Leveraging Google SafeBrowsing to Characterize Web-based Attacks

Peter Likarish and Eunjin Jung
The University of Iowa

Intro

Websites hosting malware:

- Increased 46%(Jan/08 – Jan/09)
- 77% compromised legitimate websites
- When detected, moved to new dom

Our research:

•Goal: Predict likelihood attack is hosted at domain

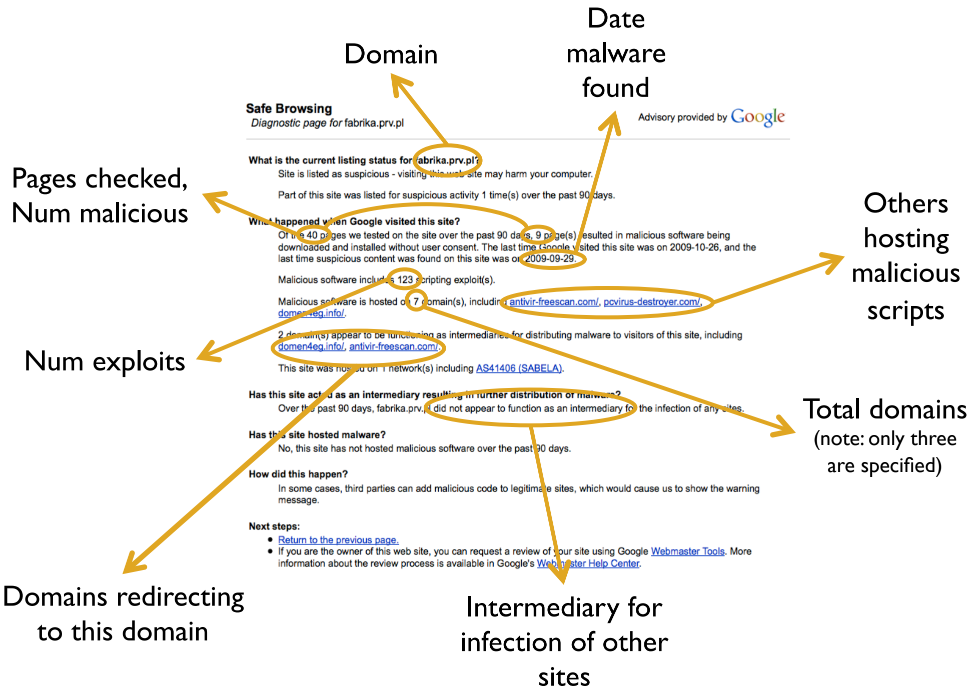
- Observe structure of web attacks using inter-domain attack graph
- Leverage Google SafeBrowsing tool (GSBDt)

Method

Re-construct attack graph from GSBDt

- Need seed sites to query GSBDt
  - http://www.malwaredomains.com
  - 3,465 blacklisted domains in last 90 days
- Construct domain-level attack graph
  - Directed edge = redirection
  - Two types of nodes: intermediaries and malicious software hosts
  - Recursively fetch new domains

Google SafeBrowsing Diagnostic tool (GSBDt) info


Domain
Date malware found
Pages checked, Num malicious
Others hosting malicious scripts
Num exploits
Total domains
(note: only three are specified)
Domains redirecting to this domain
Intermediary for infection of other sites

Safe Browsing
Diagnostic page for fabrika.prv.pl

Advisory provided by Google

What is the current listing status for fabrika.prv.pl?

Site is listed as suspicious - visiting this web site may harm your computer.

Part of this site was listed for suspicious activity 1 time(s) over the past 90 days.

What happened when Google visited this site?

Of the 40 pages we tested on the site over the past 90 days, 9 page(s) resulted in malicious software being downloaded and installed without user consent. The last time Google visited this site was on 2009-10-26, and the last time suspicious content was found on this site was on 2009-09-29.

Malicious software includes 123 scripting exploit(s).

Malicious software is hosted on 7 domain(s), including antivir-freescan.com/, pcvirus-destroyer.com/, domen4eg.info/.

2 domain(s) appear to be functioning as intermediaries for distributing malware to visitors of this site, including domen4eg.info/, antivir-freescan.com/.

This site was hosted on 1 network(s) including AS41406 (SABELA).

Has this site acted as an intermediary resulting in further distribution of malware?

Over the past 90 days, fabrika.prv.pl did not appear to function as an intermediary for the infection of any sites.

Has this site hosted malware?

No, this site has not hosted malicious software over the past 90 days.

How did this happen?

In some cases, third parties can add malicious code to legitimate sites, which would cause us to show the warning message.

Next steps:

- Return to the previous page.
- If you are the owner of this web site, you can request a review of your site using Google Webmaster Tools. More information about the review process is available in Google's Webmaster Help Center.

Preliminary results

Comparison of attack domains on Malwaredomains and GSBDt

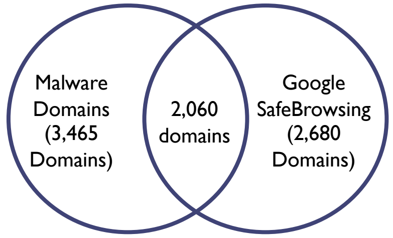

Malware Domains (3,465 Domains)
2,060 domains
Google SafeBrowsing (2,680 Domains)

Figure 1.The number of malicious domains according to each source. The intersection is the number of attacks detected by both sources.

Blacklist "coverage"

- Combined: detected attacks at 2,060 domains.
- Both missed many domains blacklisted by other service.
  - 44% of domains on Malwaredomains had *no* malicious activity on GSBDt
  - 32% of domains with malicious activity on GSDBt had *no* malicious activity on Malwaredomains

Attack details

- 1,063 disjoint attacks
- Average attack size: 5.37 domains
- Number of singleton attacks: 345 domains
- Attack size - singletons: 7.47 domains

Dealing with unknown attack domains

- Reduce with larger set of known attacks.
- Only generate subgraph of fully discovered attacks

Attack graph example

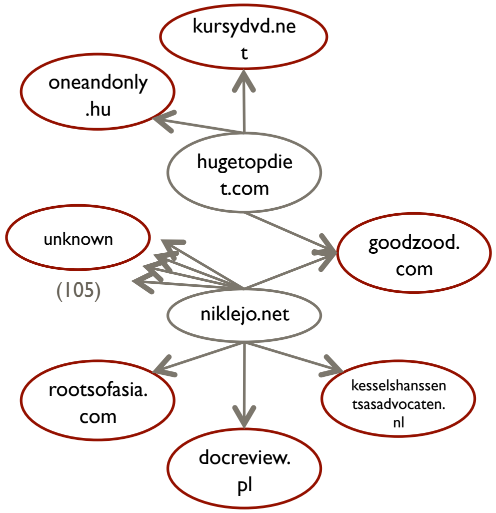


Figure 2. An example attack graph. Grey domains/links are redirections. Red domains hosted malware or exploits.

Conclusion

- Can develop models for attack structures.
- Use model to predict likelihood of attacks.
- Better mechanism to deal with unknown attack domains.

Future Work

- Need larger number of seed sides to discover larger portion of GSBDt domains.
- Develop model for evolution of attacks over time.
  - Revisiting: few domains still malicious (> 90 days).