Every Second Counts:
Quantifying the Negative Externalities of Cybercrime via Typosquatting

Mohammad Taha Khan*, Xiang Huo*, Zhou Li† & Chris Kanich*
University of Illinois at Chicago* & RSA Labs†
Cybersecurity In General...

• Generally focus on:
  ▪ Detecting malicious programs
  ▪ Finding and fixing bugs and flaws
  ▪ Economic analyses

**Why Is This Important?**

• The ultimate goal:
  ▪ Minimize the harm caused to users
  ▪ Harm: Monetary, *wasted effort, loss of time*
Typosquatting

One of the common forms of cybercrime

Typosquatting Attacker Model:

§ Identify and register common typos of established websites
§ Host advertisements, malware or competing content

Not really popular from malware standpoint

Typosquatting causes visitor loss and a loss of user time.

An Example...

This offer is available today: Thursday, April 23, 2015

Congratulations!

Dear User,

You have been selected to take part in our anonymous survey! Complete this 30 second questionnaire, and to say "thank you", we'll offer you an exclusive prize. Today's product is a $1000 Gift Card or $150 Visa Gift Card.
Typosquatting

- Evidence that typosquatting is PERVERSIVE:
  - Large organizations invest into defensive registrations
  - Internet users continue to make typos

- What makes it FEASIBLE to study:
  - Observable from a network level
  - Can infer User intent from available data
Our Contributions

- Passive detection of typosquatting domains using a conditional probability model
- Present a harm metric in the form of loss of time and users
- Apply this metric to quantify the cybercrime of typosquatting
- Our work uses an open methodology with fine grained measurements
Data-Sets

• Passive Sources
  • HTTP data logs
  • DNS logs from recursive resolver
  • Enterprise proxy data

• Active Sources
  • High Fidelity Crawler
User Intent

- User Intent to visit the destination website generates similar pairs of domains.

- User intent is manifested in various discovery methods.
Conditional Probability

Characteristics of typo domain pairs $(d_1,d_2)$:

- Damerau-Levenshtein distance = 1
- Pairs at most 33 seconds apart
- Latter $d_2$, must be a resolvable domain
- Appeared at least 3 times in the dataset

Percentage Accuracy

Bins:
- $(0, 0.1]$
- $(0.1, 0.2]$
- $(0.2, 0.3]$
- $(0.3, 0.4]$
- $(0.4, 0.5]$
- $(0.5, 0.6]$
- $(0.6, 0.7]$
- $(0.7, 0.8]$
- $(0.8, 0.9]$
- $(0.9, 1]$

Threshold Cutoff = 0.2

- Eba.com followed by Ebay.com: 90%
- nhl.com followed by nfl.com: 0.08%

Total Distinct Typo Domains = 34,400
Typo Characterization

- Adversarial registrations
  - Parked Domains
  - Malicious Websites
  - Other

- Cooperative registrations
  - JavaScript and 3xx redirections
  - Defensive registrations

- Unregistered websites
  - NX Domains
What Next...??
Quantifying Harm

Typo Delay (seconds)

Proportion of Typo Pairs

- Adversarial
- Cooperative
- Unregistered

Average Delay (s)

- Cooperative: 2.87
- Adversarial: 9.58
- Unregistered: 10.38

Average User Loss (%)

- Cooperative: 3.30
- Adversarial: 16.81
- Unregistered: 11.53
Search VS Typein Delays

- Different discovery methods show varying delay trends

Unregistered Domains
Target Domain Category

- Most Typos exist in the long tail of popularity

- Most distinct typos belonged to Adult and Blogs

![Bar Chart]

- Normalized Typo Visits
- Categories: Search Engines, Social Networks, Adult, Blogs, Digital Media, News, Online Shopping
- Y-axis: \(10^{-3}\)
Delay & Success Clustering

- Some typo domains help users to get faster to their destination websites

- Top service providers are also popular for domain parking

Loss of Revenue

• Convert time and user loss into dollars.

• Intended site owner has a negative externality ratio of **18:1** against the typosquatter

• Using per capita income an average user loses **$0.29** to typosquatting per year

• For defenders, the effort ratio is **4.62:1**, far lower than non-violent crime*

Conclusions

• Typosquatting is much less societally damaging than other non-violent crimes

• Defensive registrations do help against mistyping but not much against typosquatting

• Special technical or policy interventions are not necessarily required to deal with it
Thank You! Questions?