SOPHOS

simple + secure

Strategies for Monitoring Fake AV Distribution Networks

Onur Komili, Kyle Zeeuwen, Matei Ripeanu, Konstantin Beznosov

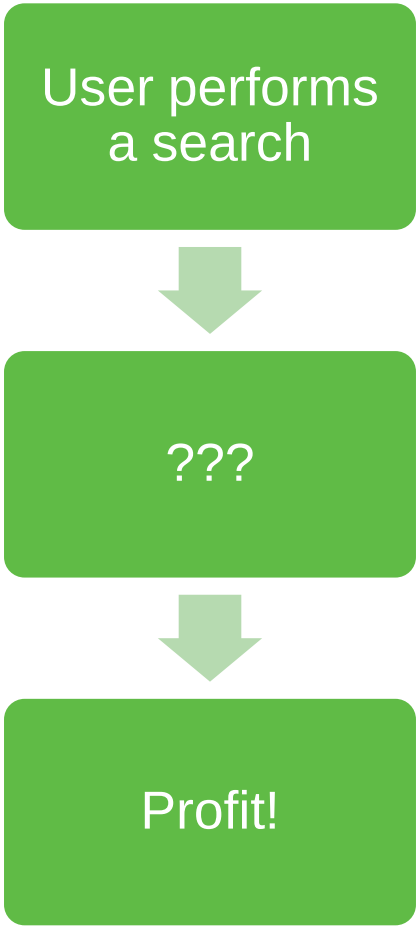
Introduction

- Security researchers study MDNs to counter malware
  - Block and take down network components
  - Detect malware binaries
  - Understand intent of malware
- MDNs are designed to evade the scrutiny of security researchers
  - Rapidly updating MDN structure and MDN content
  - Obfuscating content across various components
  - Identifying and blacklisting researchers through various methods

Introduction

- These behaviours introduce two problems for researchers:
  - **Unbounded growth in working set:** There are too many URLs to (re)evaluate
  - **Accuracy of content:** URL evaluation can be tainted by blacklisting
- **Our goal:** Optimize re-evaluation algorithm to reduce the number of evaluations per URL

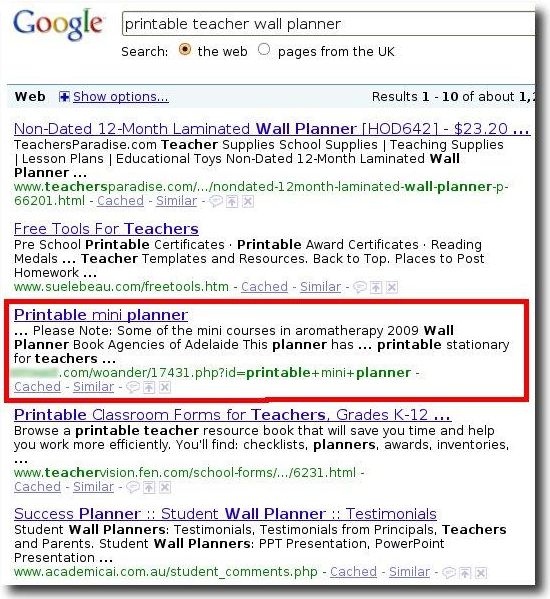
Typical FakeAV Attack



SOPHOS

Doorways

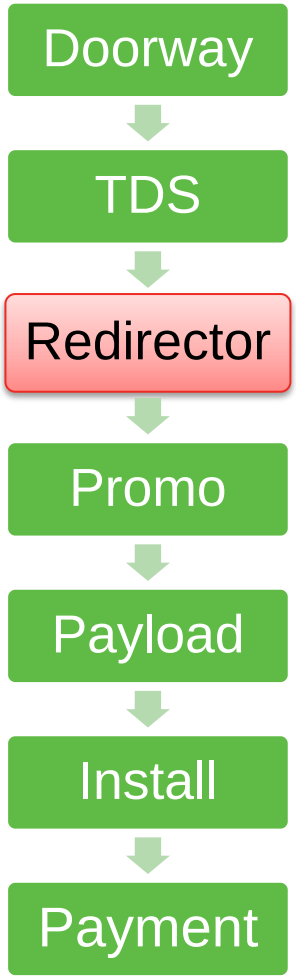
- Often on legitimate but compromised sites
- Serves keyword stuffed content to crawlers
- Google Trends or Auto Complete abuse
- Redirects user to next hop in redirect chain
- We focus on social engineering, not drive-by



Traffic Direction System (TDS)

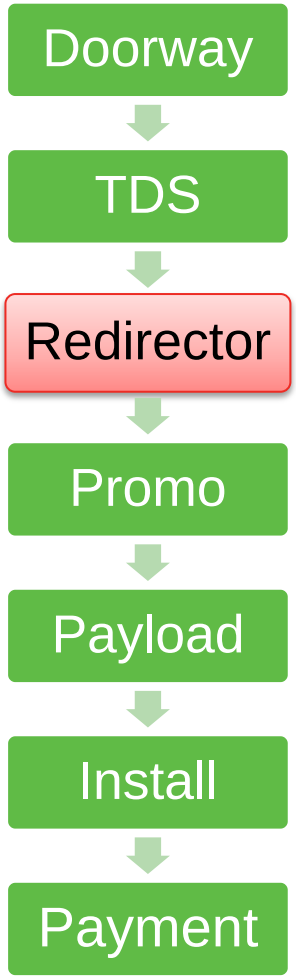
- Not always present
- SutraTDS, SimpleTDS, …
- Redirect traffic to malicious content based on
  - Country
  - Browser
  - OS
  - Keywords searched
  - etc…

Redirector



- Again, not always present
- Adds to the level of complexity
- Could have multiple redirectors
- Often hosted on bulk subdomain service sites

Bulk Subdomain Service

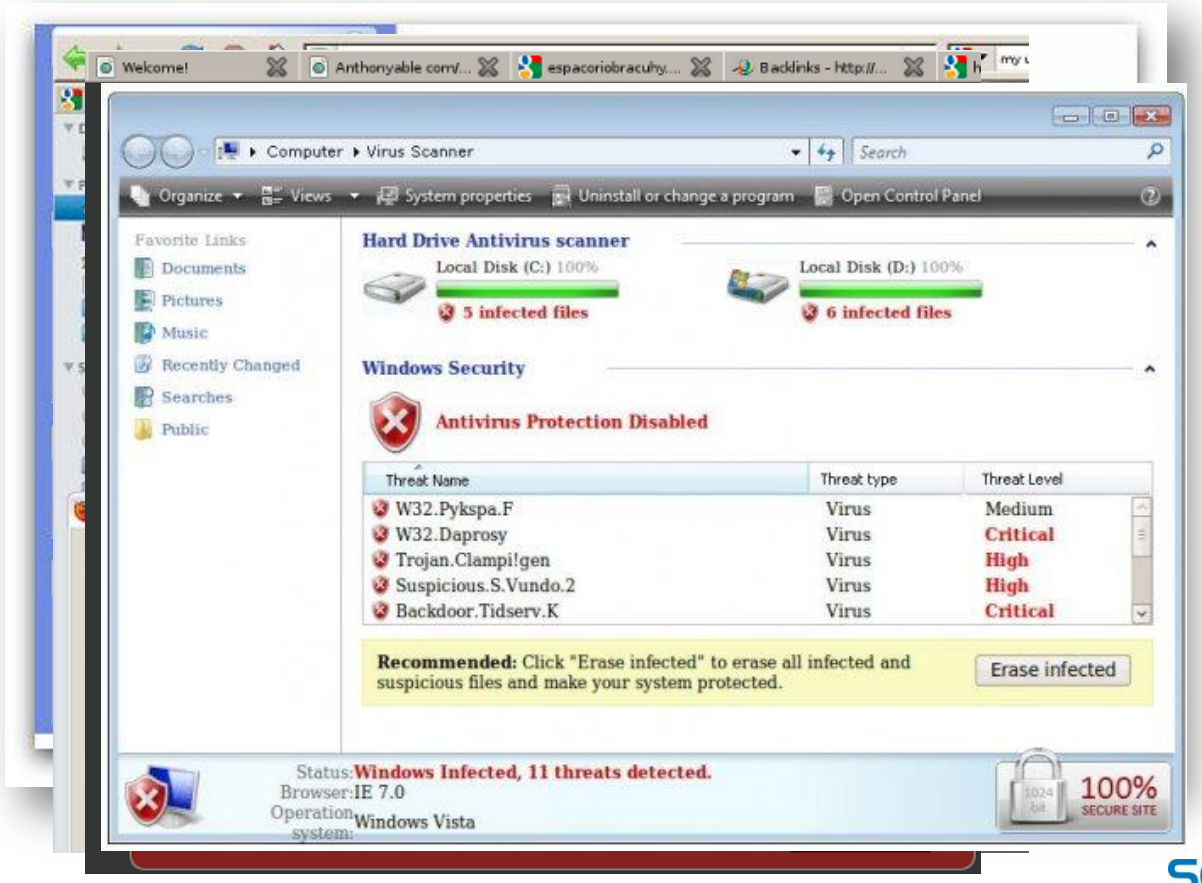


- Popular sites: co.cc, cx.cz, co.be, rr.nu, cz.cc, etc...
- Offer low-cost subdomains - as low as $0.07-$0.10 each
- Some AV vendors reluctant to block these services due to FP risk
- In June, Google began flagging many of the popular domains on their Google Safebrowsing list

Promotional Page

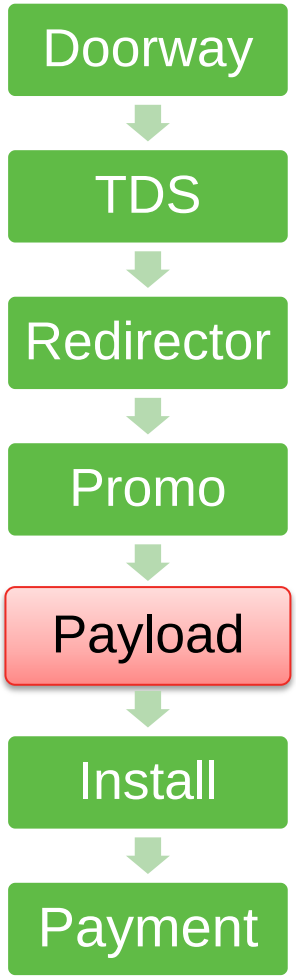
- Doorway
- TDS
- Redirector
- **Promo**
- Payload
- Install
- Payment

- Fake “My Computer” (or Finder) scanner page

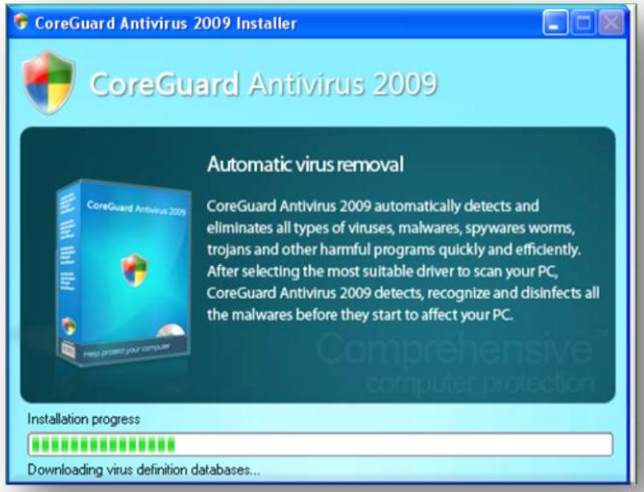


SOPHOS

Payload

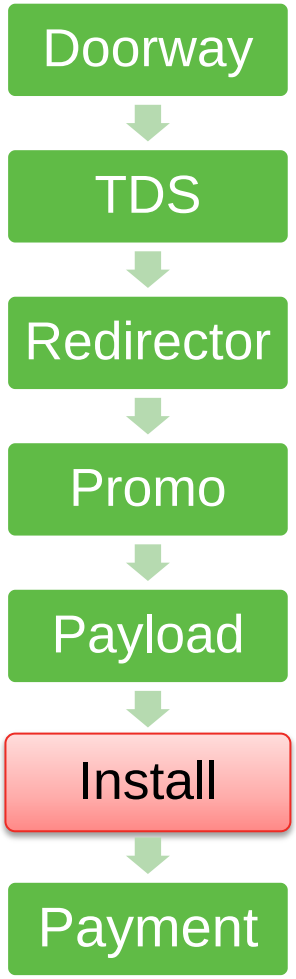


- Often hosted on the same location as the promo page
- Fully or partially polymorphic



SOPHOS

Install



- Not concerned with anything past the binary for this research

Payment

As an aside, a few points worth mentioning on the payment front:

- FakeAV is extremely profitable
- “The Underground Economy of Fake Antivirus Software” by Stone-Gross et al.
- Targeting them at the payment level has proven to be an effective method of reducing the amount of FakeAV seen in the wild

Why is Re-evaluation a problem?

- Daily volume of new potentially malicious URLs needing analysis is in the 100,000's
  - A subset provide additional value upon subsequent evaluation
  - Which URLs yield value, and for how long the URLs yield value is not well understood
- Every re-evaluation exposes the research IP pools to adversary
  - Assumption: More re-evaluation increases probability of blacklisting

Approach

- Systematically study malicious web sites over time to identify the distribution and prevalence of malware update behaviours
- Develop and evaluate optimizations to the re-evaluation logic based on update behaviours observed
- Identify the prevalence of IP blacklisting by MDNs and propose strategies to avoid blacklisting

Tachyon Detection Grid

- In house research tool developed to fetch and reschedule monitoring of URLs fed into the system
- Used to detect cloaking behaviour used by MDNs
- Uses a high interaction fetcher also developed in house
  - Virtual Box
  - Windows XP
  - Firefox + various addons
  - Sikuli
  - Number of custom scripts
- Generated PCAP files for post-analysis

Identifying MDNs

Three characteristics used to identify specific MDNs:

1) Each MDN identified had only one repository active at a given time

2) The repository URLs contain distinct patterns

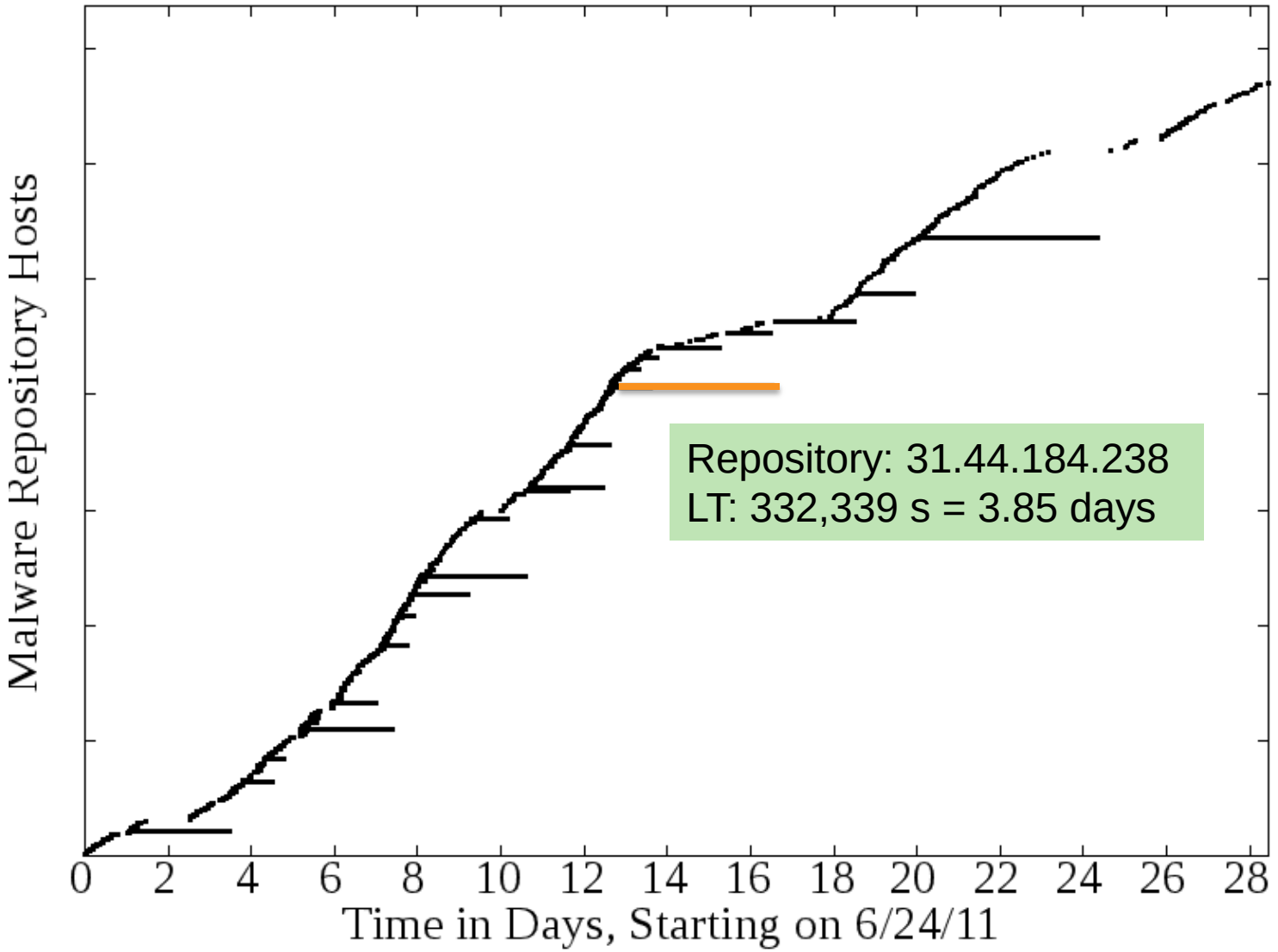
- e.g., /^http:\/\/www[0-9]\.[a-z]+\.rr\.nu\//

3) The injected code at the doorway is distinct for each MDN

Side Note…

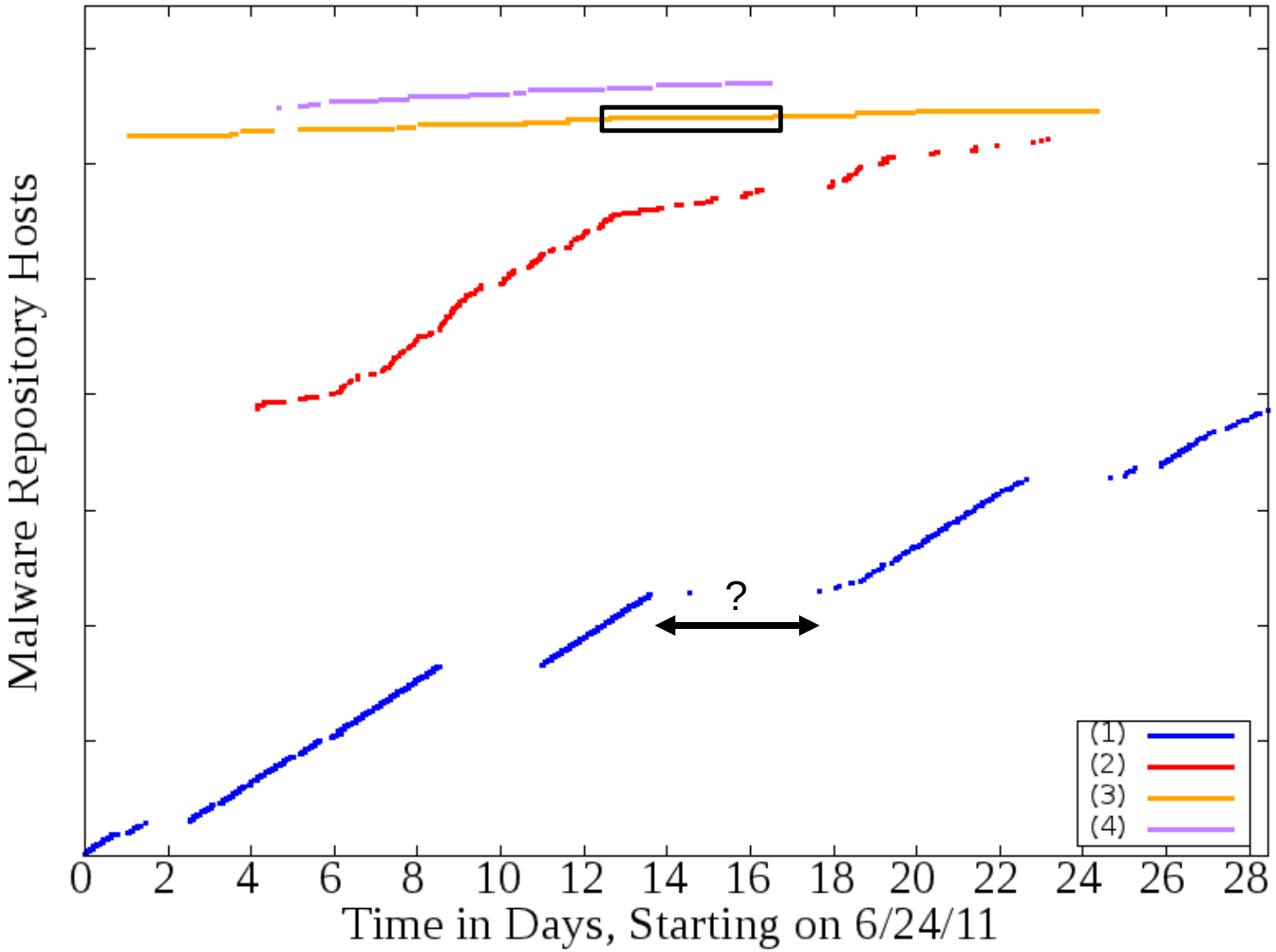
- Early days of black SEO, scripts used server side 302 redirects
- Lately they tend to use Javascript/Meta instead
- Why the change?
- Easy for researchers to fetch with a low interaction fetcher by spoofing referrer checks, need a “real” browser for Javascript handling

Ungrouped Repository Lifetimes



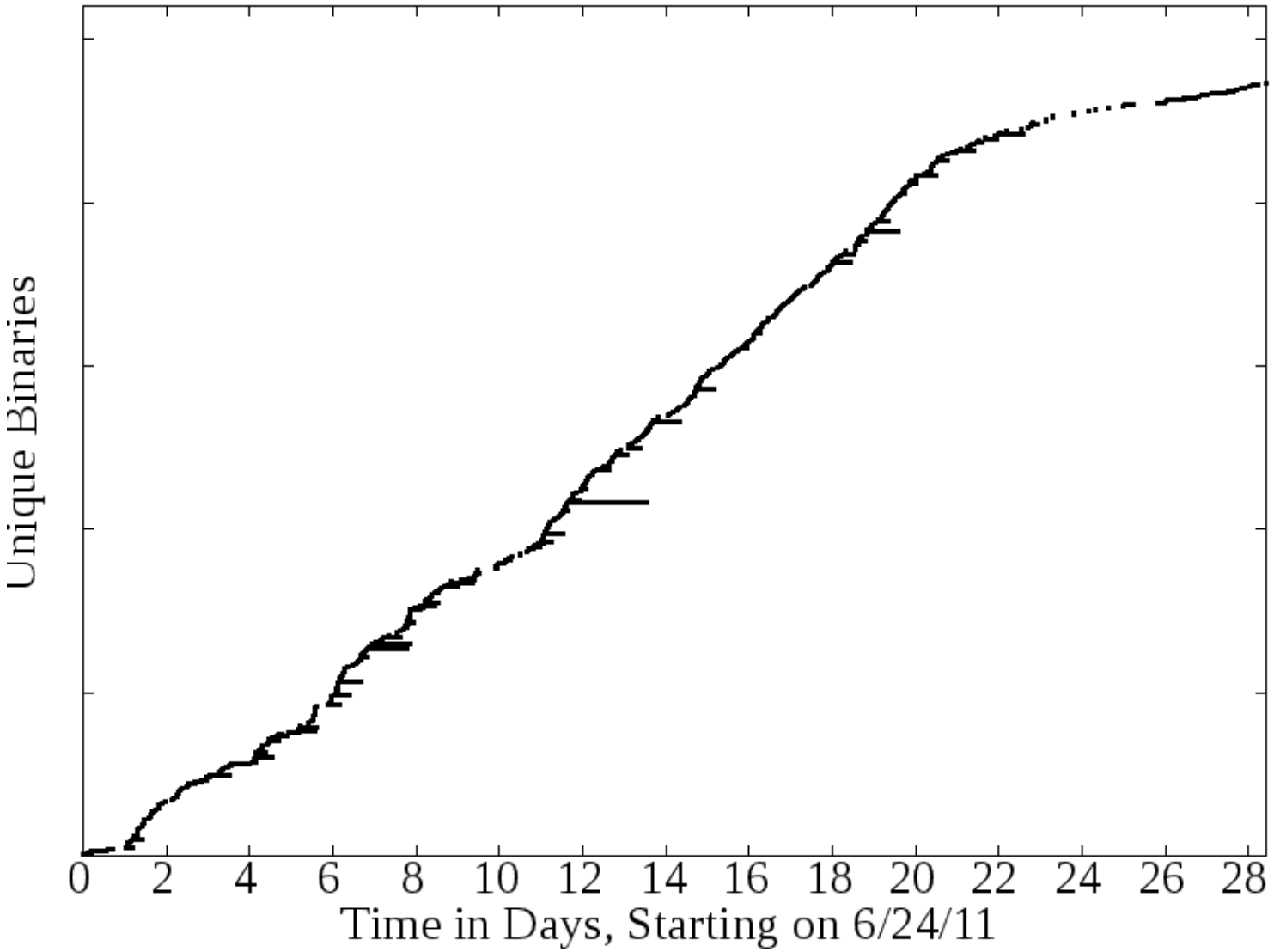
SOPHOS

Grouped Repository Lifetimes



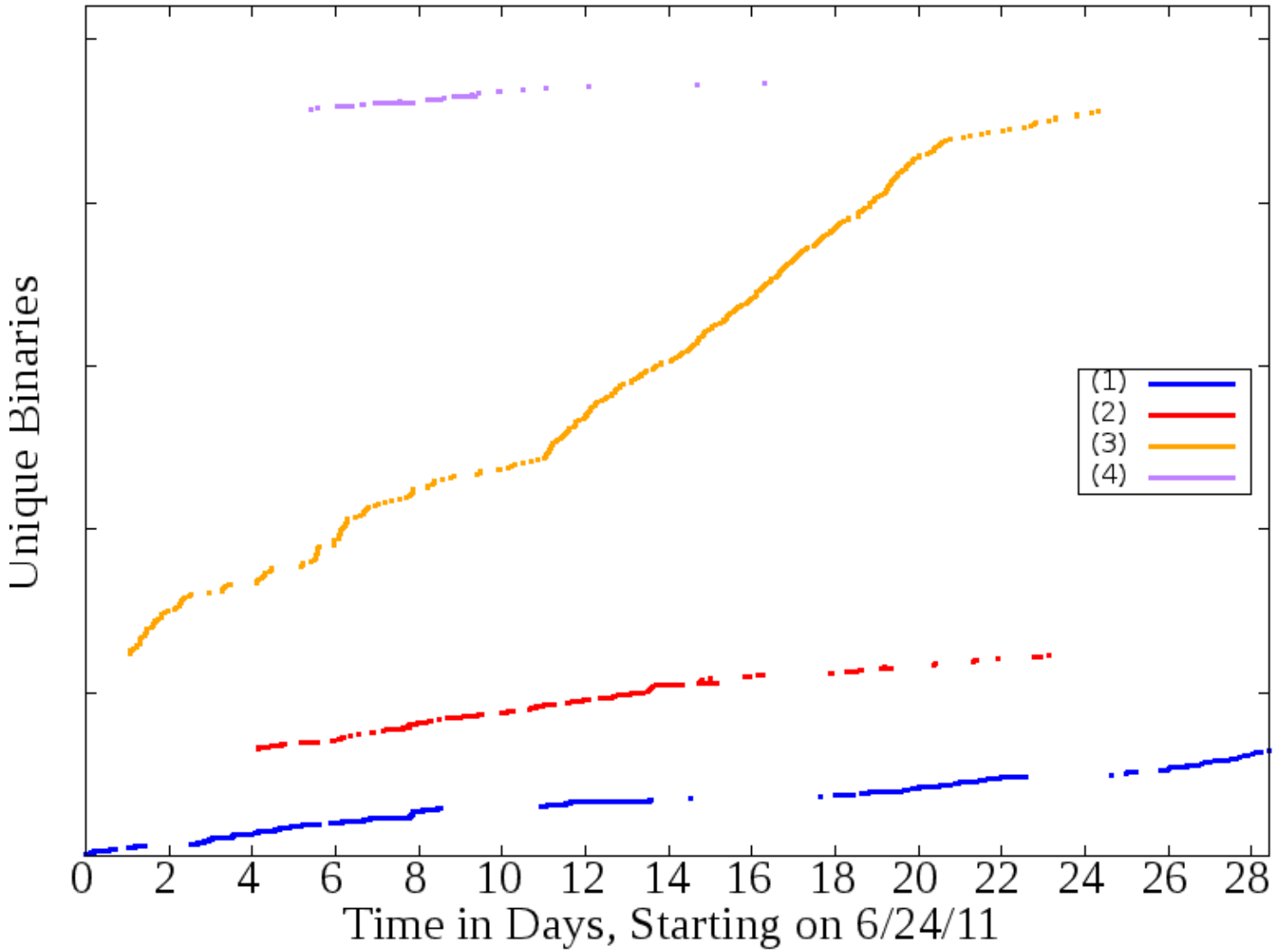
SOPHOS

Ungrouped Binary Lifetimes



SOPHOS

Grouped Binary Lifetimes



SOPHOS

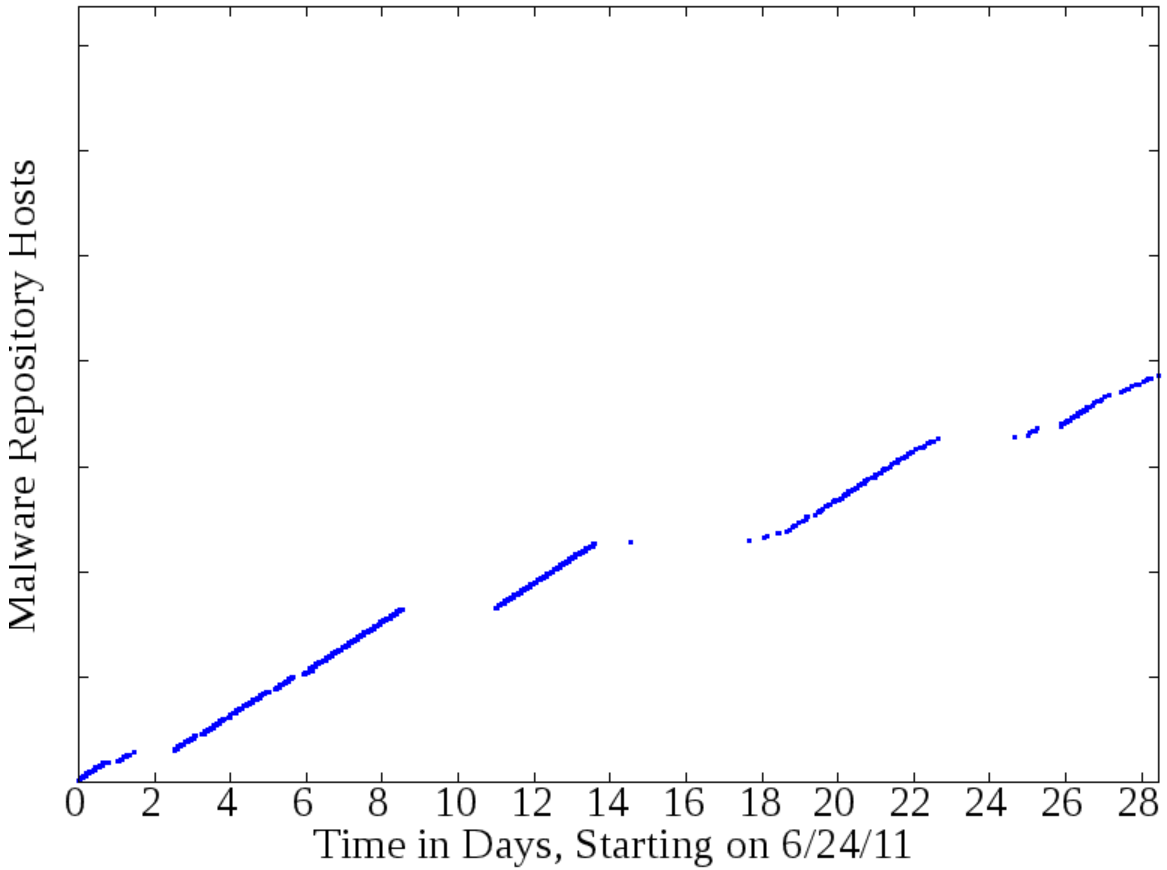
MDNs

MDN #	Mal Repo Details	Binary Update Behaviour	Blacklisting?
1	Randomly generated strings of .info TLD	Periodic updates	IP Blacklisting, redirection to non-malicious sites
2	Hosted on findhere.org, rr.nu	Periodic updates	None observed
3	Snowshoeing through multiple ranges of Ips	Fully polymorphic	Possible
4	Static base string incremented numbers appended, .info TLD	Periodic updates	None observed

Blacklisting

SOPHOS

IP Blacklisting



Gaps indicate no successful fetches of binaries were made during that time

SOPHOS

Some Kind Words

Fetch ID	755667
Fetch Time	2011-07-05 18:09
Experiment	high_interaction_best_protection_july5
Downloader	high_interaction_best_protection_july5-rotating_proxies
Client	pink
Get JSON	get raw json
Back to URL Summary	url summary

Structured Pages

200 - meltemdaysal.com/teaching/econ305/abnormal-qrs-complex&page=6
stream OC f780189342e607451591fbdd86a2597c3b8cbce8, detection:Mal/SEORed-A, filetype:ASCII text
Link via jsvar

200 - iozbireddest.info/fast-scan/
stream OC d5efd6bf836ae815bcb02066c6edc0cd0128055b, detection:Undetected, filetype:HTML
Link via img

200 - iozbireddest.info/fast-scan/img4/loading.gif
stream OC b3feb85f8b3587a591538c87a1596716f331e8e6, detection:Undetected, filetype:GIF

Link via iframe

302 - iozbireddest.info/fast-scan/download.php?q=av-sucks
stream OC adc83b19e793491b1c6ea[redacted]2fc, detection:Undetected, filetype:very short file (no magic)
Link via server_redirect

iozbireddest.info/fastantivirus2011.exe

SOPHOS

Not Quite Blacklisting

- On July 20th, domains from the same MDN started resolving to an IP belonging to Denis Sinegubko's unmaskparasites.com
- Site was not hacked, just trying to confuse us
- From July 8-19, 3.9 million requests were made to his server, have full server logs
- At one point people were being redirected to: *http://blog.unmaskbullshits.com/penis-sinegubko-was-found-shit-in-the-park-even-worst-than-previous-one-part-3213.html*





Possible Blacklisting

document.write("<img src='http://counter.yadro.ru/hit;JohnDeer?t52.6;r"+escape(document.referrer)+((typeof(screen)=="undefined")?"":";s"+screen.width+""+screen.height+"*"+(screen.colorDepth?screen.colorDepth:screen.pixelDepth))+";u"+escape(document.URL)+";"+Math.random()+"'"+"border='0' width='88' height='31'>");*

- JohnDeer perhaps a reference to John Deere Tractors used in harvesting fields
- Screen width and height when run in headless mode had values of 0

Optimizations

SOPHOS

Fake AV MDN Structure and Optimizations

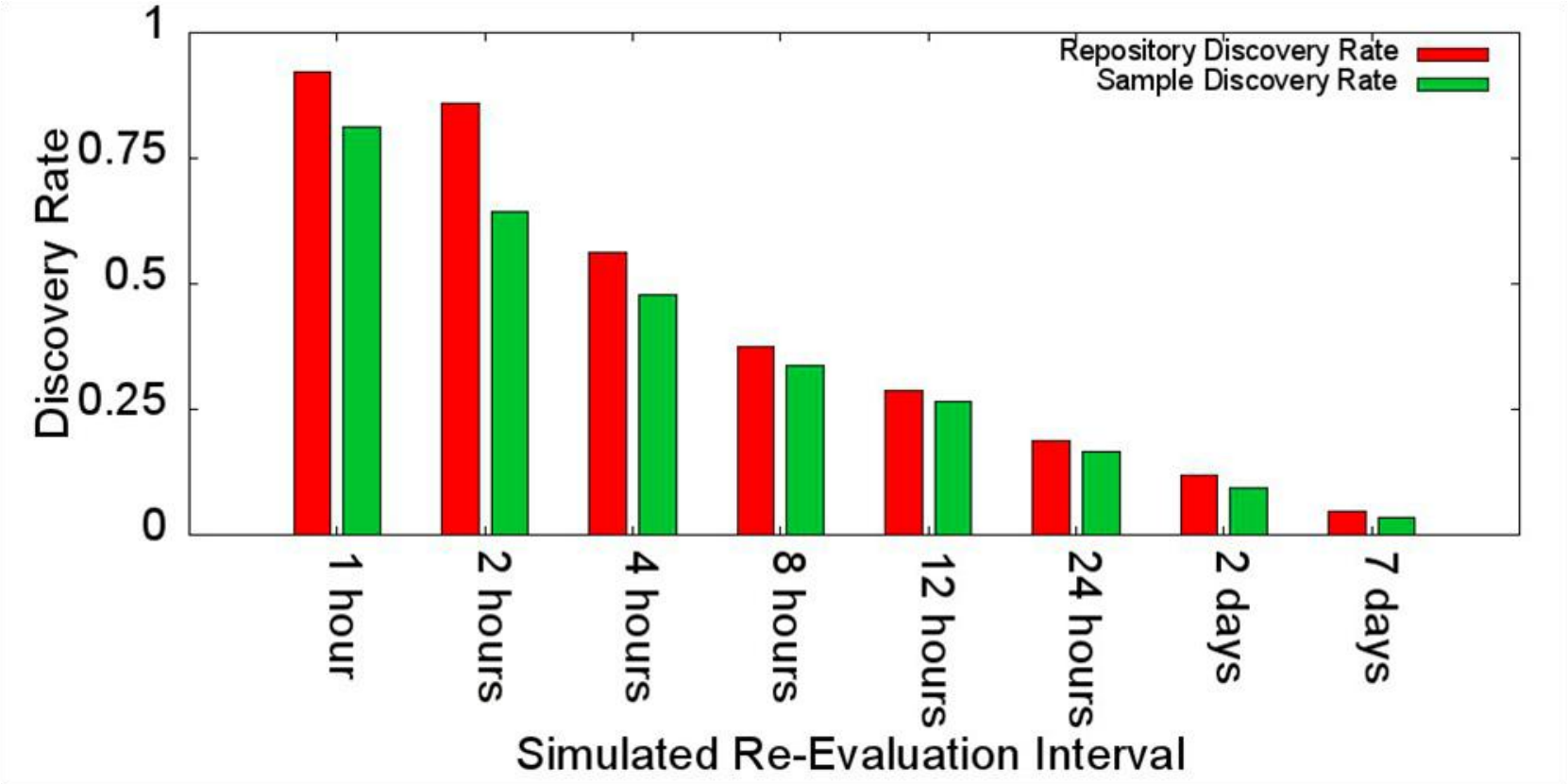
Two proposed optimizations:

1) Exploit high ’Fan In Factor’ between Landing Pages and Repositories in an MDN

2) Reduce exposures to repositories if we suspect the binary has not updated

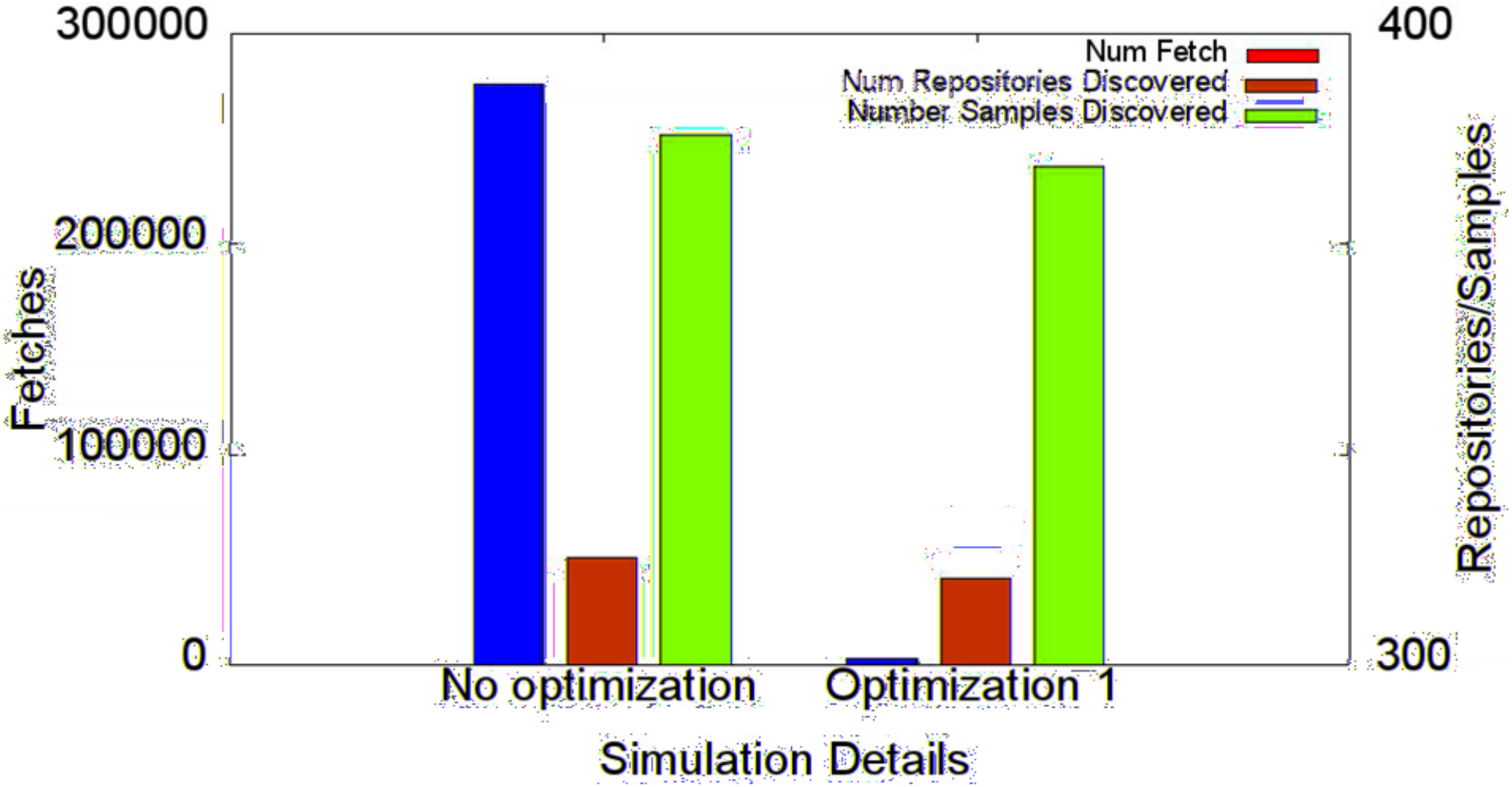
We performed a number of simulations to evaluate their effectiveness.

Re-Evaluation Interval vs Sample/Network Discovery



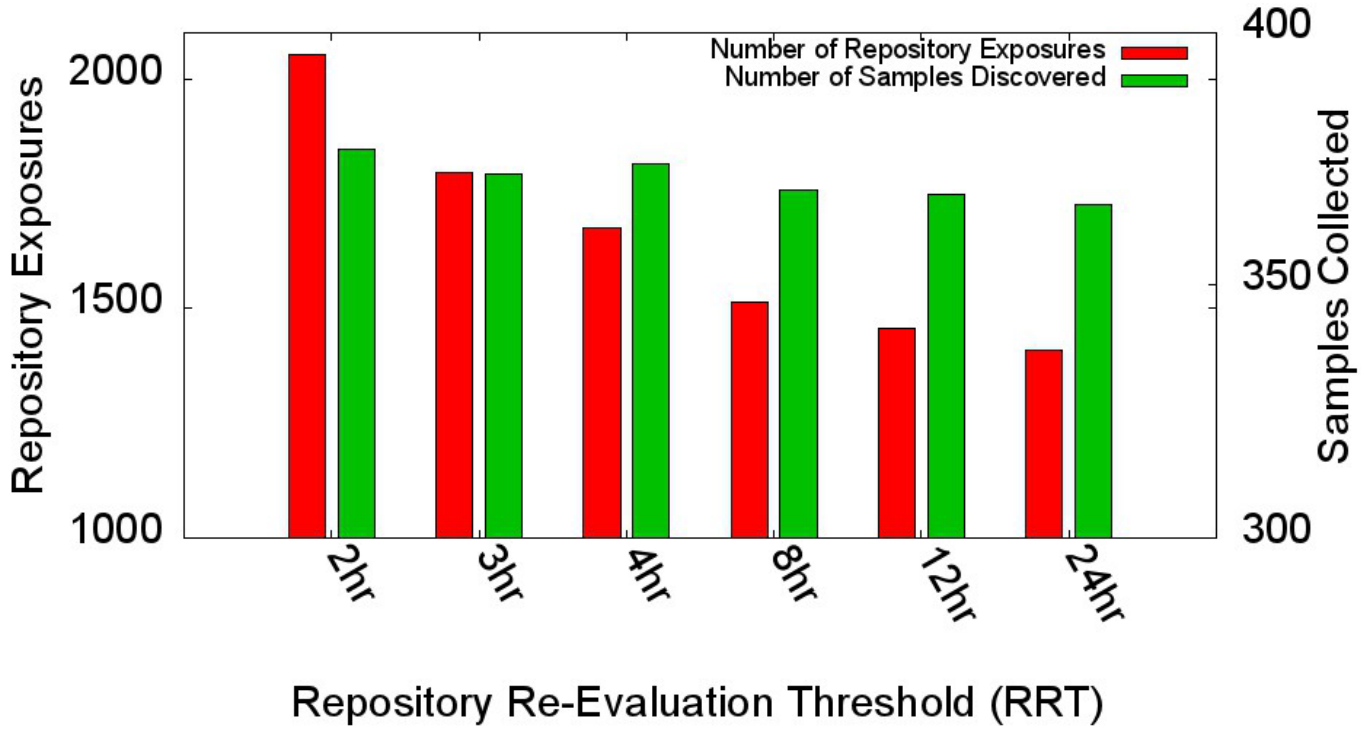
Sample and network discovery rates quickly decline as the re-evaluation interval increases

Impact of First Optimization



Drastic reduction in fetch volume at small discovery cost

Impact of Second Optimization

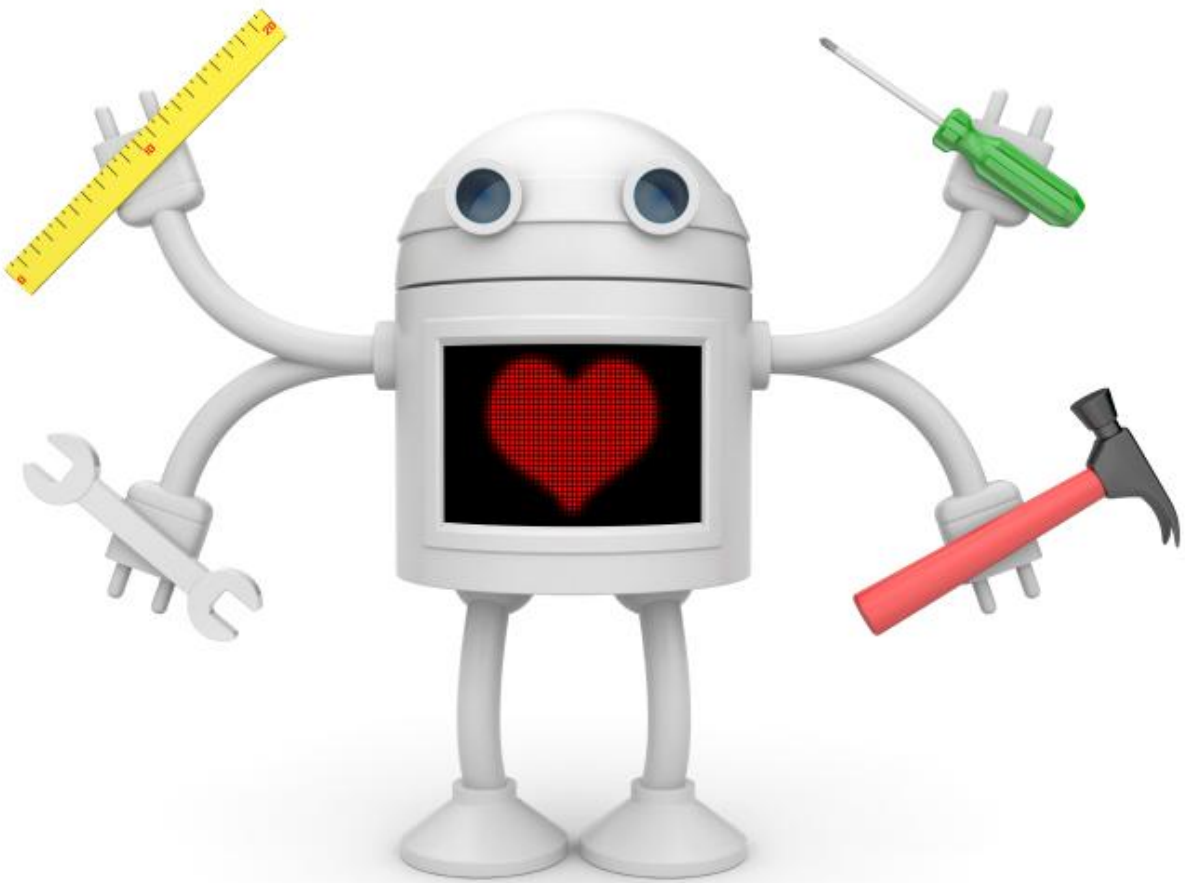


- RRT = Do not revisit repository more frequently than X
- Further reduction in fetch volume, with minor cost to sample discovery rate
- This optimization does not work well with fully polymorphic MDNs

SOPHOS

Conclusions

- Blacklisting is a valid concern when developing automated systems to monitor MDNs
- URLs should be grouped into MDNs and a re-evaluation strategy should be applied wherever possible to reduce resource requirements and chance of blacklisting
- Grouping binary samples by MDN is also an effective strategy when writing detection
- Using knowledge of MDNs and their lifetimes allows you to prioritize which samples require immediate attention versus which can potentially use a quick checksum detection



Thank you!