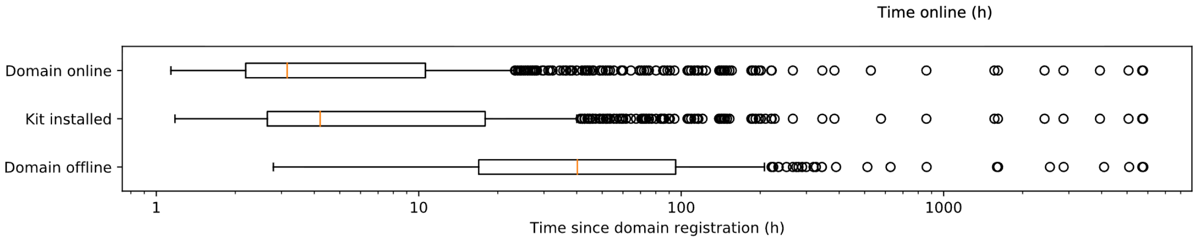


Domain feature	Example & references	Score
Punycode usage	`xn-pypl-loac.com` [11, 30]	30
Suspicious TLDs	`.xyz`, `.icu`, `.top` [16, 41]	20
TLD as subdomain	`x.com.domain.net` [16, 27]	20
Brand name	`brand.domain.net` [16, 27]	40-150
Typosquatted brand	`paypa1.com` [22, 27]	0-110
Suspicious keyword	`login`, `verify` [27, 31]	25-50
Hyphens count	`brand-n--ame.net` [18, 27]	$3x$
Subdomain count	`sub.x.domain.net` [27, 32]	$3x$
Free certificate	`Let's Encrypt` [16, 48]	20
Fake `www`	`wwwbrand.com` [22]	45

- Every HTTPS domain registration appears in a CT log
    - Great measurement resource!
- Scan for “suspicious” domains
    - Score based on features
- Monitor domain content for a week
    - Look for phishing fingerprint, changes
- Evidence of attackers located in the Netherlands!

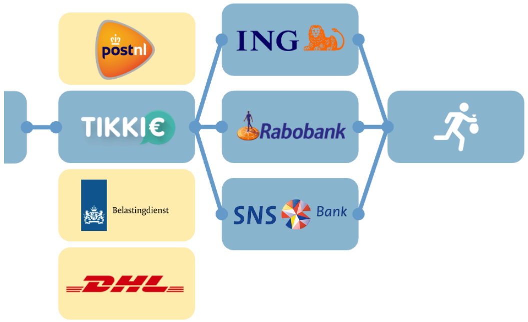
Timeline

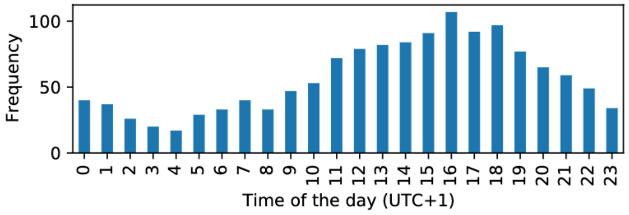
- 24h median, 45h average uptime



Interesting Observations

- Namecheap (takes BTC) and Let’s Encrypt big enablers
- Multi-step attacks common
  - Visits start at a non-banking site, then follow clicks
- Anti-scan measures
  - Blank page, redirect, “site taken down”
- Psychological analysis
  - Scarcity and consistency





Strengths and Weaknesses

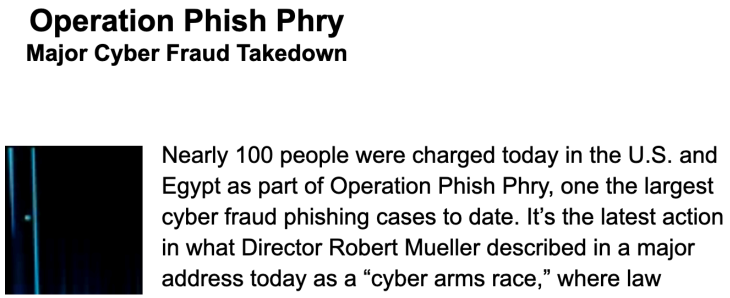
- Strong measurement paper
- Lifecycle focus
- Manual validation to avoid Fps
- Good scoping

- Incomplete perspective
- Fingerprints seem easy to defeat
- Ad-hoc approach
- Manual effort required to make things work

Contributions

- Is their methodology….
  - Merely an enabler for their results?
  - Usable in future research?
  - Provide a new *detection* strategy?
- Compare and contrast the scale and validity of this paper with the previous one

Intervention and Risk



Operation Phish Phry
Major Cyber Fraud Takedown

Nearly 100 people were charged today in the U.S. and Egypt as part of Operation Phish Phry, one the largest cyber fraud phishing cases to date. It's the latest action in what Director Robert Mueller described in a major address today as a "cyber arms race," where law

- What would be a usable *fast intervention* signal
  - Is suspicious domain + kit fp enough to avoid false positives?
- Who should intervene? Who has the incentives?
  - Hosting providers? Domain registrars? Law enforcement?
- Can we create barriers to some of the "choke points"?

Comprehensive Look?

- What biases are introduced by the methodology?
- What conclusions can / cannot be drawn based on the results?
- Can we generalize from the Dutch experience?

Other Discussion