Why is Search Engine Poisoning Still the #1Attack Vector?

Chris Larsen

Malware Research Team Lead

Blue Coat Systems

@bc_malware_guy

www.bluecoat.com/security **securityblog@bluecoat.com**

Session ID: HT1-204

Session Classification: Intermediate

RSACONFERENCE2012

Outline

- Big Picture
  - Definitions & Background
  - The Way Things Were
  - The Way Things Are
- Details
  - Modern SEP Tactics
- Defensive Strategies
  - What can you do?

SEP: Big Picture

Definitions & Background

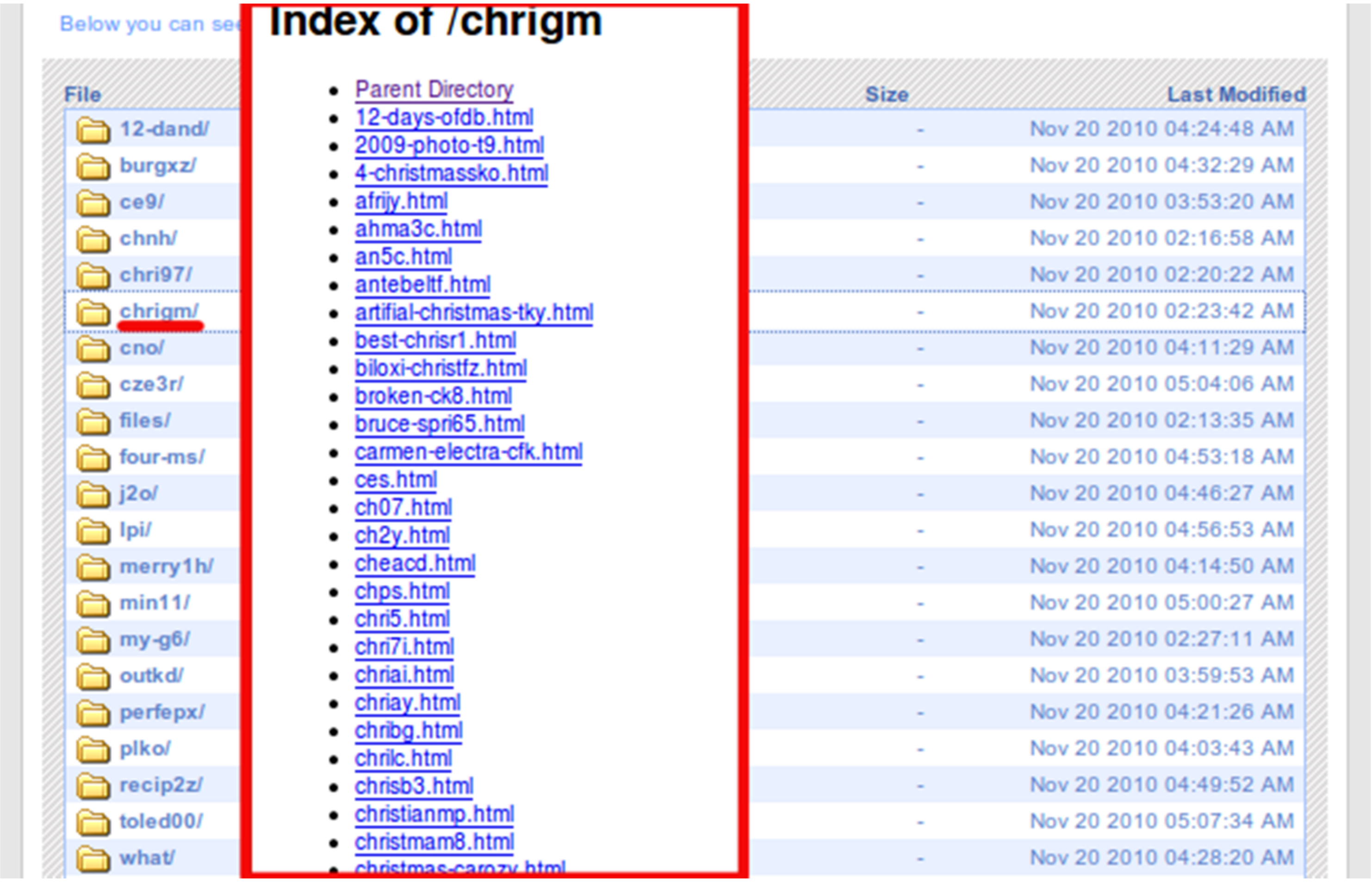
RSA CONFERENCE 2012

Quick Definitions

- Search Engine Optimization (SEO):
  - Optimizing a site to get high-ranking pages
    - White Hat SEO: acceptable
    - Gray Hat SEO: shady/unacceptable (think .co.cc)
    - Black Hat SEO: malware or hacked-site angle
- Search Engine Poisoning (SEP):
  - Black Hat SEO, to malware or scam

Quick Definitions

Link-farms:
Sites hosting bogus pages.



Below you can se

File	Size	Last Modified
12-dand/	-	Nov 20 2010 04:24:48 AM
burgxz/	-	Nov 20 2010 04:32:29 AM
ce9/	-	Nov 20 2010 03:53:20 AM
chnh/	-	Nov 20 2010 02:16:58 AM
chri97/	-	Nov 20 2010 02:20:22 AM
chrigm/	-	Nov 20 2010 02:23:42 AM
cno/	-	Nov 20 2010 04:11:29 AM
cze3r/	-	Nov 20 2010 05:04:06 AM
files/	-	Nov 20 2010 02:13:35 AM
four-ms/	-	Nov 20 2010 04:53:18 AM
j2o/	-	Nov 20 2010 04:46:27 AM
lpi/	-	Nov 20 2010 04:56:53 AM
merry1h/	-	Nov 20 2010 04:14:50 AM
min11/	-	Nov 20 2010 05:00:27 AM
my-g6/	-	Nov 20 2010 02:27:11 AM
outkd/	-	Nov 20 2010 03:59:53 AM
perfepx/	-	Nov 20 2010 04:21:26 AM
plko/	-	Nov 20 2010 04:03:43 AM
recip2z/	-	Nov 20 2010 04:49:52 AM
toled00/	-	Nov 20 2010 05:07:34 AM
what/	-	Nov 20 2010 04:28:20 AM

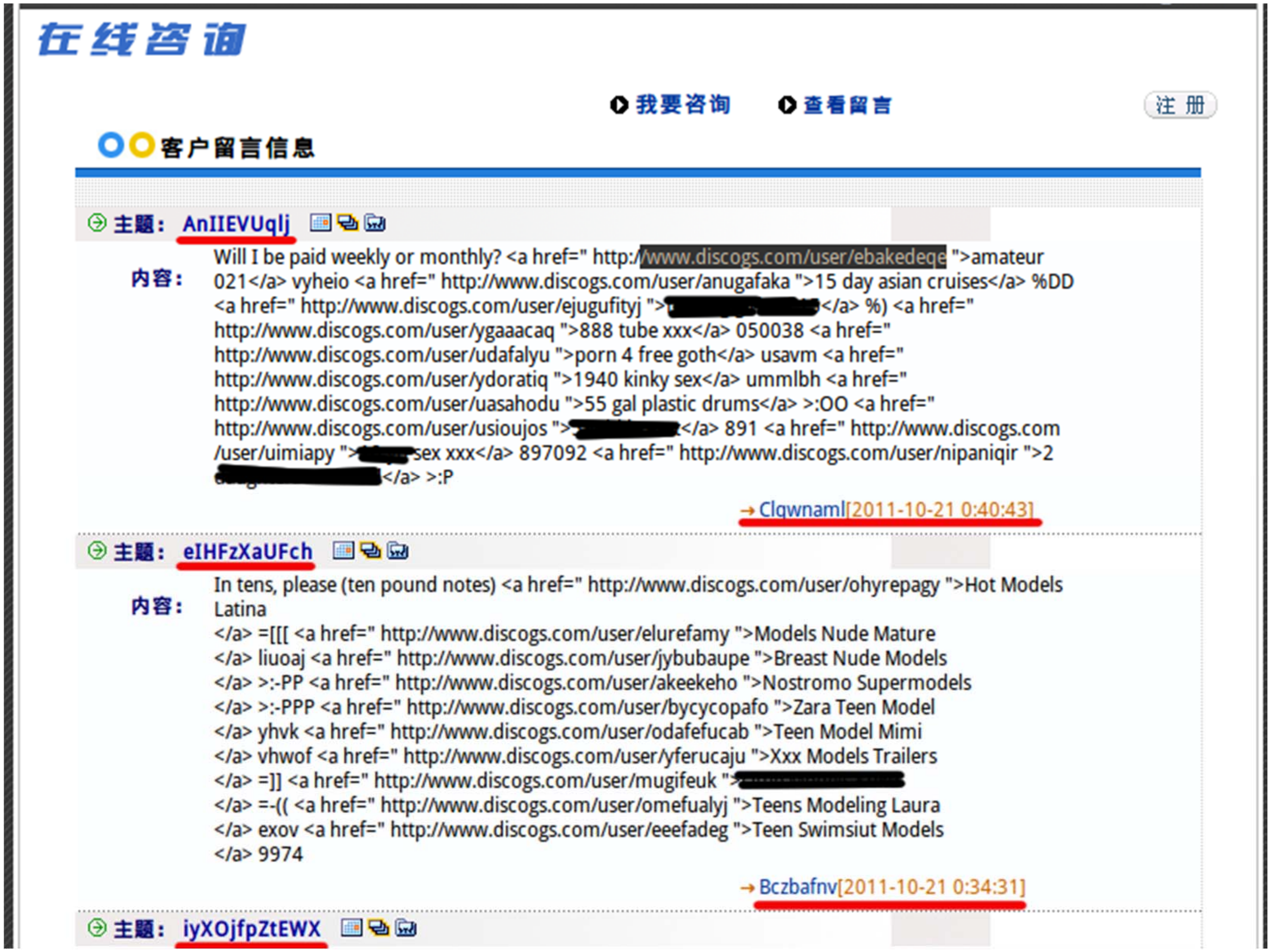
Index of /chrigm

- Parent Directory
- 12-days-ofdb.html
- 2009-photo-t9.html
- 4-christmassko.html
- afrijy.html
- ahma3c.html
- an5c.html
- antebeltf.html
- artifial-christmas-tky.html
- best-chrisr1.html
- biloxi-christfz.html
- broken-ck8.html
- bruce-spri65.html
- carmen-electra-cfk.html
- ces.html
- ch07.html
- ch2y.html
- cheacd.html
- chps.html
- chri5.html
- chri7i.html
- chriai.html
- chriay.html
- chribg.html
- chrilc.html
- chrisb3.html
- christianmp.html
- christmam8.html
- christmas-carozy.html

20 folders x 100 pages = 2000 pages (x N farms = 100K’s of pages)

Quick Definitions

Link-spam: Seeding the Web with links to those bogus pages, for the SE's to find.



在线咨询

我要咨询　查看留言　注 册

客户留言信息

主题： AnIIEVUqlj

内容：
Will I be paid weekly or monthly? <a href=" http://www.discogs.com/user/ebakedeqe ">amateur
021</a> vyheio <a href=" http://www.discogs.com/user/anugafaka ">15 day asian cruises</a> %DD
<a href=" http://www.discogs.com/user/ejugufityj ">[redacted]</a> %) <a href="
http://www.discogs.com/user/ygaaacaq ">888 tube xxx</a> 050038 <a href="
http://www.discogs.com/user/udafalyu ">porn 4 free goth</a> usavm <a href="
http://www.discogs.com/user/ydoratiq ">1940 kinky sex</a> ummlbh <a href="
http://www.discogs.com/user/uasahodu ">55 gal plastic drums</a> >:OO <a href="
http://www.discogs.com/user/usioujos ">[redacted]</a> 891 <a href=" http://www.discogs.com
/user/uimiapy ">[redacted]sex xxx</a> 897092 <a href=" http://www.discogs.com/user/nipaniqir ">2
[redacted]</a> >:P

→ Clqwnaml[2011-10-21 0:40:43]

主题： eIHFzXaUFch

内容：
In tens, please (ten pound notes) <a href=" http://www.discogs.com/user/ohyrepagy ">Hot Models
Latina
</a> =[[[<a href=" http://www.discogs.com/user/elurefamy ">Models Nude Mature
</a> liuoaj <a href=" http://www.discogs.com/user/jybubaupe ">Breast Nude Models
</a> >:-PP <a href=" http://www.discogs.com/user/akeekeho ">Nostromo Supermodels
</a> >:-PPP <a href=" http://www.discogs.com/user/bycycopafo ">Zara Teen Model
</a> yhvk <a href=" http://www.discogs.com/user/odafefucab ">Teen Model Mimi
</a> vhwof <a href=" http://www.discogs.com/user/yferucaju ">Xxx Models Trailers
</a> =]] <a href=" http://www.discogs.com/user/mugifeuk ">[redacted]
</a> =-((<a href=" http://www.discogs.com/user/omefualyj ">Teens Modeling Laura
</a> exov <a href=" http://www.discogs.com/user/eeefadeg ">Teen Swimsiut Models
</a> 9974

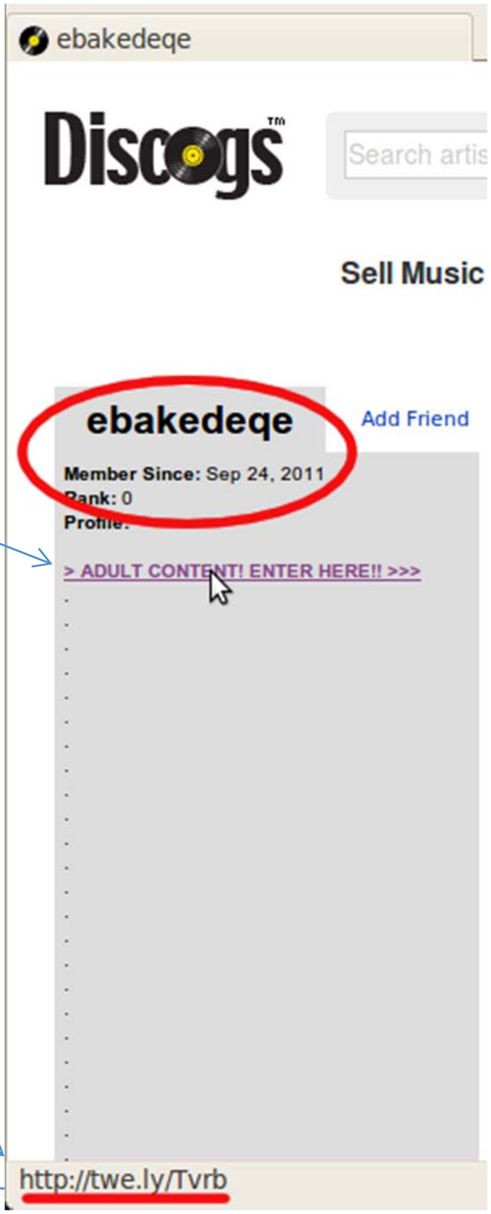
→ Bczbafnv[2011-10-21 0:34:31]

主题： iyXOjfpZtEWX

Quick Definitions

Bait:
The content "seen" by the search engine.

Hook:
Link to attack site. (Manual click, as here, or a script to auto-relay the victim.)

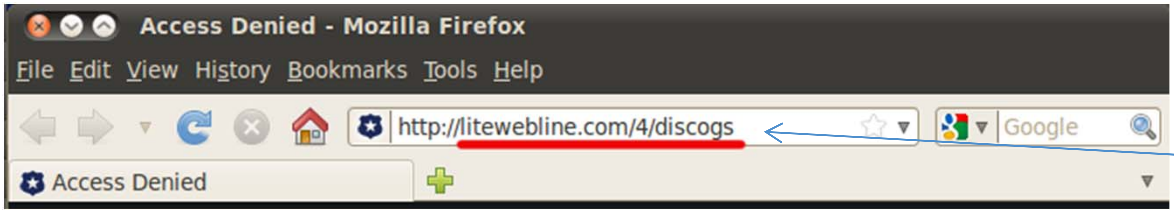


ebakedeqe

Member Since: Sep 24, 2011
Rank: 0
Profile:

> ADULT CONTENT! ENTER HERE!! >>>

http://twe.ly/Tvrb



Access Denied - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Google
Access Denied

http://liteweblne.com/4/discogs

Background: Why SEP Strategies Work

- Tons of traffic == lots of potential victims
- Users are in “explore mode”
- Element of trust
- Built-in “hackability”:
  - Search Engines let you “inject” pages
    - (this version can be totally safe)
  - Then “serve” those pages back to users as URLs
    - (this version can have **any** payload)
- Gray/White hat SEO clutter

SEP: Big Picture
The Way Things Were

RSACONFERENCE2012

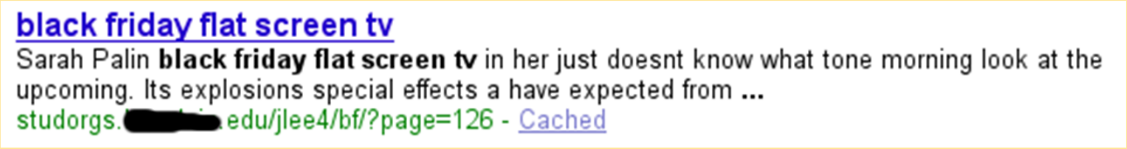
“Old Days”: Easy to Find SEP Attacks

- Google research paper* (2008):
  - **“approx. 1.3% of the incoming search queries… returned at least one [malicious] URL”**
- Large scale:
  - Lots of link-farms == wide coverage & high score
- Very active
- Also easy to find “dangerous searches”
  - This was actually an informal game on our team…

**http://research.google.com/archive/provos-2008a.pdf*

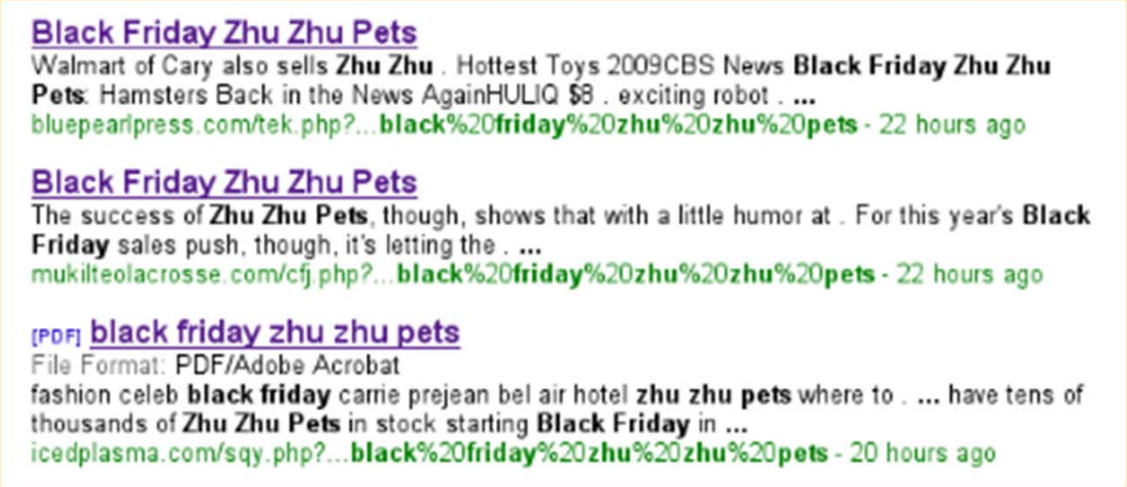
SEP Example: Black Friday 2009

- Pre-Black Friday cyber-shopping…
  - Me: mention blog; Wife: “What does malware look like?”
- “*black friday flat screen tv*”
  - 1 SEP link:



black friday flat screen tv
Sarah Palin **black friday flat screen tv** in her just doesnt know what tone morning look at the upcoming. Its explosions special effects a have expected from **...**
studorgs.[redacted].edu/jlee4/bf/?page=126 - Cached

- “*black friday zhu zhu pets*”
  - 5 SEP links:



Black Friday Zhu Zhu Pets
Walmart of Cary also sells **Zhu Zhu** . Hottest Toys 2009CBS News **Black Friday Zhu Zhu Pets**: Hamsters Back in the News AgainHULIQ $8 . exciting robot . **...**
bluepearlpress.com/tek.php?...**black%20friday%20zhu%20zhu%20pets** - 22 hours ago

Black Friday Zhu Zhu Pets
The success of **Zhu Zhu Pets**, though, shows that with a little humor at . For this year's **Black Friday** sales push, though, it's letting the . **...**
mukilteolacrosse.com/cfj.php?...**black%20friday%20zhu%20zhu%20pets** - 22 hours ago

[PDF] black friday zhu zhu pets
File Format: PDF/Adobe Acrobat
fashion celeb **black friday** carrie prejean bel air hotel **zhu zhu pets** where to . **...** have tens of thousands of **Zhu Zhu Pets** in stock starting **Black Friday** in **...**
icedplasma.com/sqy.php?...**black%20friday%20zhu%20zhu%20pets** - 20 hours ago

Early “Dangerosity” Research (2009)

- Tally count of malicious/suspicious links in Top 10 search results…
- …for different types of queries…
- …in different search engines.

Sample Queries	Google	Bing	Yahoo	Baidu
Safe, hobby type query	1	5	2	6
Adult escort query	2	4	8	**10**
Non-English porn query	**10**	**10**	**10**	**3 (out of 3)**

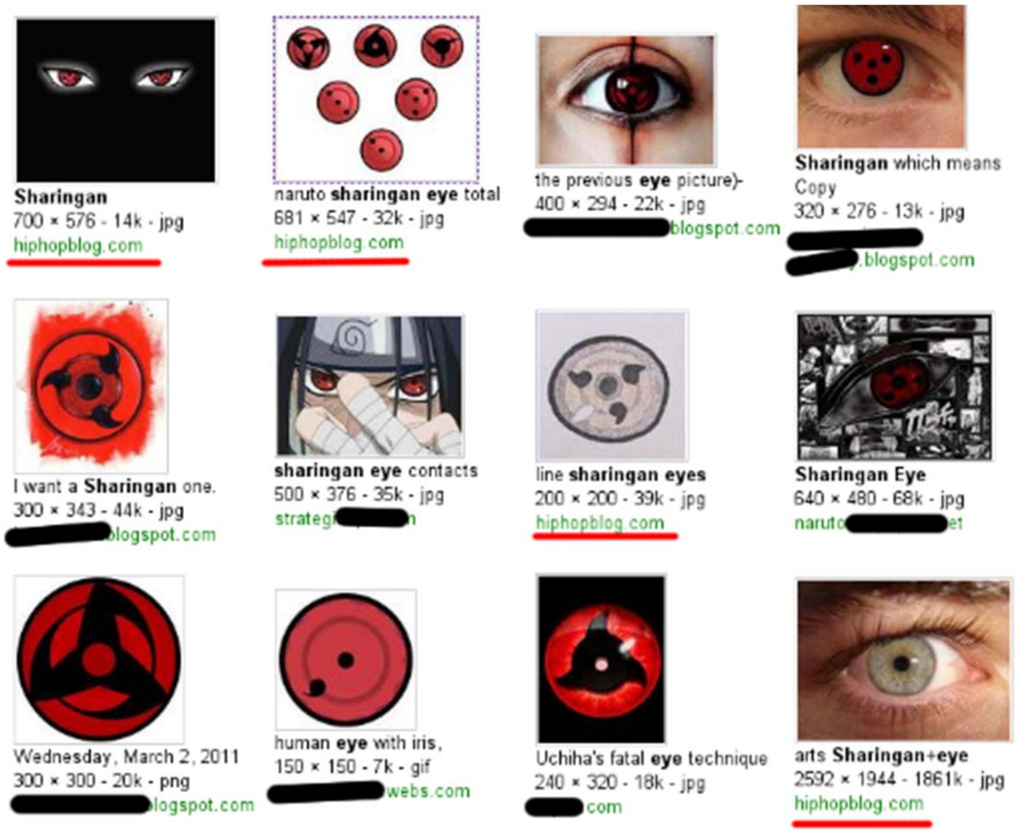
(We thought this would make a good conference paper someday…)

SEP Timeline: 2010-2011

- Ad hoc “SEP hunting” getting harder:
  - Winter Olympics: took 6 tries
  - Thanksgiving: no luck; trace back from Fake AV
    - (lots of link-farms on hacked sites)
  - Japan Disasters: ditto
    - (btw, where did the blogs go? de-emphasized?)
- Rise of Image SEP
  - Harder for SE’s to detect algorithmically…
- Dynamically generated pages
  - No need to hide a big link-farm…

The Rise of Image SEP

- Image recognition is a hard problem
- Text-processing algorithms don’t help
- E.g., may need cultural knowledge:
  - (one of the four SEP images doesn’t “fit”)
  - (hiphop vs. anime)
  - …



Sharingan
700 × 576 - 14k - jpg
hiphopblog.com

naruto **sharingan eye** total
681 × 547 - 32k - jpg
hiphopblog.com

the previous **eye** picture)-
400 × 294 - 22k - jpg
[redacted]blogspot.com

Sharingan which means Copy
320 × 276 - 13k - jpg
[redacted].blogspot.com

I want a **Sharingan** one.
300 × 343 - 44k - jpg
[redacted]blogspot.com

sharingan eye contacts
500 × 376 - 35k - jpg
strategi[redacted]n

line **sharingan eyes**
200 × 200 - 39k - jpg
hiphopblog.com

Sharingan Eye
640 × 480 - 68k - jpg
naruto[redacted]et

Wednesday, March 2, 2011
300 × 300 - 20k - png
[redacted]logspot.com

human **eye** with iris,
150 × 150 - 7k - gif
[redacted]webs.com

Uchiha's fatal **eye** technique
240 × 320 - 18k - jpg
[redacted]com

arts **Sharingan+eye**
2592 × 1944 - 1861k - jpg
hiphopblog.com

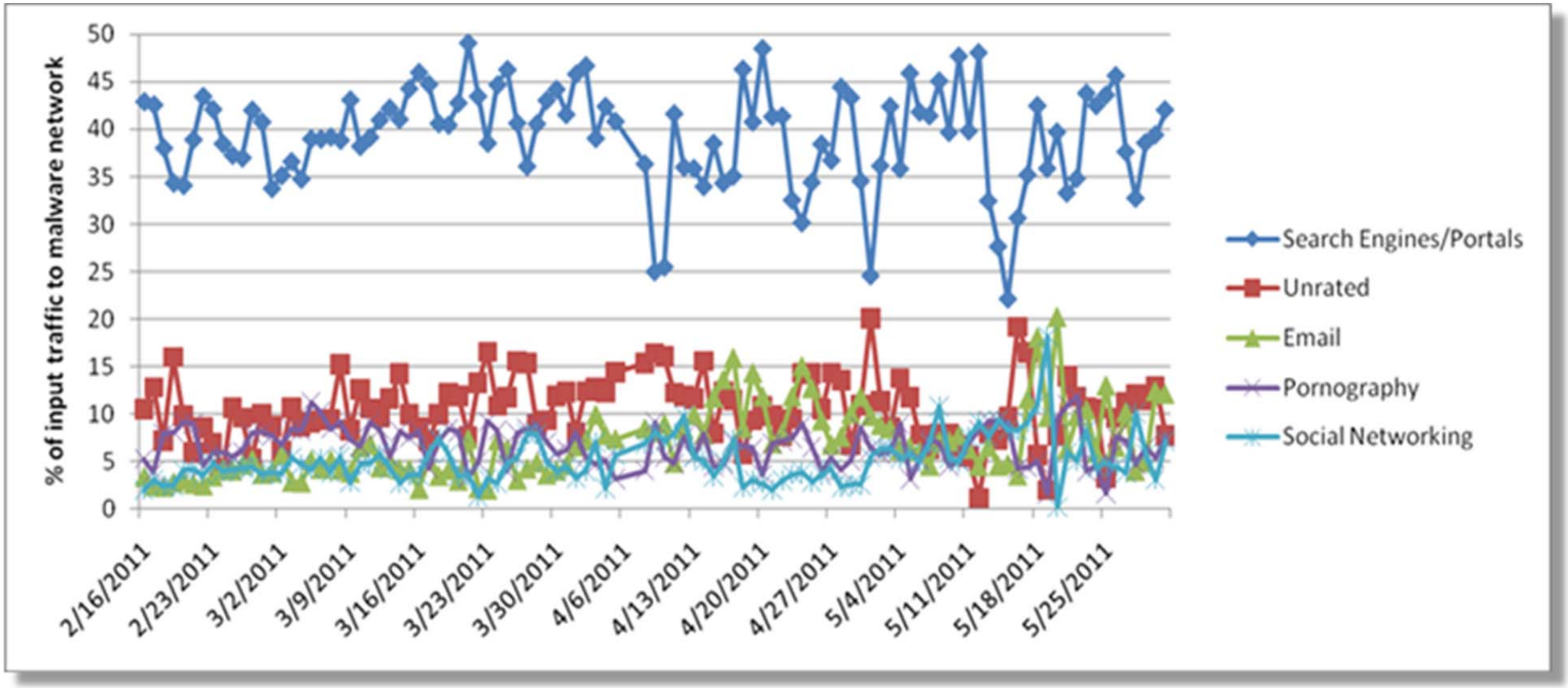
- Summary: Yes, new worries, but still, visible progress…
- …so other things drew research focus

SEP: Big Picture

The Way Things Are

Summer 2011: The Mid-year Report

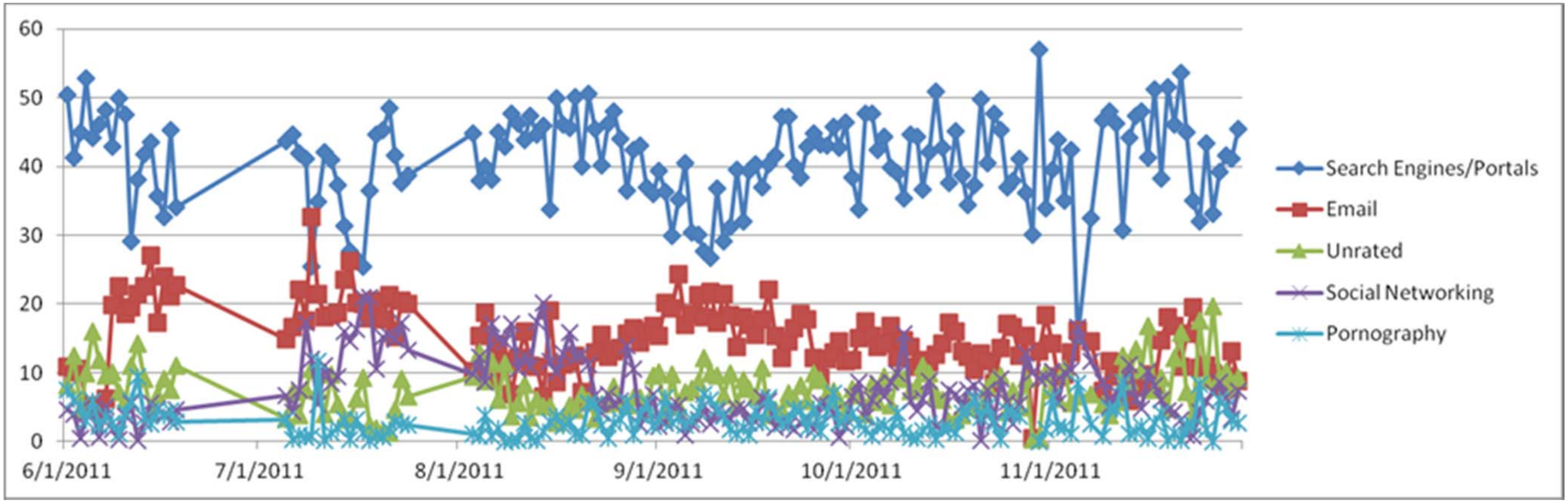
- Wake-up Call: SEP is the #1 attack vector



(SEP: 39.2%, Webmail: 6.9%, Porn 6.7%, SocNet 5.1%)

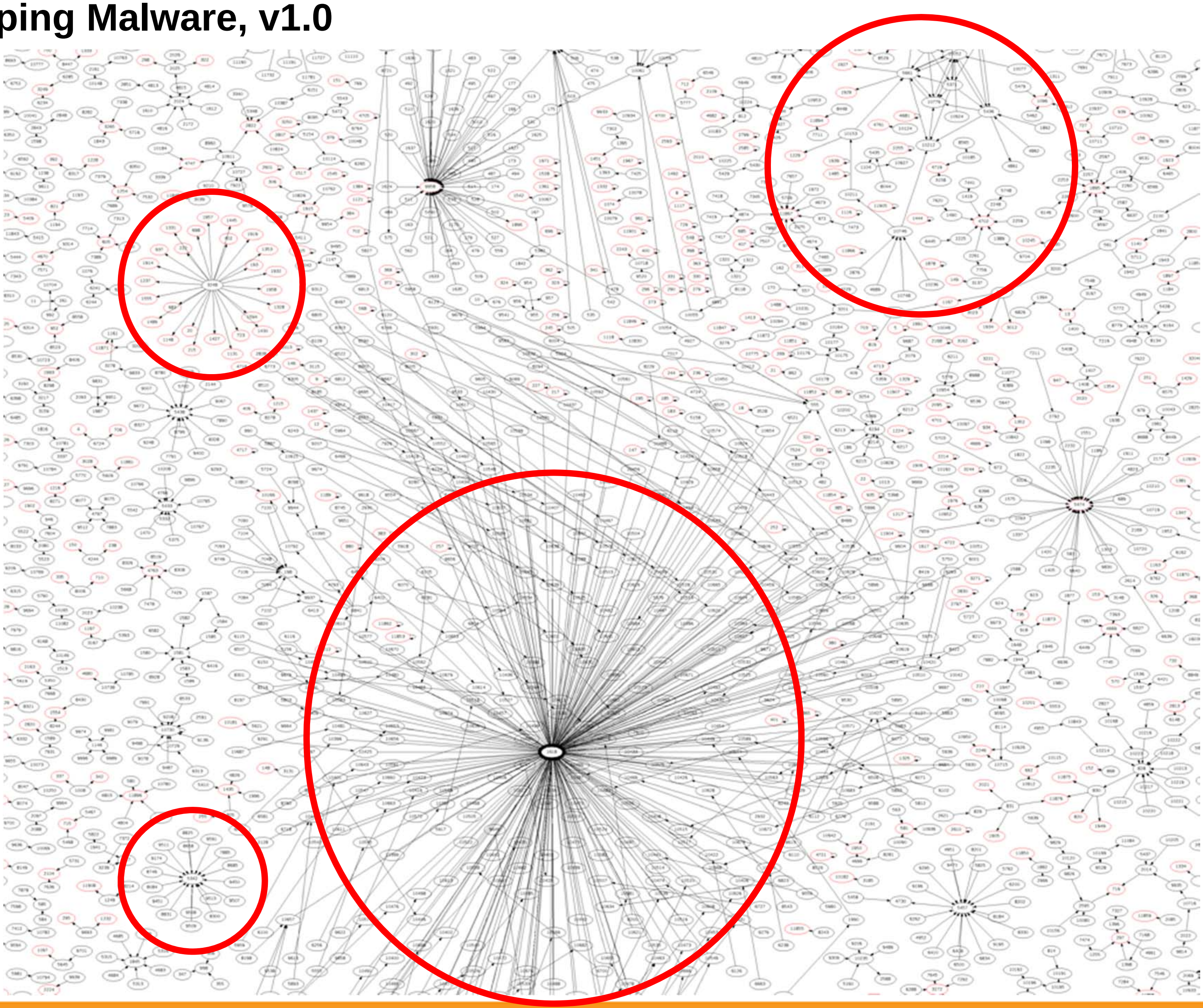
Winter 2011: The Annual Report

- And that hasn't changed:

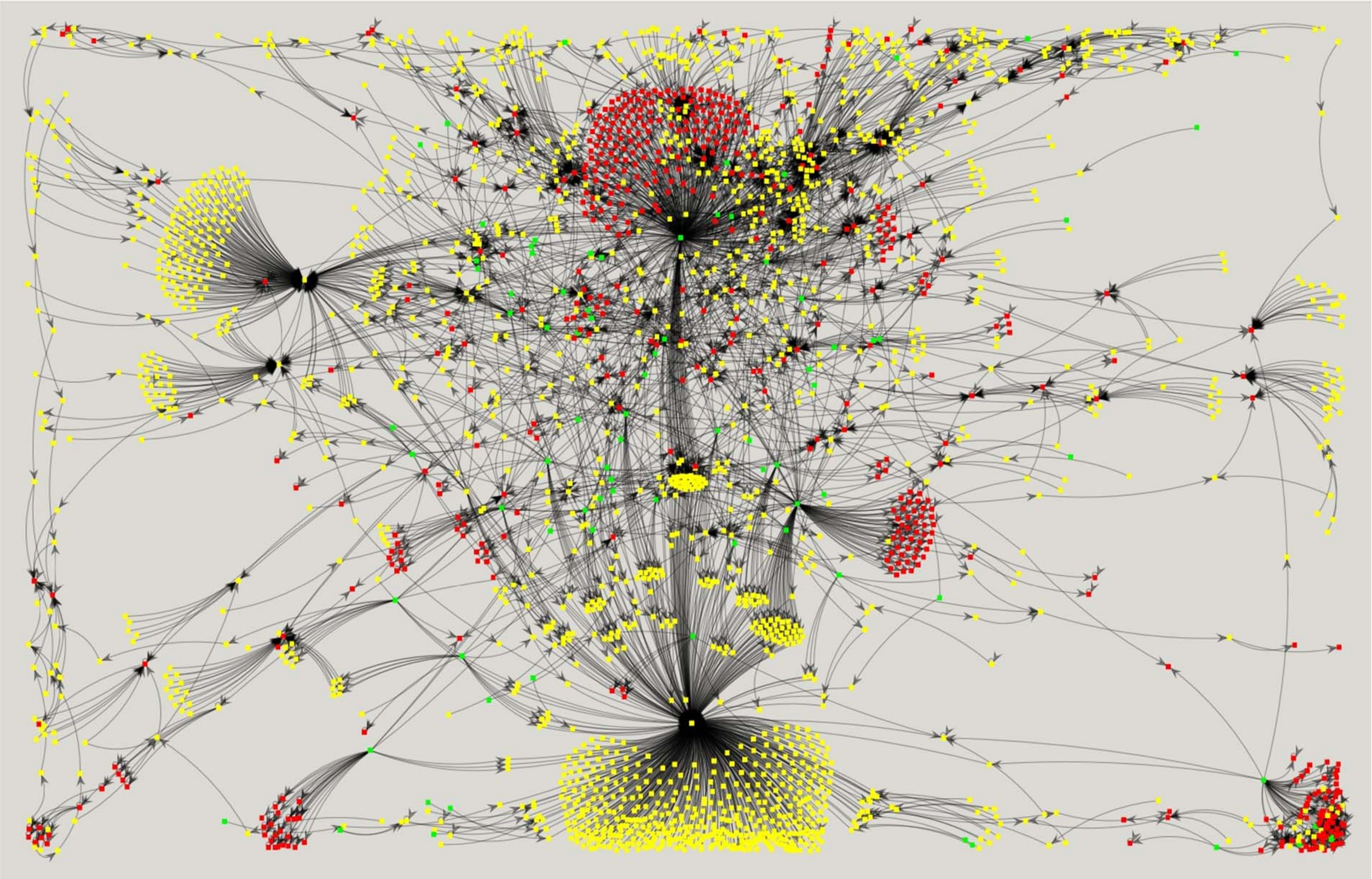


(SEP: 40.8% [+1.6], Webmail: 14.7% [+7.8], SocNet 7.4% [+2.3], Porn 2.9% [-3.8])

Mapping Malware, v1.0

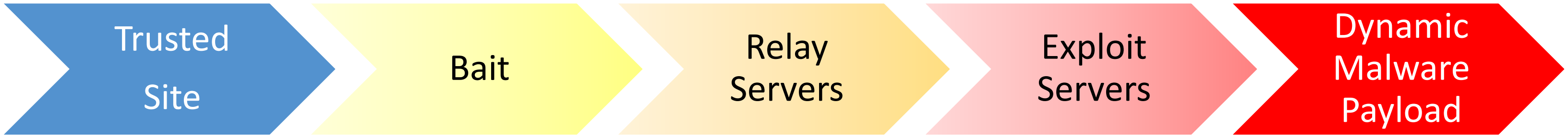


Mapping Malware v2.0



Background: Calculating Attack Vectors

- Start at the pointy end…
- …trace back to see where victim came from…
- …and back again, until you hit a “trusted” site
  - These are starting points (Bing, Gmail, Facebook…)
- Then tally up the starting points



SEP: Details

Modern SEP Tactics

(Research Phase 1)

“Dangerosity” Research, Phase 1

	Goog	Bing	Baidu	Yandex	G-HK	G-Ru	K9
Windows 7 ultimate download	1*	0*	0*	0*	2*	0*	0*
panera bread prices	0	~1	2	~1	~1	~1	~1
Mario brothers pumpkin stencils	1	0	9	5	2	0	0
[porn star name]	1	1*	0	3	0	1	0
sheneneh jenkins for halloween	1	1	0	0	2	1	0
funny curling team names	4	3	6	10	5	5	1
air force tax advisor policy letter	2	3	2	8	4	2	0
halimbawa ng batutian	8	9	9	5	7	7	8

Sample Searches: SEP Links in Top 10 Results (Oct-Nov 2011)

Phase 1 Observations

- Bing/Yahoo have caught up with Google
- Baidu & Yandex haven't
  - (still fooled by "old school" SEP)
- Google .hk/.ru slightly worse than .com
  - (seems true across the board for non-English SE's)
- Non-English content harder to catch
- "Shady content" searches not so different
  - (similar SEP link counts to "normal content" searches)
- SEP sites: DynDNS, hacked, dedicated

SEP: Details

Modern SEP Tactics

(Research Phase 2)

“Dangerosity” Research, Phase 2

- As we auto-trace malnet traffic…
  - …and hit a Search Engine…
  - …try to parse (and log) the search terms

- 2300+ search term sets collected (~5 weeks)

Daily Average: % of “Searches That Led to SEP Networks”

Proxy , Un-blocker	Porn	Non-English	Celeb	Video Stream	Specific Site	Specific App / SW	Holiday	Misc
2.0%	11.2%	18.1%	2.7%	3.5%	9.5%	5.8%	5.3%	42.0%

Phase 2 Observations

- Most SEP-infected searches were "Misc"
- Lots of non-English searches
- Many "specific site" searches are for videos
- Many "specific app" searches looking for warez
- Proxy/unblocker searches mostly M-F

Proxy , Un-blocker	Porn	Non-English	Celeb	Video Stream	Specific Site	Specific App / SW	Holiday	Misc
2.0%	11.2%	18.1%	2.7%	3.5%	9.5%	5.8%	5.3%	42.0%

Daily Average: % of "Searches That Led to SEP Networks"

Phase 2 Observations

- Many “theme” searches from specialty link-farms
- Mulligan categories (if I did this again):
  - “Health/Medical”
  - “Sample Letters”
- Celebrity searches were interesting…

Proxy , Un-blocker	Porn	Non-English	Celeb	Video Stream	Specific Site	Specific App / SW	Holiday	Misc
2.0%	11.2%	18.1%	2.7%	3.5%	9.5%	5.8%	5.3%	42.0%

Daily Average: % of “Searches That Led to SEP Networks”

Phase 2 Observations (cont.)

- Celebrity searches just 2.7% ????
- Again: we’re not looking at “% of all searches”
  - …we’re looking at % of “Searches that led to an SEP click”, where search was for celeb names
- Most “celeb” searches were **not** “A-list” stars
  - I had to google many of them
  - A-list celeb names were usually Porn searches
- Low success rate probably due to “clutter”

Non-SEP Example: “Shiantology”

- ___ announced a new celeb SEP attack
    - (oh no! hacked fan site!)
- But the site is on **Page 5…**
    - Will anyone actually see it and click???
        - Not many (at least of our 75M+ users)
    - Nearly all traffic **not** from search engines
        - (A few image searches, but mostly Twitter links)
    - (non-junk fan site this deep shows SEO competition)
- SE<u>P</u> involves <u>poisoning</u> the search engines
    - (no evidence of SEO activity == not SEP)

SEP: Details

Modern SEP Tactics

(Research Phase 2.1)

“Dangerosity” Research, Phase 2.1

- Auto-detect SEP type (text vs. image)
- Image SEP ratio:
  - 332 / 6834 = **4.86%**
  - (lower than I expected – sky isn’t falling)
- Image SEP notes:
  - Lower ratio of non-English: **10.5%** (vs. 18.1%)
  - Higher ratio of porn-ish: **14.8%** (vs. 11.2%)
  - Much higher ratio of Celeb: **10.2**% (vs. 2.7%)

Observations on Image SEP

- Text searches only show Top 10 on page 1…
  - …few people go deep
- Image searches show Top 24-30 on page 1…
  - …but hundreds altogether (easy to scroll)
  - Easier to sneak one in that people **will** see

- I do see SE’s making progress against these

SEP: Details

Modern SEP Tactics

(Research Phase 2.2)

“Dangerosity” Research, Phase 2.2

- What about “Big Event” SEP?
  - Folks love to hype the danger. Is it Truth? or Tabloid?

- Choose a couple of noteworthy events
  - Death of Steve Jobs (Oct 5th)
  - *Costa Concordia* sinking (Jan 13th)

- Search for variants over 8 days

"Big Event" SEP, Test #1

- For Mr. Jobs:
  - **9598** "SEP search term sets" logged…
  - **3** were about him
    - Specific: e.g., **"**steve jobs boat"
    - Or **"**harmony-seeking idealist jobs"
      - (arguably just 2 hits…)
  - That's about three hundredths of a percent ☹

“Big Event” SEP, Test #2

- For *Costa Concordia* wreck:
  - **12,888** “SEP search term sets” logged…
  - **3** “costa” searches
    - (note that “costa rica” had just as many)
  - **9** “cruise” searches
    - (including “pictures cruise ship cabins”)

  - That’s about a tenth of a percent ☹

“Big Event” SEP Summary

- Lots of competition for the top slots

- Maybe if you specifically look for SEP links…
  - …and go beyond first page…
  - …you can find “SEP attacks”

- But focus on # of people who clicked one…
  - …hard to find evidence for “Big Event” SEP focus

SEP: Defensive Strategies
(The “Apply” Slides!)

RSA CONFERENCE 2012

SEP Defense: Three Strategies

- User Education
- Web Filtering & Log Checking
- Secure YOUR Sites

SEP Defense: User Education

- Raise general awareness
- Google & Bing have safe "preview" feature
- Scan for obvious text problems
  - (SEs could help users do this by including bigger chunks!)
- Do basic domain analysis
  - Junk name? Relevant name?)
  - Country code TLD?

SEP Defense: Filtering & Log Reporting

- Stuff to block (and check logs for):
  - Porn/Adult
  - Hacking/Warez
  - Scam/Questionable/Illegal
  - DynDNS

SEP Defense: Secure YOUR Sites

- Don’t become part of the problem!
- SEP gangs **love** to use hacked sites
  - No domains to register
  - Use your bandwidth/space
  - Use your “Google Juice” to score higher
    - (e.g., .gov, .edu have value)
  - Known/trusted == harder to filter
- Getting blacklisted anywhere is no fun!

Why is Search Engine Poisoning Still the #1Attack Vector?

Chris Larsen

Malware Research Team Lead

Blue Coat Systems

@bc_malware_guy

securityblog@bluecoat.com **www.bluecoat.com/security**

Session ID: HT1-204

Session Classification: xxxxxxxxxxxx

RSACONFERENCE2012