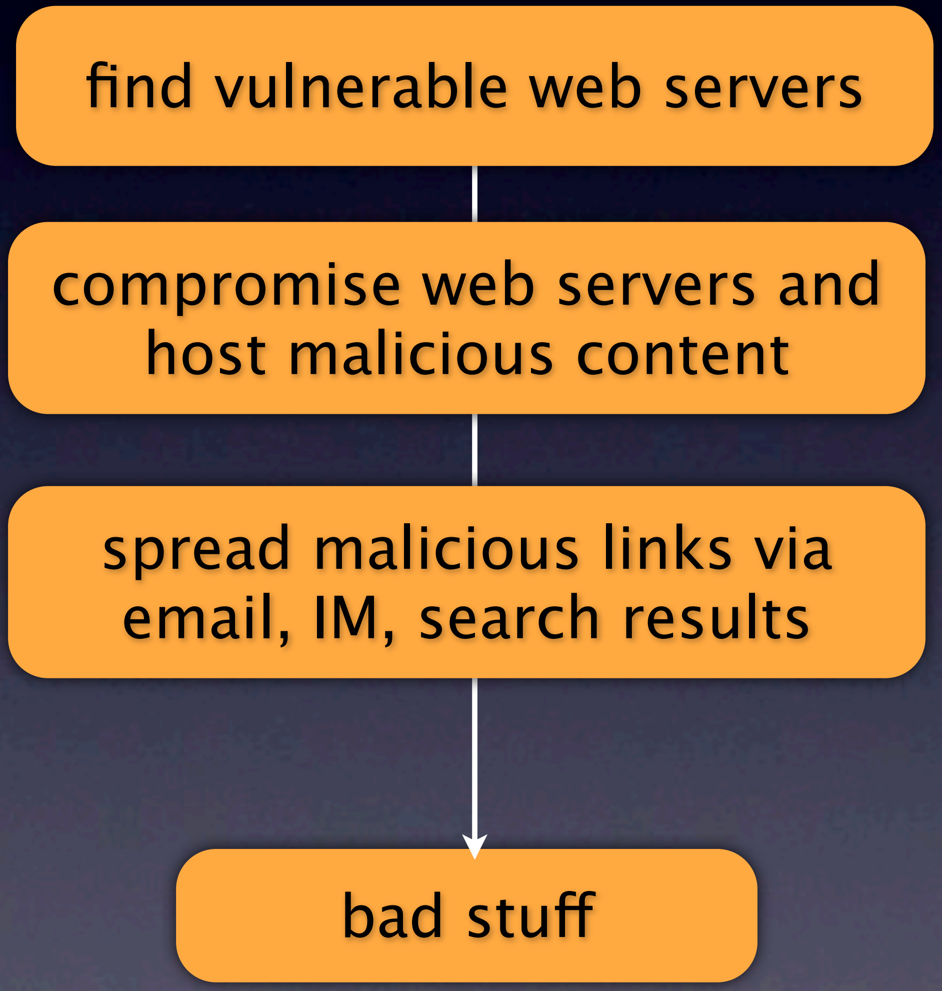
deSEO: Combating Search-Result Poisoning

John P John
Fang Yu, Yinglian Xie,
Arvind Krishnamurthy, Martin Abadi
University of Washington & MSR, Silicon Valley

The malware pipeline



find vulnerable web servers

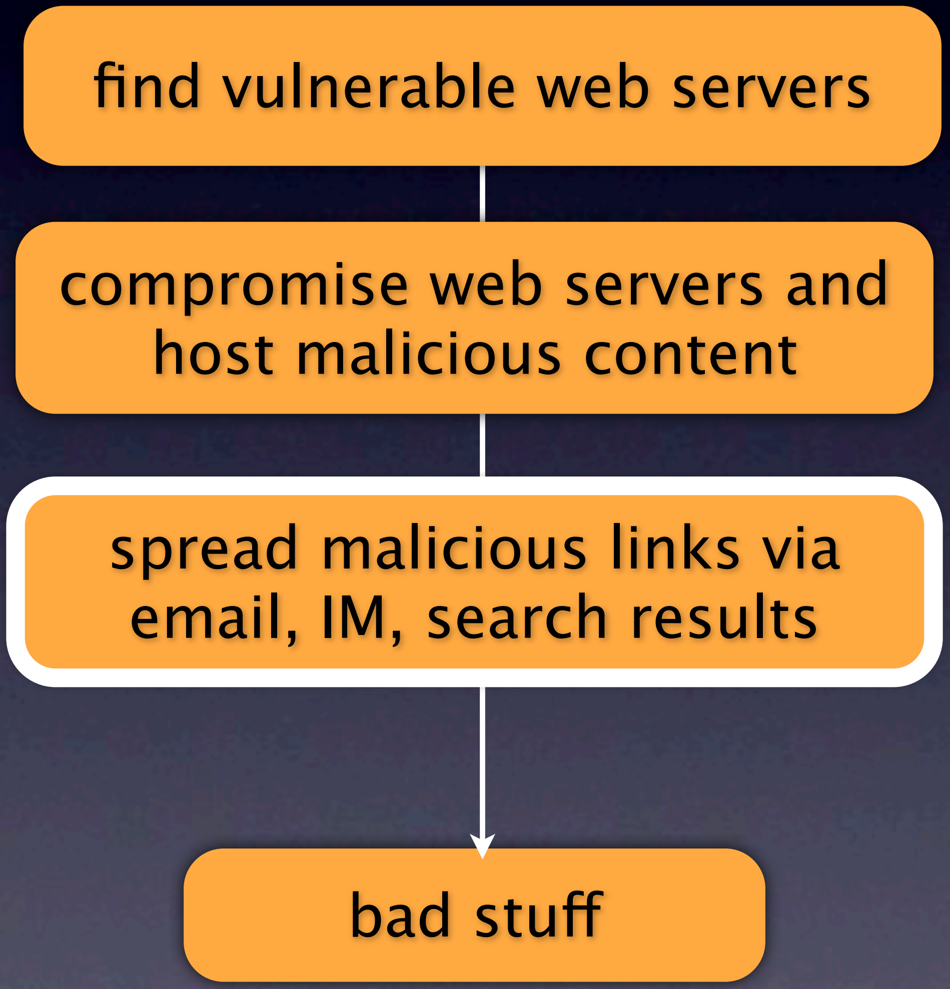
compromise web servers and host malicious content

spread malicious links via email, IM, search results

bad stuff

The malware pipeline

- Malware links spread through:
  - spam emails, spam IMs, social networks, search results, etc.
- We look at *search results*

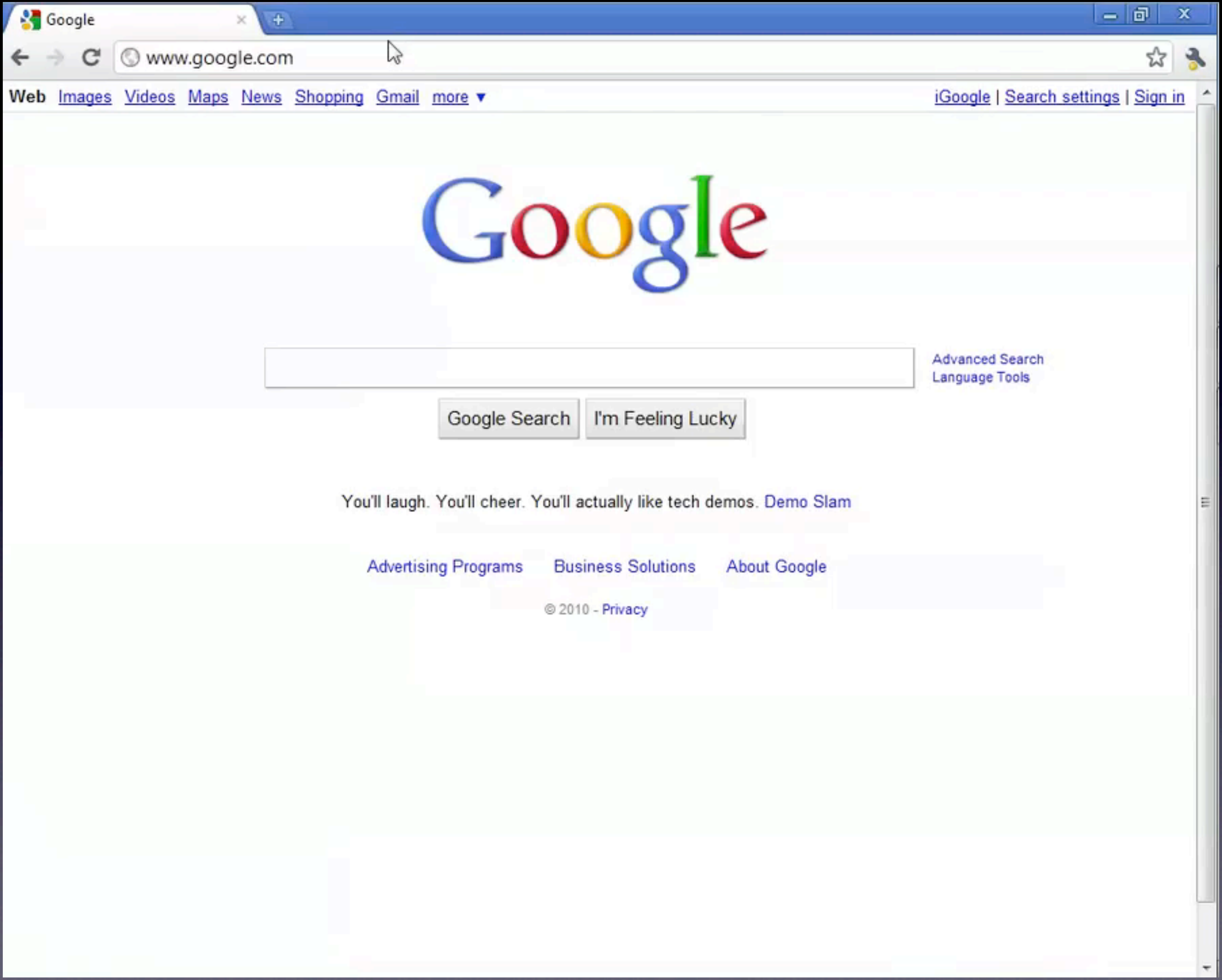


find vulnerable web servers

compromise web servers and host malicious content

spread malicious links via email, IM, search results

bad stuff



You'll laugh. You'll cheer. You'll actually like tech demos. Demo Slam

Advertising Programs Business Solutions About Google

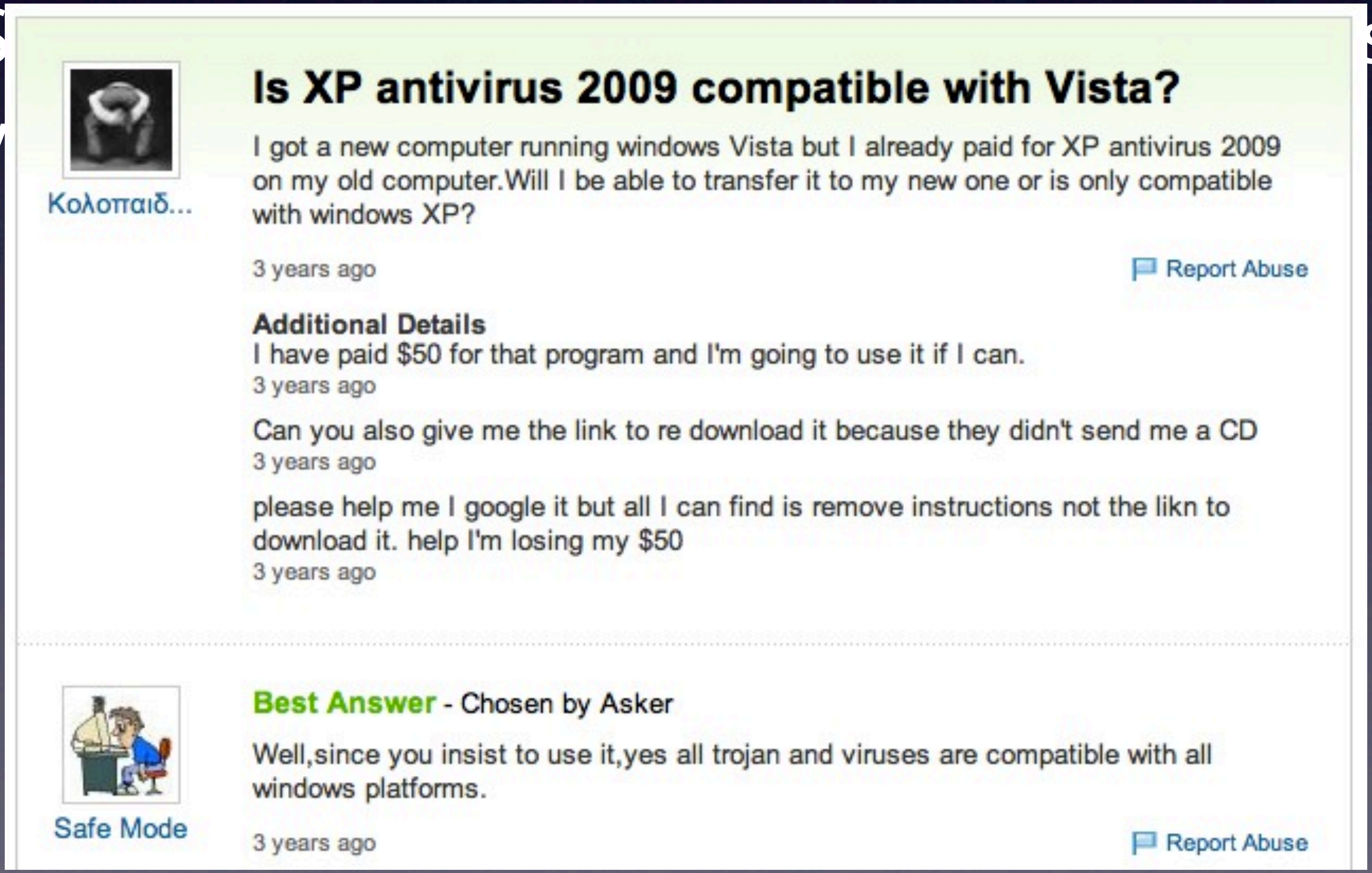
© 2010 - Privacy

Is this really a problem?

- ~40% of popular searches contain at least one malicious link in top results
- Scareware fraud made $150 m. in profit last year

Is this really a problem?

- ~40% of popular searches contain at least one malicious link in top results
- S st y



Is XP antivirus 2009 compatible with Vista?

I got a new computer running windows Vista but I already paid for XP antivirus 2009 on my old computer.Will I be able to transfer it to my new one or is only compatible with windows XP?

3 years ago

Additional Details
I have paid $50 for that program and I'm going to use it if I can.
3 years ago

Can you also give me the link to re download it because they didn't send me a CD
3 years ago

please help me I google it but all I can find is remove instructions not the likn to download it. help I'm losing my $50
3 years ago

Best Answer - Chosen by Asker

Well,since you insist to use it,yes all trojan and viruses are compatible with all windows platforms.

3 years ago

Contributions

- How does the search poisoning attack work?

  -examined a live attack involving 5,000 compromised sites

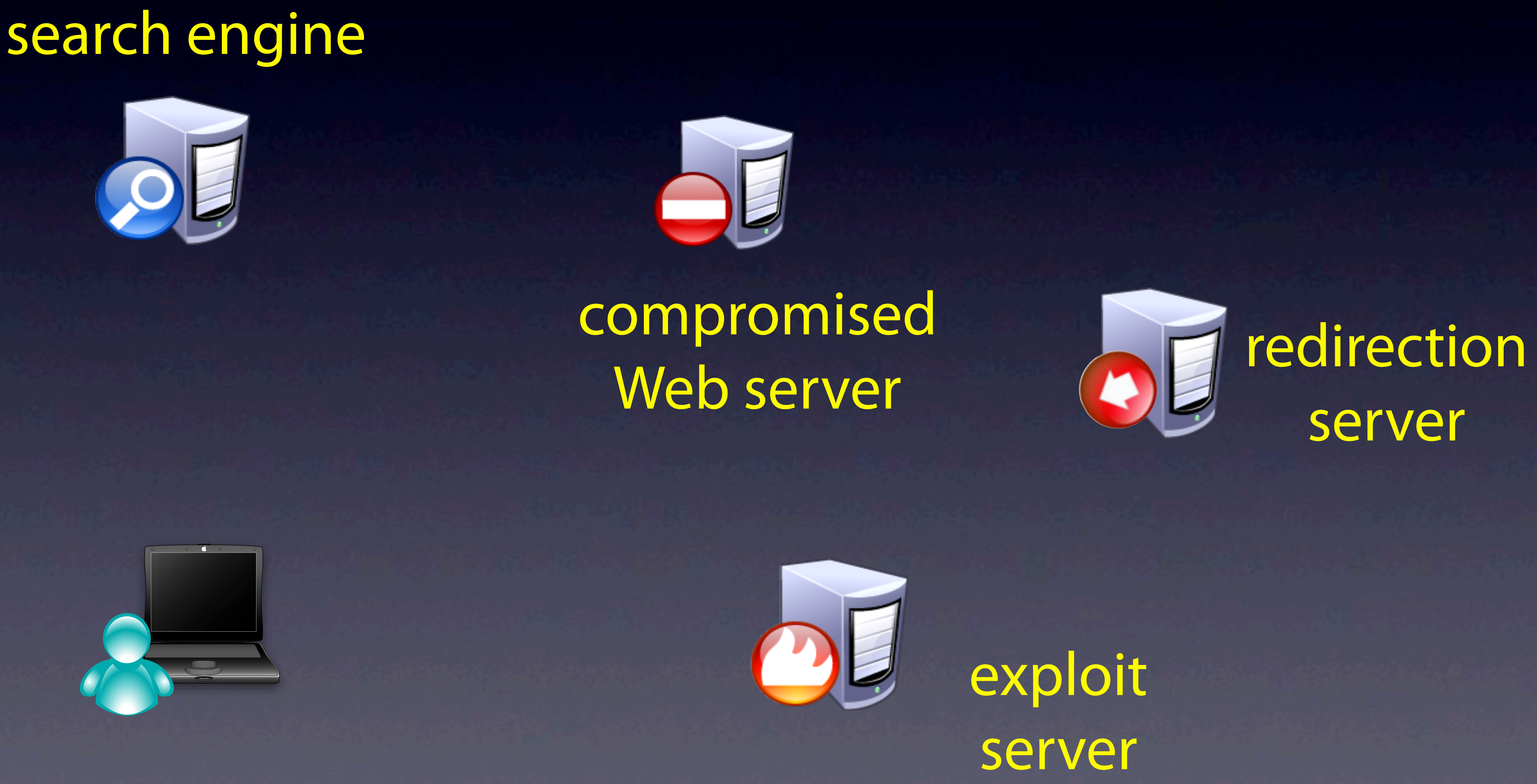
- What can we learn about such attacks?

  -identified common features in search poisoning attacks

- How can we defend against them?

  -developed deSEO, which detected new live SEO attacks on 1,000+ domains

Anatomy of SEO attack



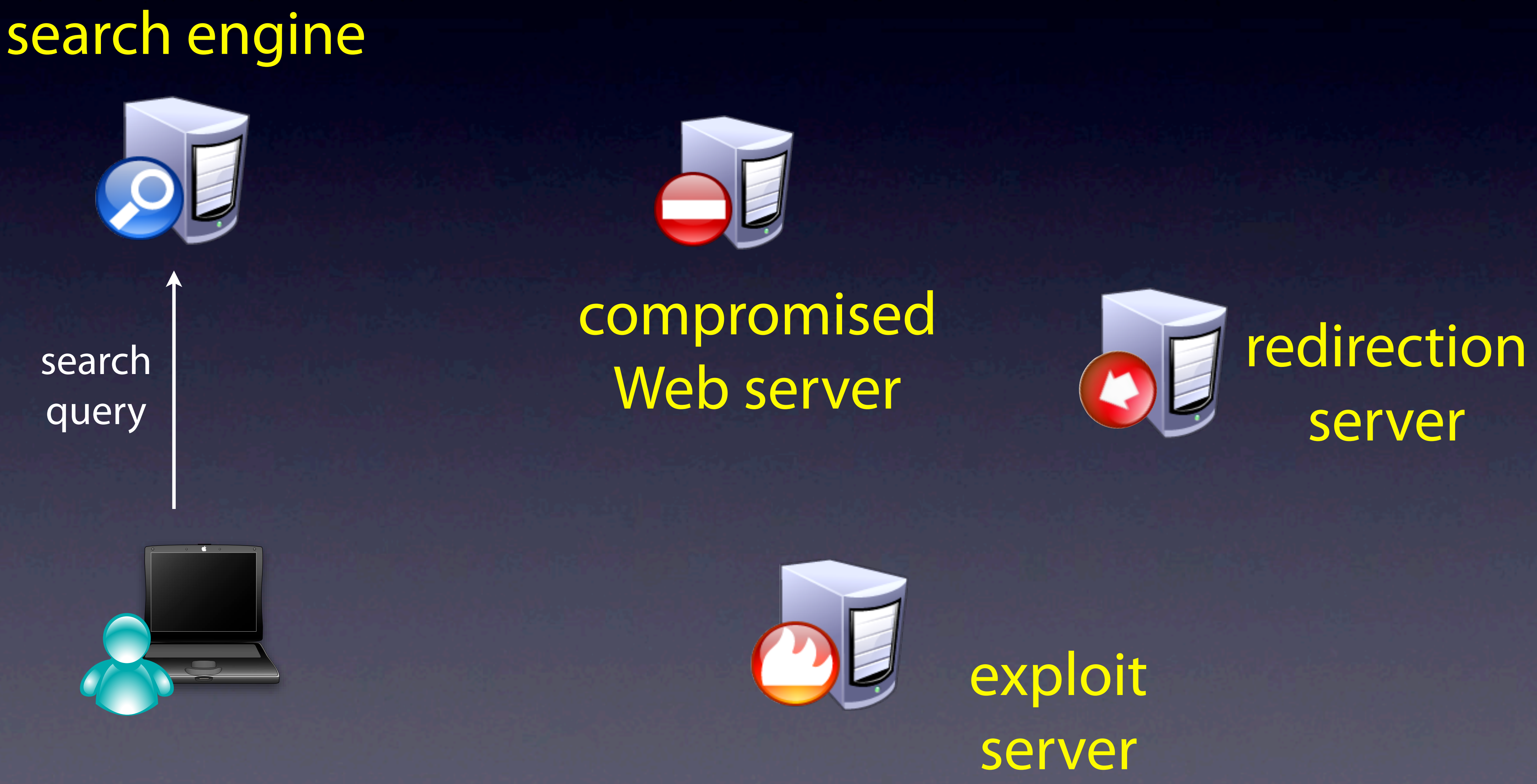
search engine

compromised
Web server

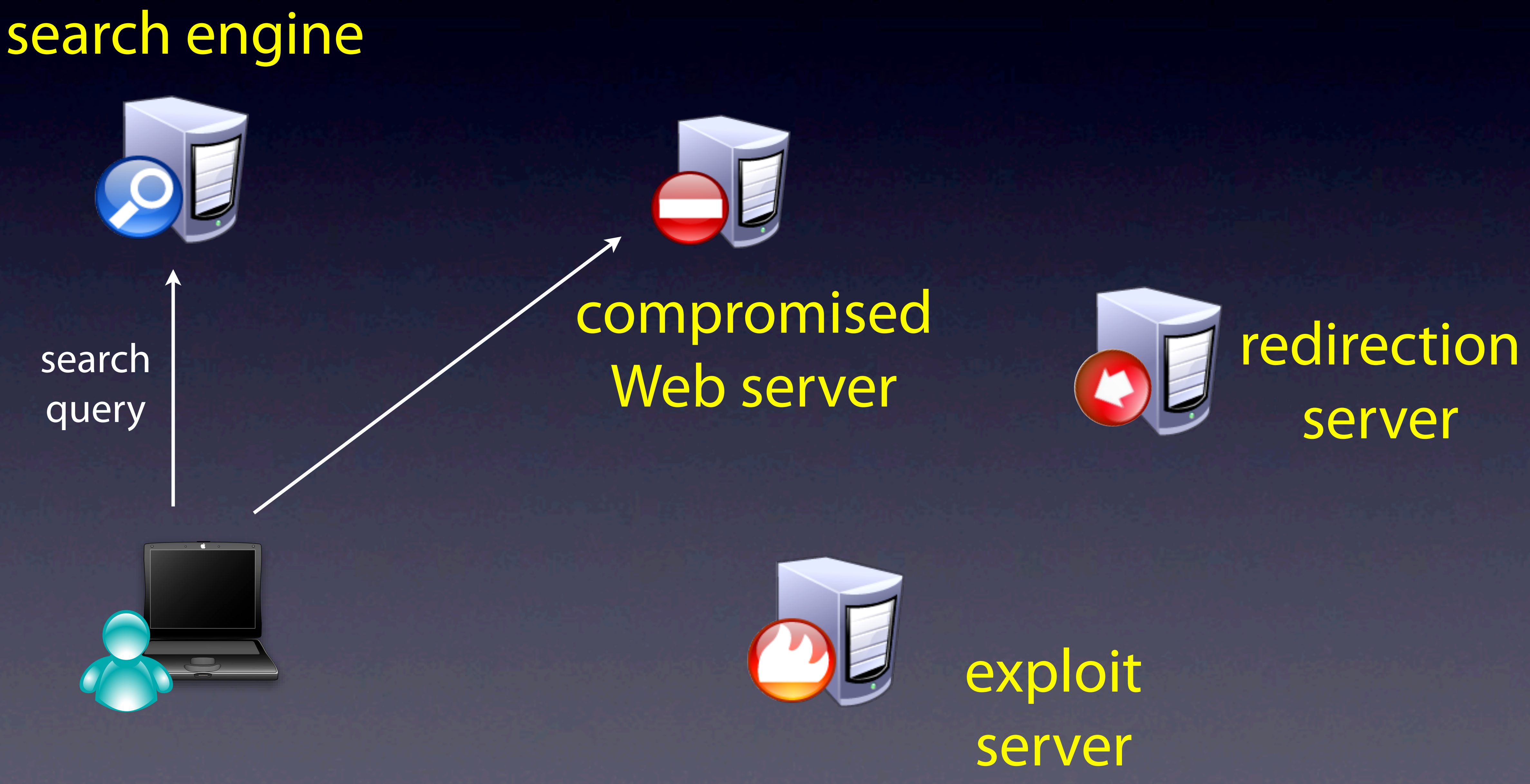
redirection
server

exploit
server

Anatomy of SEO attack



Anatomy of SEO attack



search engine

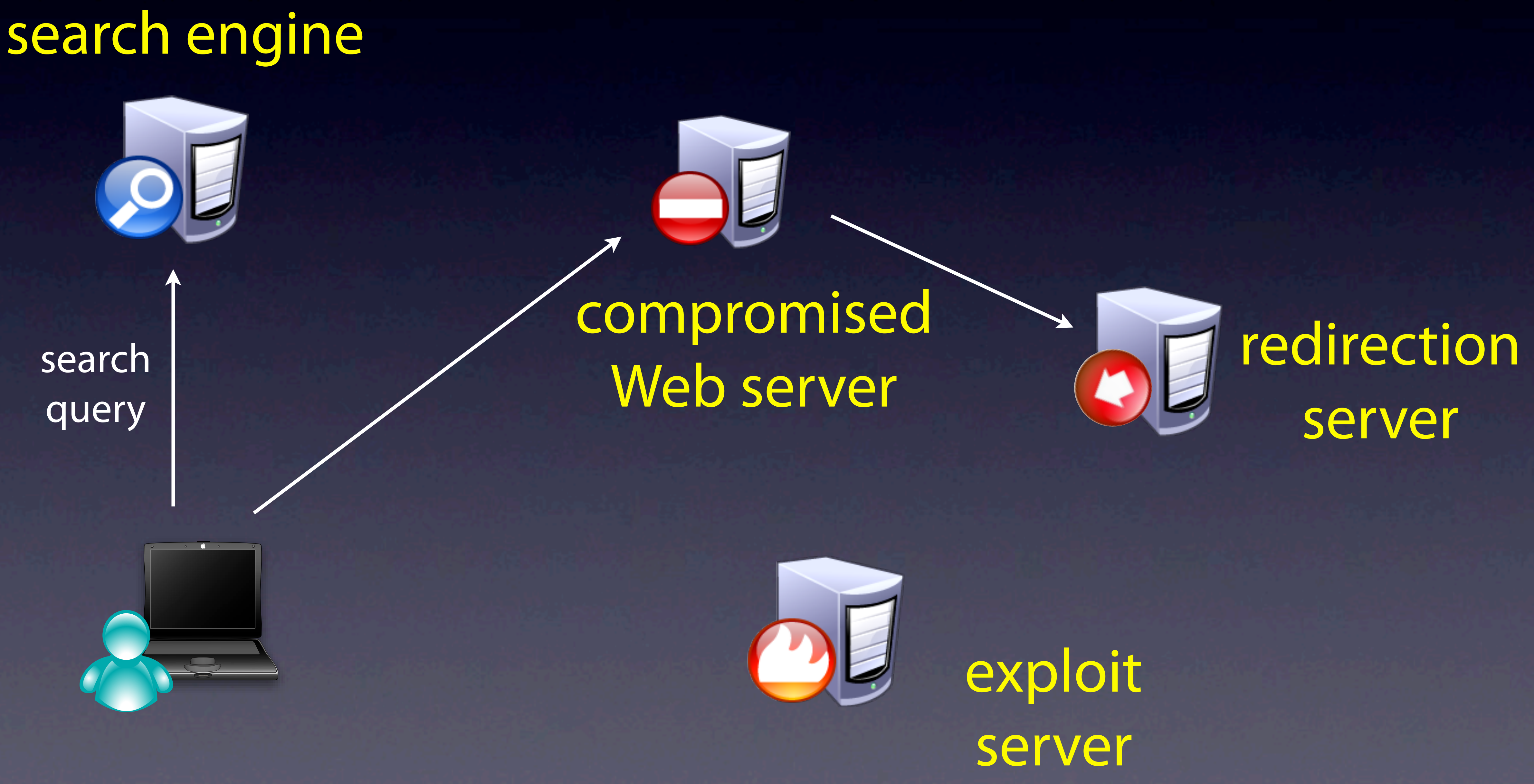
search query

compromised Web server

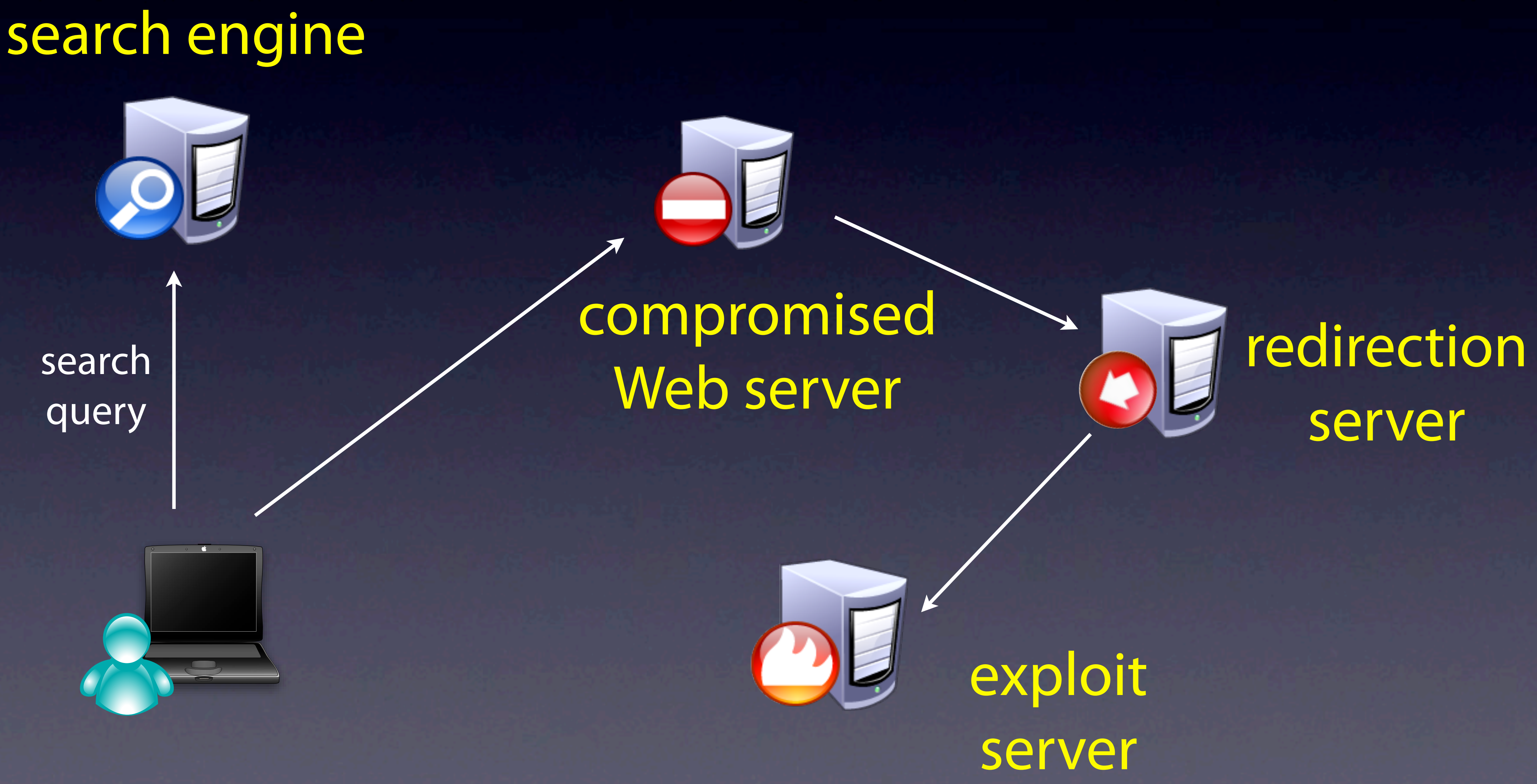
redirection server

exploit server

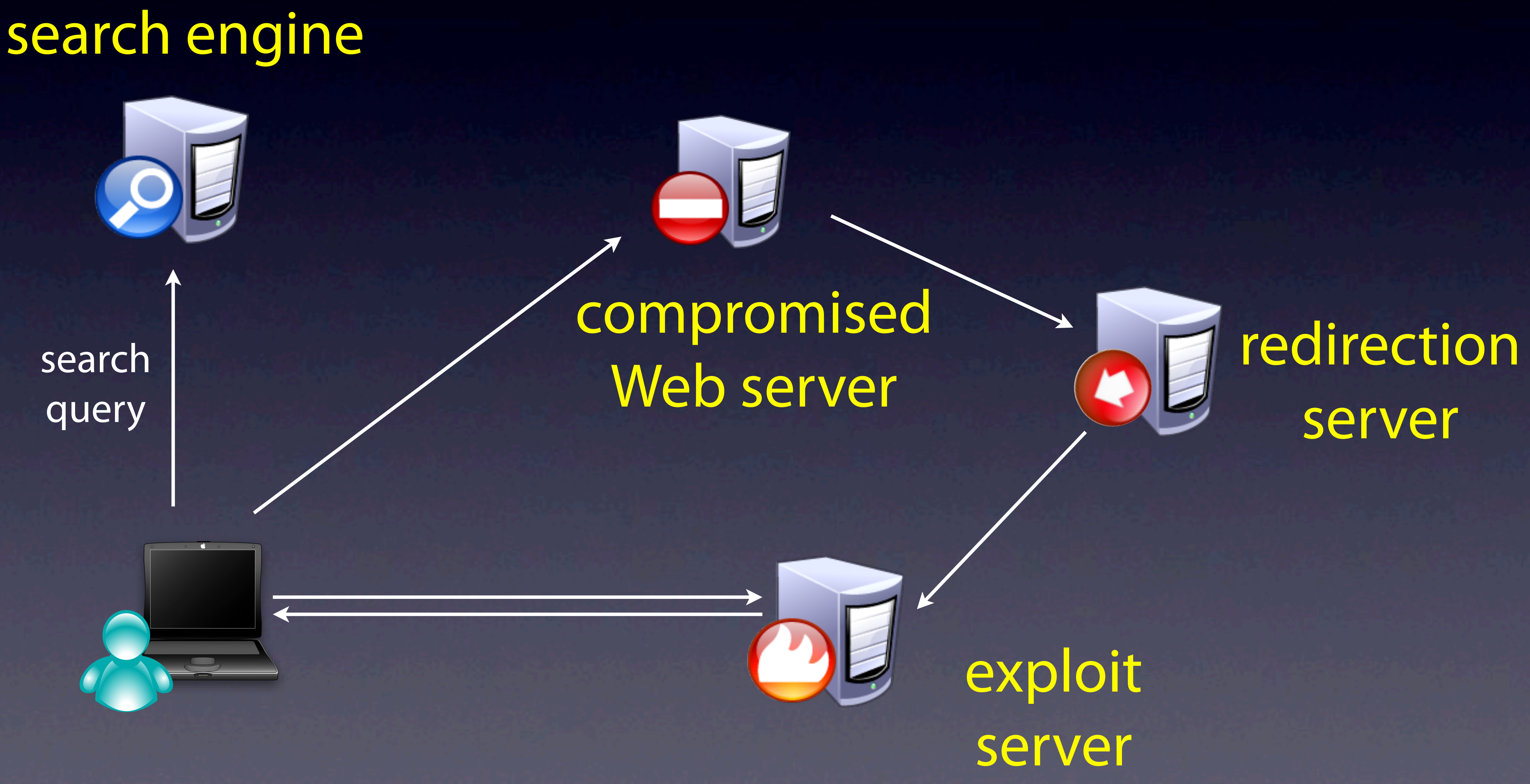
Anatomy of SEO attack



Anatomy of SEO attack



Anatomy of SEO attack



Analysis of an attack

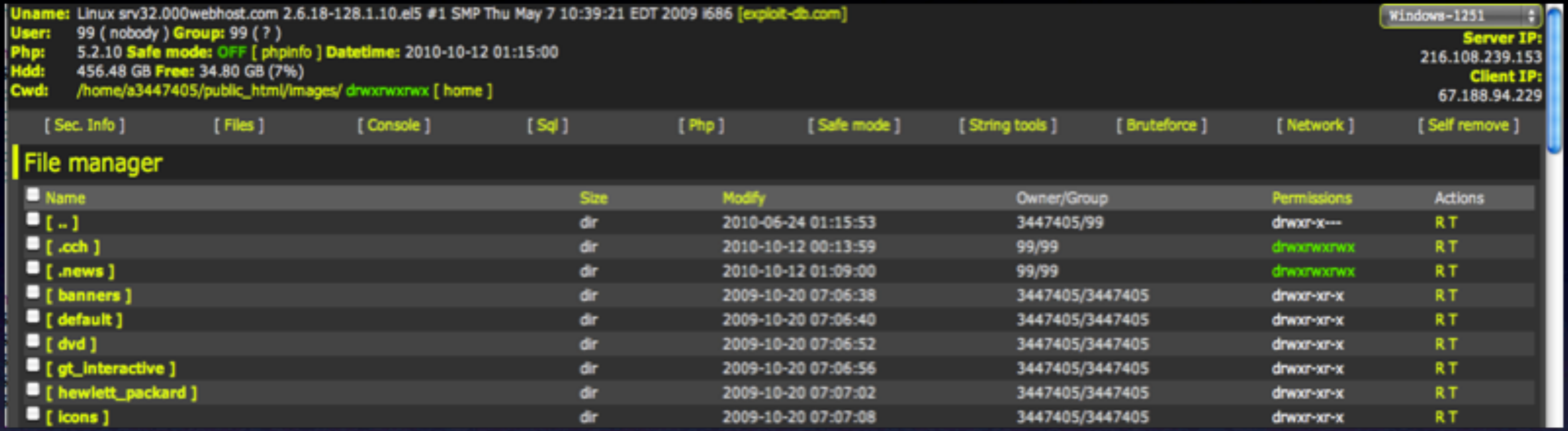
- Examine a specific attack
  - August - October 2010
  - 5,000 compromised domains
  - Tens of thousands of compromised keywords
  - Millions of SEO pages generated

How are servers compromised?

- Sites running osCommerce
- Unpatched vulnerabilities
- Allows attackers to host any file on the Web server - including executables

  `www.example.com/admin/file_manager.php/login.php?action=processuploads`

What files are uploaded?



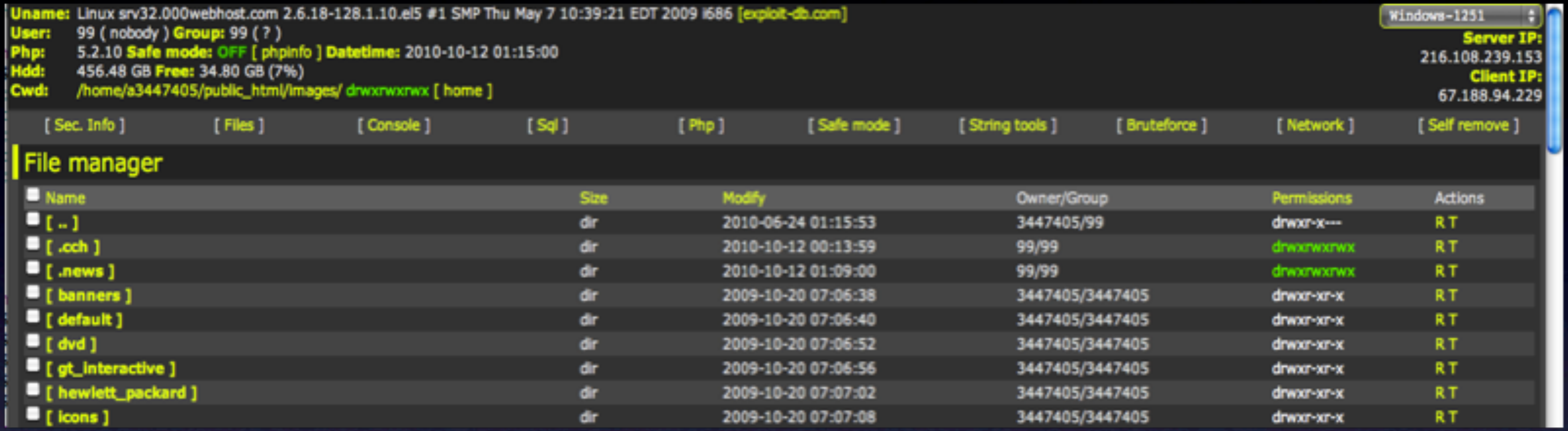
Uname: Linux srv32.000webhost.com 2.6.18-128.1.10.el5 #1 SMP Thu May 7 10:39:21 EDT 2009 i686 [exploit-db.com]
User: 99 (nobody) **Group:** 99 (?)
Php: 5.2.10 **Safe mode:** OFF [phpinfo] **Datetime:** 2010-10-12 01:15:00
Hdd: 456.48 GB **Free:** 34.80 GB (7%)
Cwd: /home/a3447405/public_html/images/ drwxrwxrwx [home]

Server IP:
216.108.239.153
Client IP:
67.188.94.229

File manager

□ Name	Size	Modify	Owner/Group	Permissions	Actions
□ [..]	dir	2010-06-24 01:15:53	3447405/99	drwxr-x---	R T
□ [.cch]	dir	2010-10-12 00:13:59	99/99	drwxrwxrwx	R T
□ [.news]	dir	2010-10-12 01:09:00	99/99	drwxrwxrwx	R T
□ [banners]	dir	2009-10-20 07:06:38	3447405/3447405	drwxr-xr-x	R T
□ [default]	dir	2009-10-20 07:06:40	3447405/3447405	drwxr-xr-x	R T
□ [dvd]	dir	2009-10-20 07:06:52	3447405/3447405	drwxr-xr-x	R T
□ [gt_interactive]	dir	2009-10-20 07:06:56	3447405/3447405	drwxr-xr-x	R T
□ [hewlett_packard]	dir	2009-10-20 07:07:02	3447405/3447405	drwxr-xr-x	R T
□ [icons]	dir	2009-10-20 07:07:08	3447405/3447405	drwxr-xr-x	R T

What files are uploaded?



```
Uname: Linux srv32.000webhost.com 2.6.18-128.1.10.el5 #1 SMP Thu May 7 10:39:21 EDT 2009 i686 [exploit-db.com]
User:  99 ( nobody ) Group: 99 ( ? )
Php:   5.2.10 Safe mode: OFF [ phpinfo ] Datetime: 2010-10-12 01:15:00
Hdd:   456.48 GB Free: 34.80 GB (7%)
Cwd:   /home/a3447405/public_html/images/ drwxrwxrwx [ home ]
```

Windows-1251

Server IP:
216.108.239.153
Client IP:
67.188.94.229

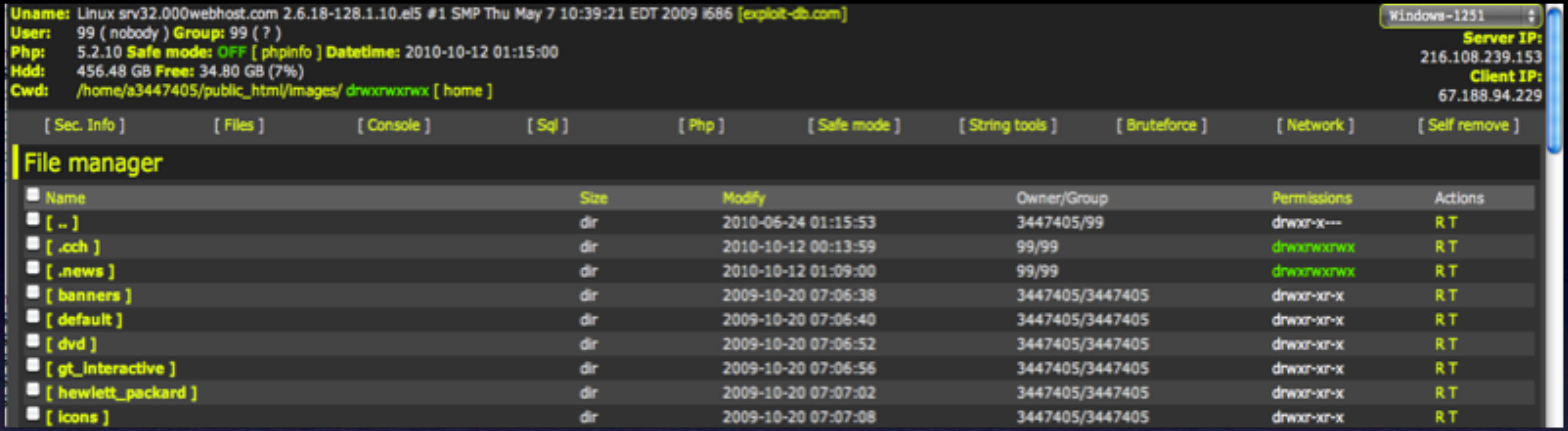
[Sec. Info] [Files] [Console] [Sql] [Php] [Safe mode] [String tools] [Bruteforce] [Network] [Self remove]

File manager

□ Name	Size	Modify	Owner/Group	Permissions	Actions
□ [..]	dir	2010-06-24 01:15:53	3447405/99	drwxr-x---	R T
□ [.cch]	dir	2010-10-12 00:13:59	99/99	drwxrwxrwx	R T
□ [.news]	dir	2010-10-12 01:09:00	99/99	drwxrwxrwx	R T
□ [banners]	dir	2009-10-20 07:06:38	3447405/3447405	drwxr-xr-x	R T
□ [default]	dir	2009-10-20 07:06:40	3447405/3447405	drwxr-xr-x	R T
□ [dvd]	dir	2009-10-20 07:06:52	3447405/3447405	drwxr-xr-x	R T
□ [gt_interactive]	dir	2009-10-20 07:06:56	3447405/3447405	drwxr-xr-x	R T
□ [hewlett_packard]	dir	2009-10-20 07:07:02	3447405/3447405	drwxr-xr-x	R T
□ [icons]	dir	2009-10-20 07:07:08	3447405/3447405	drwxr-xr-x	R T

- php shell to manage file operations

What files are uploaded?



Uname: Linux srv32.000webhost.com 2.6.18-128.1.10.el5 #1 SMP Thu May 7 10:39:21 EDT 2009 i686 [exploit-db.com]
User: 99 (nobody) **Group:** 99 (?)
Php: 5.2.10 **Safe mode:** OFF [phpinfo] **Datetime:** 2010-10-12 01:15:00
Hdd: 456.48 GB **Free:** 34.80 GB (7%)
Cwd: /home/a3447405/public_html/images/ drwxrwxrwx [home]

Windows-1251
Server IP:
216.108.239.153
Client IP:
67.188.94.229

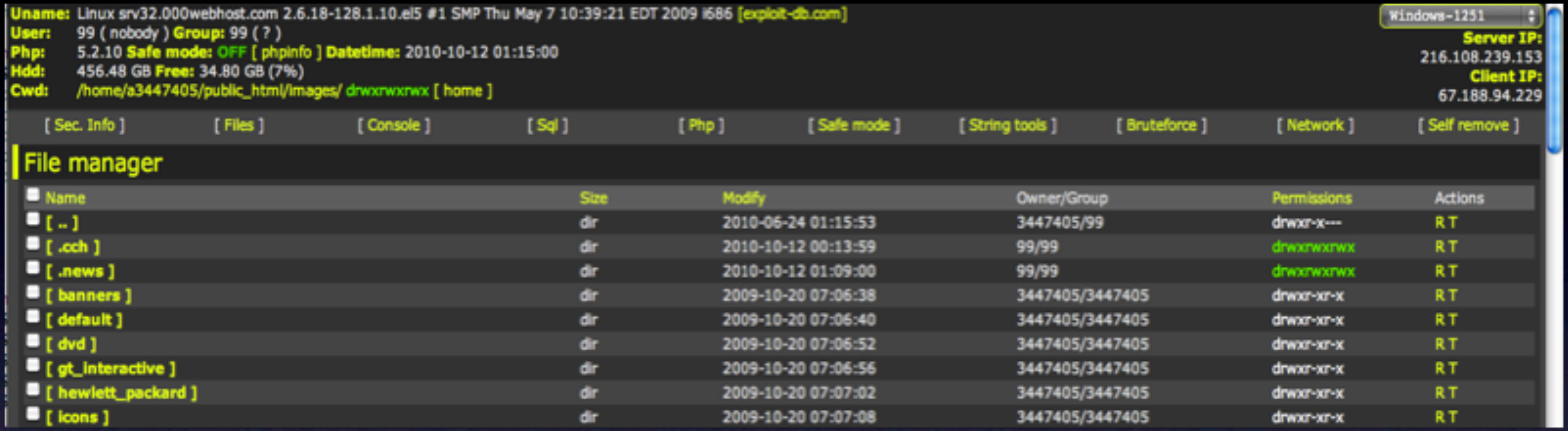
[Sec. Info] [Files] [Console] [Sql] [Php] [Safe mode] [String tools] [Bruteforce] [Network] [Self remove]

File manager

□ Name	Size	Modify	Owner/Group	Permissions	Actions
□ [..]	dir	2010-06-24 01:15:53	3447405/99	drwxr-x---	R T
□ [.cch]	dir	2010-10-12 00:13:59	99/99	drwxrwxrwx	R T
□ [.news]	dir	2010-10-12 01:09:00	99/99	drwxrwxrwx	R T
□ [banners]	dir	2009-10-20 07:06:38	3447405/3447405	drwxr-xr-x	R T
□ [default]	dir	2009-10-20 07:06:40	3447405/3447405	drwxr-xr-x	R T
□ [dvd]	dir	2009-10-20 07:06:52	3447405/3447405	drwxr-xr-x	R T
□ [gt_interactive]	dir	2009-10-20 07:06:56	3447405/3447405	drwxr-xr-x	R T
□ [hewlett_packard]	dir	2009-10-20 07:07:02	3447405/3447405	drwxr-xr-x	R T
□ [icons]	dir	2009-10-20 07:07:08	3447405/3447405	drwxr-xr-x	R T

- php shell to manage file operations
- HTML templates, images

What files are uploaded?



```
Uname: Linux srv32.000webhost.com 2.6.18-128.1.10.el5 #1 SMP Thu May 7 10:39:21 EDT 2009 i686 [exploit-db.com]
User:  99 ( nobody ) Group: 99 ( ? )
Php:   5.2.10 Safe mode: OFF [ phpinfo ] Datetime: 2010-10-12 01:15:00
Hdd:   456.48 GB Free: 34.80 GB (7%)
Cwd:   /home/a3447405/public_html/images/ drwxrwxrwx [ home ]
```

```
Server IP:
216.108.239.153
Client IP:
67.188.94.229
```

File manager

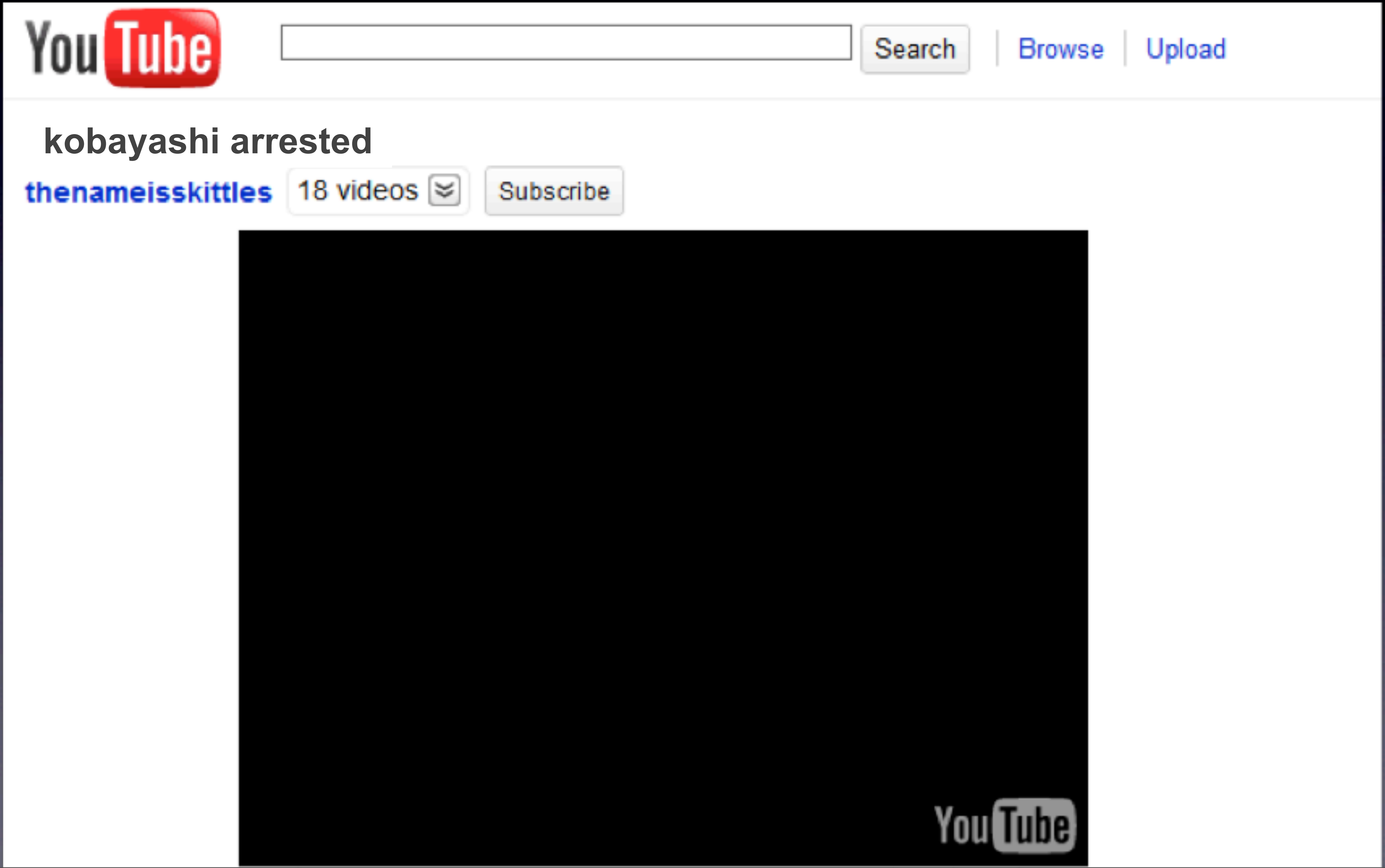
□ Name	Size	Modify	Owner/Group	Permissions	Actions
□ [..]	dir	2010-06-24 01:15:53	3447405/99	drwxr-x---	R T
□ [.cch]	dir	2010-10-12 00:13:59	99/99	drwxrwxrwx	R T
□ [.news]	dir	2010-10-12 01:09:00	99/99	drwxrwxrwx	R T
□ [banners]	dir	2009-10-20 07:06:38	3447405/3447405	drwxr-xr-x	R T
□ [default]	dir	2009-10-20 07:06:40	3447405/3447405	drwxr-xr-x	R T
□ [dvd]	dir	2009-10-20 07:06:52	3447405/3447405	drwxr-xr-x	R T
□ [gt_interactive]	dir	2009-10-20 07:06:56	3447405/3447405	drwxr-xr-x	R T
□ [hewlett_packard]	dir	2009-10-20 07:07:02	3447405/3447405	drwxr-xr-x	R T
□ [icons]	dir	2009-10-20 07:07:08	3447405/3447405	drwxr-xr-x	R T

- php shell to manage file operations
- HTML templates, images
- php script to generate SEO web pages

The main php script

www.example.com/images/page.php?page=kobayashi+arrested

The main php script

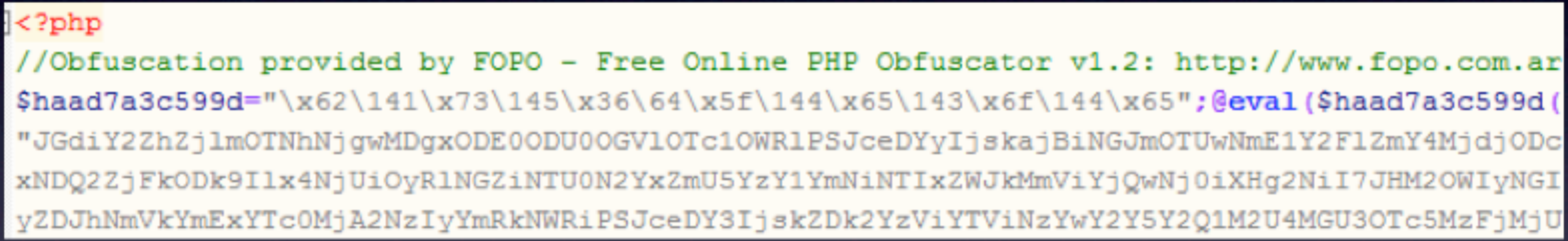


kobayashi arrested

thenameisskittles

The main php script

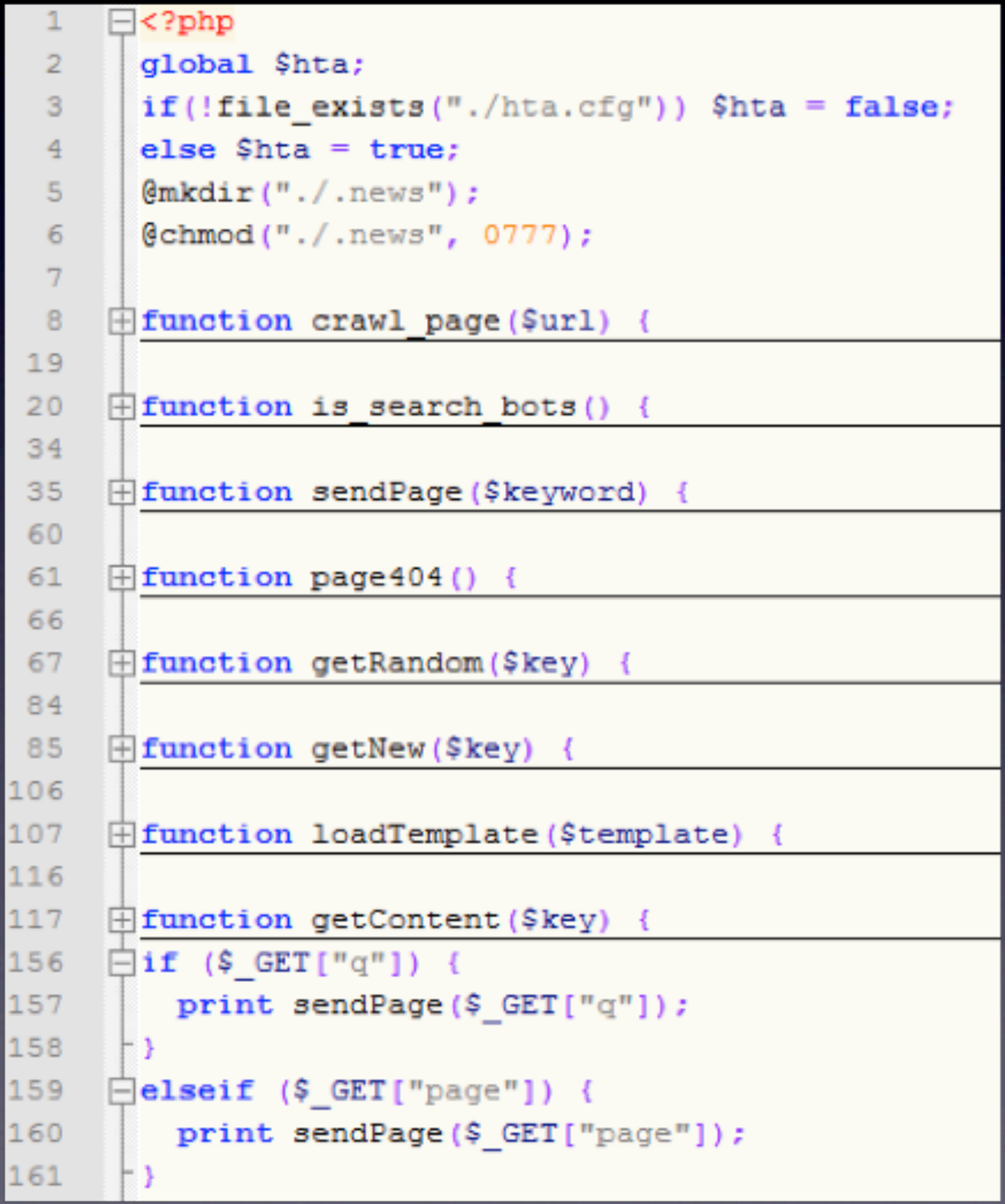
www.example.com/images/page.php?page=kobayashi+arrested



```
<?php
//Obfuscation provided by FOPO - Free Online PHP Obfuscator v1.2: http://www.fopo.com.ar
$haad7a3c599d="\x62\141\x73\145\x36\64\x5f\144\x65\143\x6f\144\x65";@eval($haad7a3c599d(
"JGdiY2ZhZjlmOTNhNjgwMDgxODE0ODU0OGV1OTc1OWR1PSJceDYyIjskajBiNGJmOTUwNmE1Y2F1ZmY4MjdjODc
xNDQ2ZjFkODk9Il x4NjUiOyR1NGZiNTU0N2YxZmU5YzY1YmNiNTIxZWJkMmViYjQwNj0iXHg2NiI7JHM2OWIyNGI
yZDJhNmVkYmExYTc0MjA2NzIyYmRkNWRiPSJceDY3IjskZDk2YzViYTViNzYwY2Y5Y2Q1M2U4MGU3OTc5MzFjMjU
```

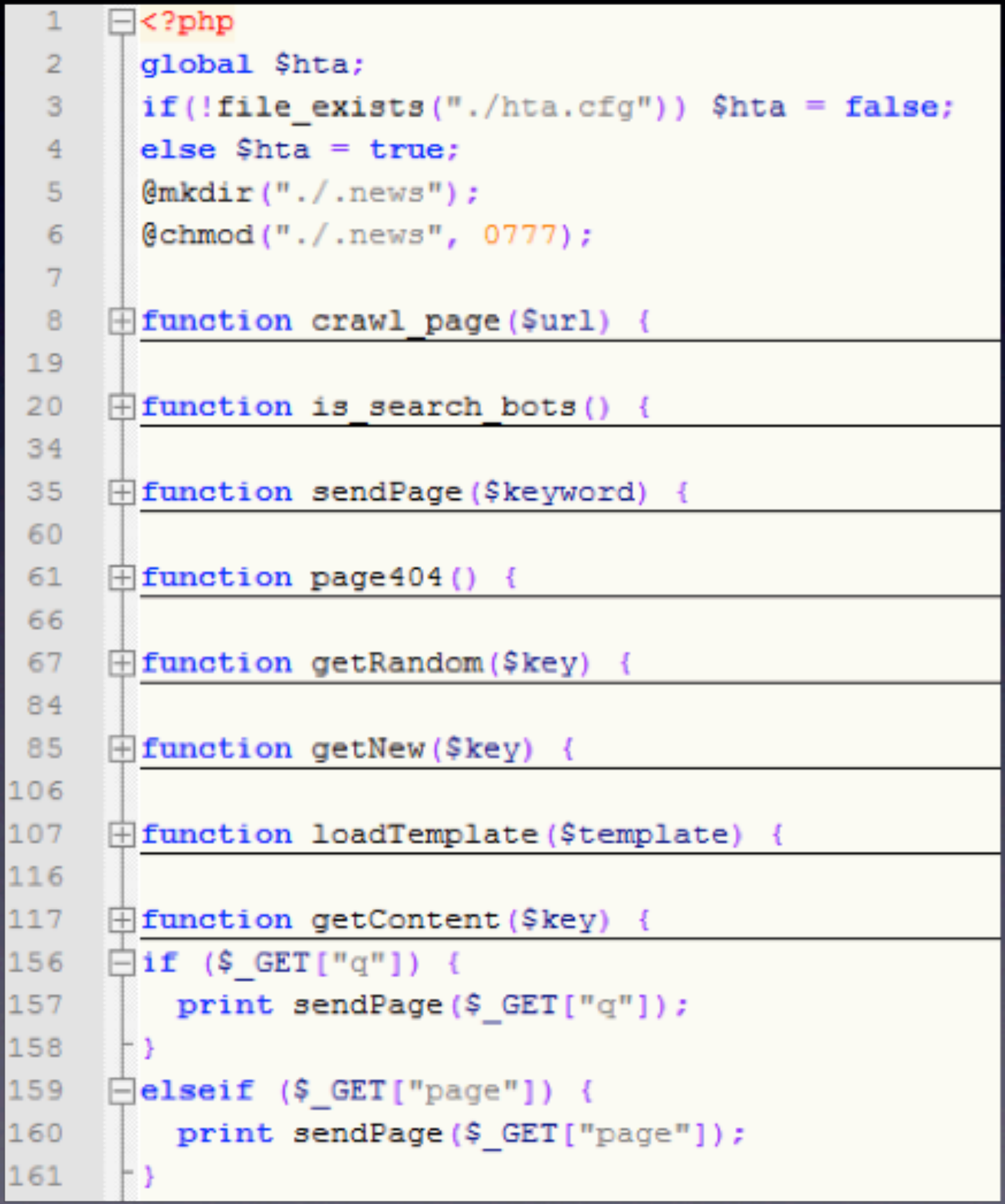
- Obfuscated script
- Simple encryption using nested *evals*

The main script (de-obfuscated)



```php
<?php
global $hta;
if(!file_exists("./hta.cfg")) $hta = false;
else $hta = true;
@mkdir("./.news");
@chmod("./.news", 0777);

function crawl_page($url) {

function is_search_bots() {

function sendPage($keyword) {

function page404() {

function getRandom($key) {

function getNew($key) {

function loadTemplate($template) {

function getContent($key) {
if ($_GET["q"]) {
  print sendPage($_GET["q"]);
}
elseif ($_GET["page"]) {
  print sendPage($_GET["page"]);
}
```

The main script (de-obfuscated)

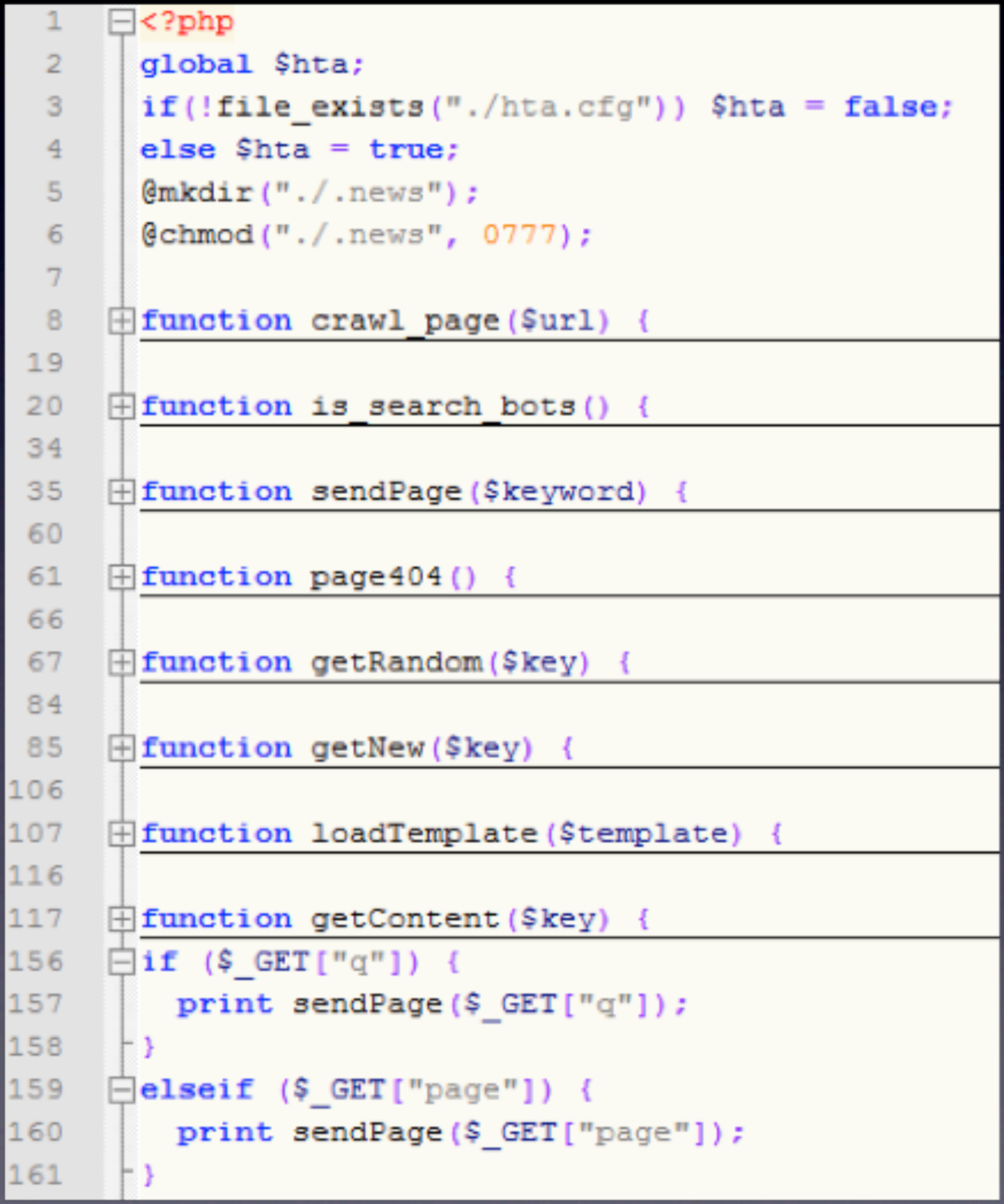


```php
<?php
global $hta;
if(!file_exists("./hta.cfg")) $hta = false;
else $hta = true;
@mkdir("./.news");
@chmod("./.news", 0777);

function crawl_page($url) {

function is_search_bots() {

function sendPage($keyword) {

function page404() {

function getRandom($key) {

function getNew($key) {

function loadTemplate($template) {

function getContent($key) {
if ($_GET["q"]) {
   print sendPage($_GET["q"]);
}
elseif ($_GET["page"]) {
   print sendPage($_GET["page"]);
}
```

Check if search crawler

Generate page for keyword

The main script (de-obfuscated)



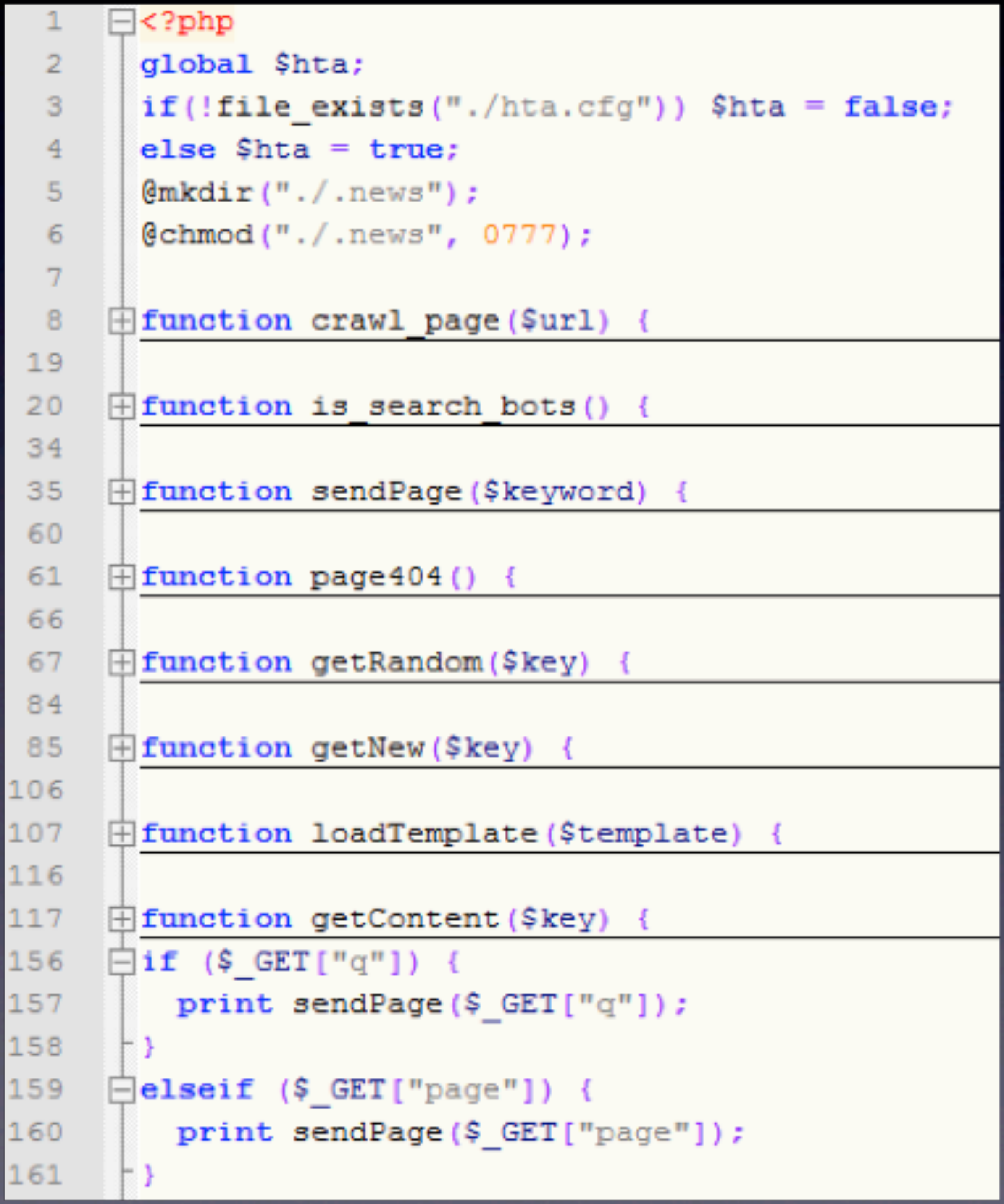
```
<?php
global $hta;
if(!file_exists("./hta.cfg")) $hta = false;
else $hta = true;
@mkdir("./.news");
@chmod("./.news", 0777);

function crawl_page($url) {

function is_search_bots() {

function sendPage($keyword) {

function page404() {

function getRandom($key) {

function getNew($key) {

function loadTemplate($template) {

function getContent($key) {
if ($_GET["q"]) {
   print sendPage($_GET["q"]);
}
elseif ($_GET["page"]) {
   print sendPage($_GET["page"]);
}
```

Check if search crawler

Generate page for keyword

Fetch:
   snippets from google
   images from bing

The main script (de-obfuscated)



```
<?php
global $hta;
if(!file_exists("./hta.cfg")) $hta = false;
else $hta = true;
@mkdir("./.news");
@chmod("./.news", 0777);

function crawl_page($url) {

function is_search_bots() {

function sendPage($keyword) {

function page404() {

function getRandom($key) {

function getNew($key) {

function loadTemplate($template) {

function getContent($key) {
if ($_GET["q"]) {
    print sendPage($_GET["q"]);
}
elseif ($_GET["page"]) {
    print sendPage($_GET["page"]);
}
```

Check if search crawler

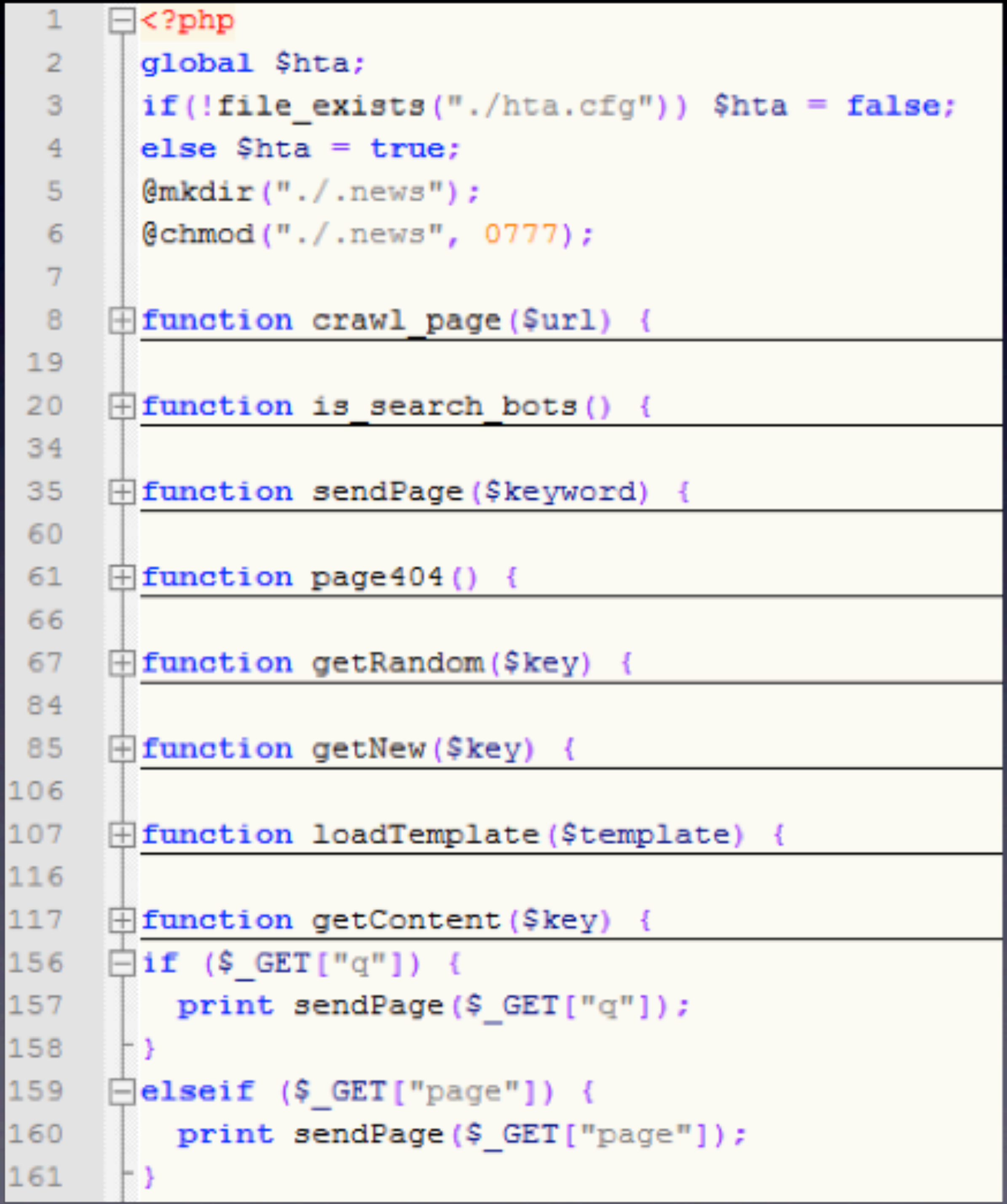
Generate page for keyword

Fetch:

- snippets from google
- images from bing

Add links to other compromised sites

The main script (de-obfuscated)



```php
<?php
global $hta;
if(!file_exists("./hta.cfg")) $hta = false;
else $hta = true;
@mkdir("./.news");
@chmod("./.news", 0777);

function crawl_page($url) {

function is_search_bots() {

function sendPage($keyword) {

function page404() {

function getRandom($key) {

function getNew($key) {

function loadTemplate($template) {

function getContent($key) {
if ($_GET["q"]) {
   print sendPage($_GET["q"]);
}
elseif ($_GET["page"]) {
   print sendPage($_GET["page"]);
}
```

Check if search crawler

Generate page for keyword

Fetch:

- snippets from google
- images from bing

Add links to other compromised sites

Cache page

Dense link structure

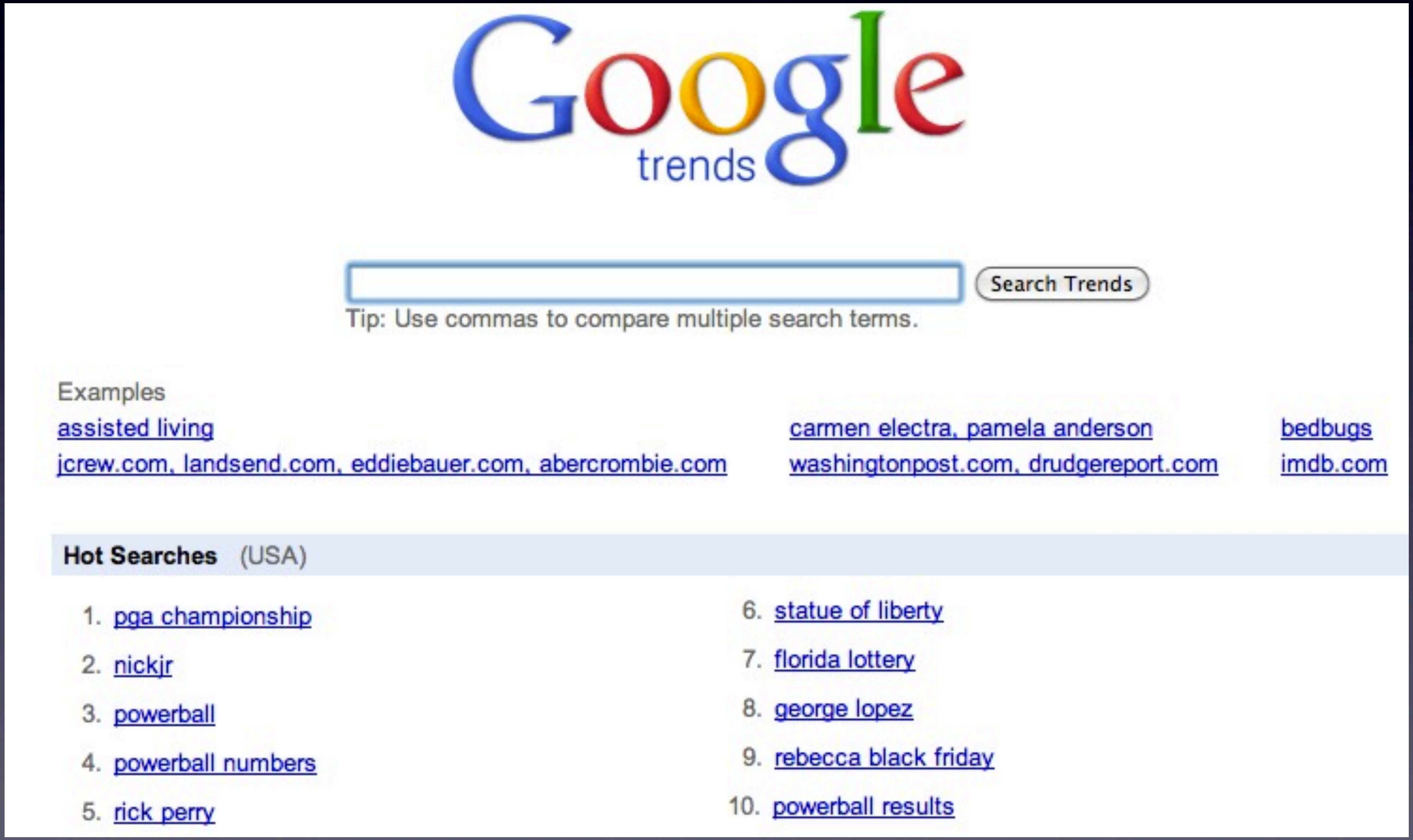
- Other compromised domains found by crawling included links
- Each site linked to 200 other sites
- ~5,000 compromised domains identified
- Each site hosted 8,000 SEO pages
  - **40 million** pages total

Poisoned keywords

- 20,000+ popular search terms poisoned

Poisoned keywords

- 20,000+ popular search terms poisoned



Tip: Use commas to compare multiple search terms.

Examples

assisted living

jcrew.com, landsend.com, eddiebauer.com, abercrombie.com

carmen electra, pamela anderson

washingtonpost.com, drudgereport.com

bedbugs

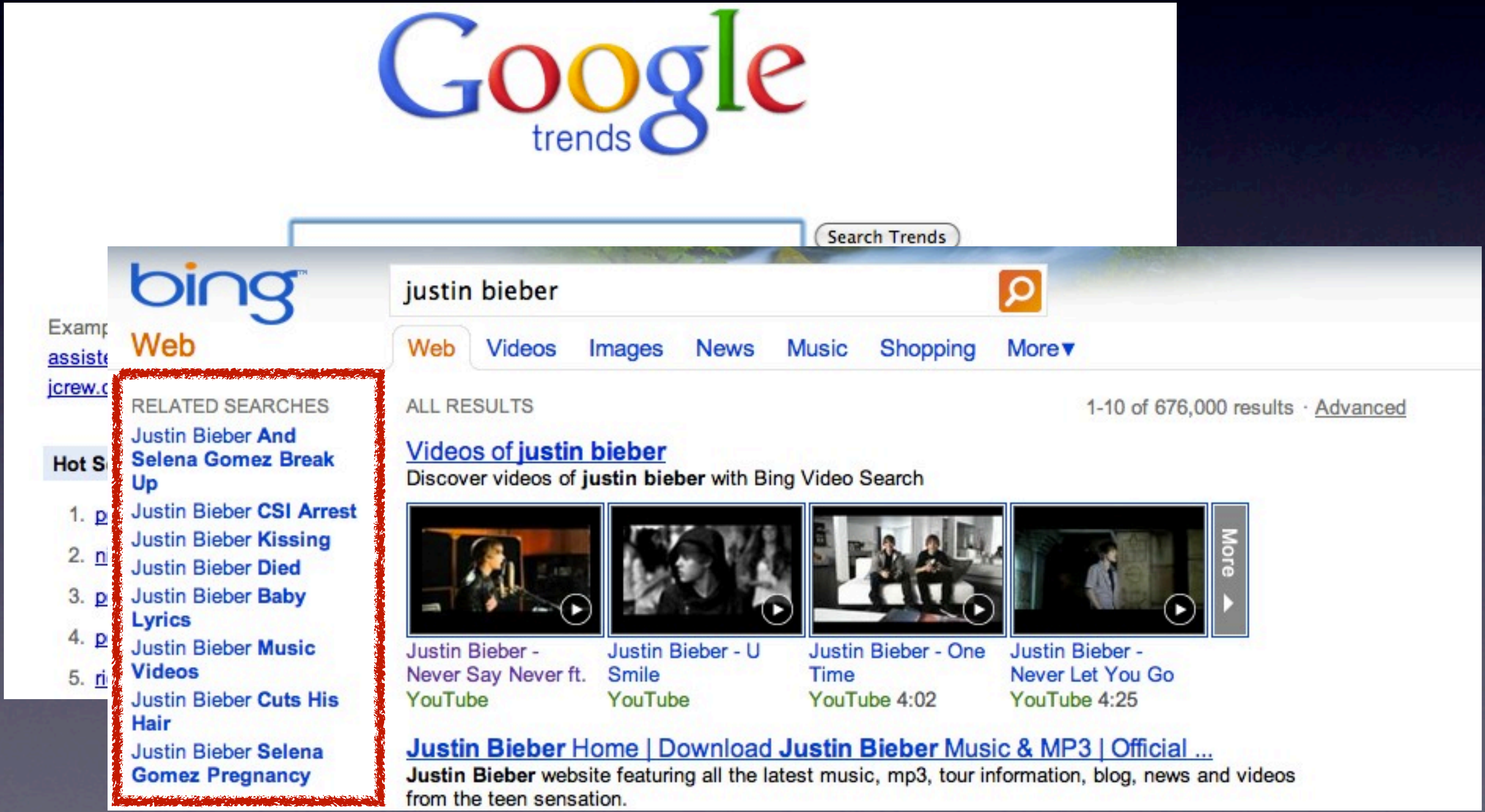
imdb.com

Hot Searches (USA)

1. pga championship
2. nickjr
3. powerball
4. powerball numbers
5. rick perry
6. statue of liberty
7. florida lottery
8. george lopez
9. rebecca black friday
10. powerball results

Poisoned keywords

- 20,000+ popular search terms poisoned



Examp
assist
jcrew.c

Hot S
1. p
2. n
3. p
4. p
5. ri

justin bieber

Web

Web | Videos | Images | News | Music | Shopping | More▾

RELATED SEARCHES

Justin Bieber **And Selena Gomez Break Up**
Justin Bieber **CSI Arrest**
Justin Bieber **Kissing**
Justin Bieber **Died**
Justin Bieber **Baby Lyrics**
Justin Bieber **Music Videos**
Justin Bieber **Cuts His Hair**
Justin Bieber **Selena Gomez Pregnancy**

ALL RESULTS

1-10 of 676,000 results · Advanced

Videos of **justin bieber**
Discover videos of **justin bieber** with Bing Video Search

Justin Bieber - Never Say Never ft.
YouTube

Justin Bieber - U Smile
YouTube

Justin Bieber - One Time
YouTube 4:02

Justin Bieber - Never Let You Go
YouTube 4:25

Justin Bieber Home | Download **Justin Bieber** Music & MP3 | Official ...
Justin Bieber website featuring all the latest music, mp3, tour information, blog, news and videos from the teen sensation.

Poisoned keywords

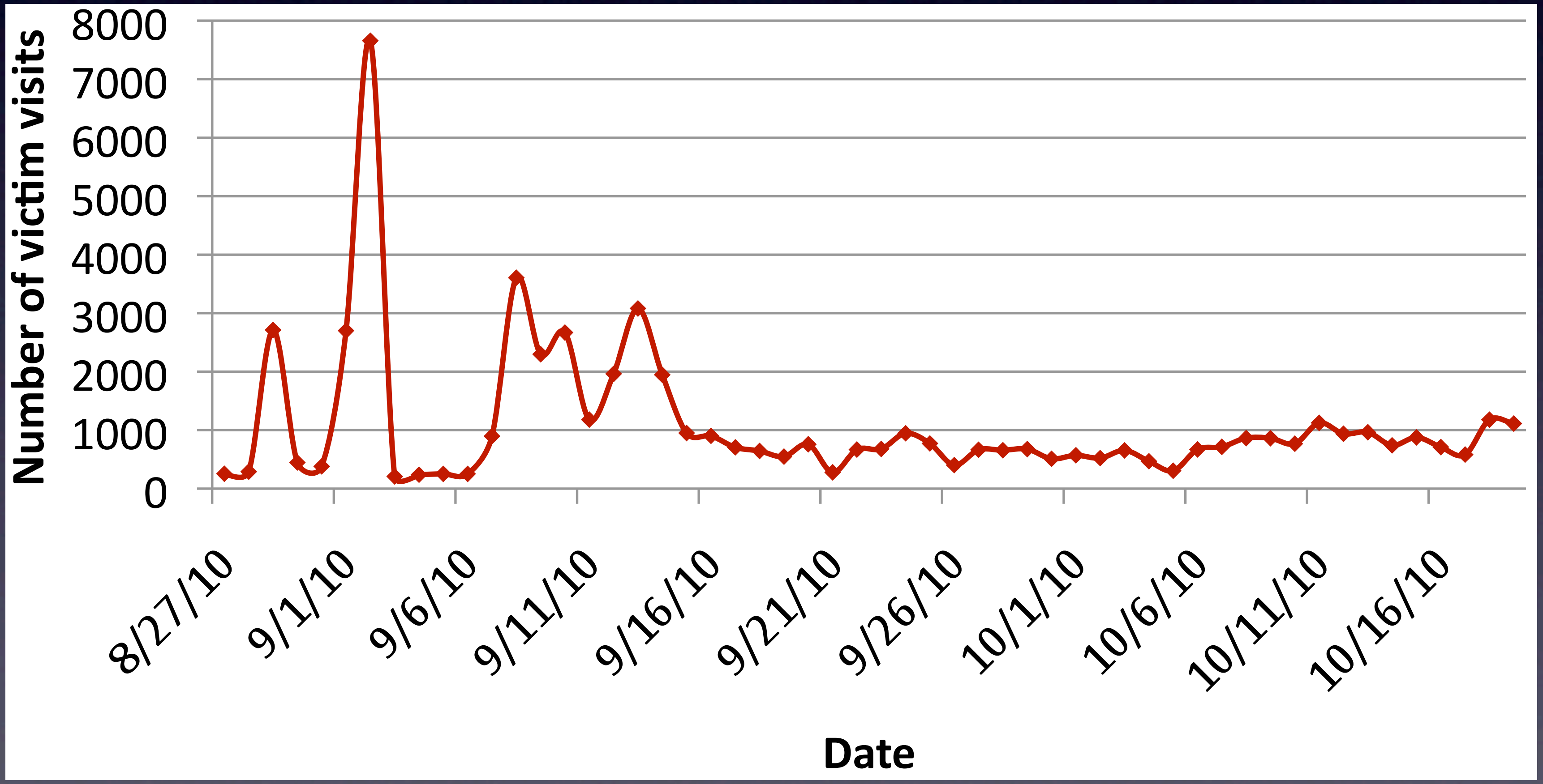
- 20,000+ popular search terms poisoned
- Google Trends + Bing related searches
  - *haiti earthquake*
  - *senate elections*
  - *veterans day 2010*
  - *halloween 2010*
  - *thanksgiving 2010 ...*

Poisoned keywords

- 20,000+ popular search terms poisoned
- Google Trends + Bing related searches
  - *haiti earthquake*
  - *senate elections*
  - *veterans day 2010*
  - *halloween 2010*
  - *thanksgiving 2010 ...*
- 95% of Google Trends keywords poisoned

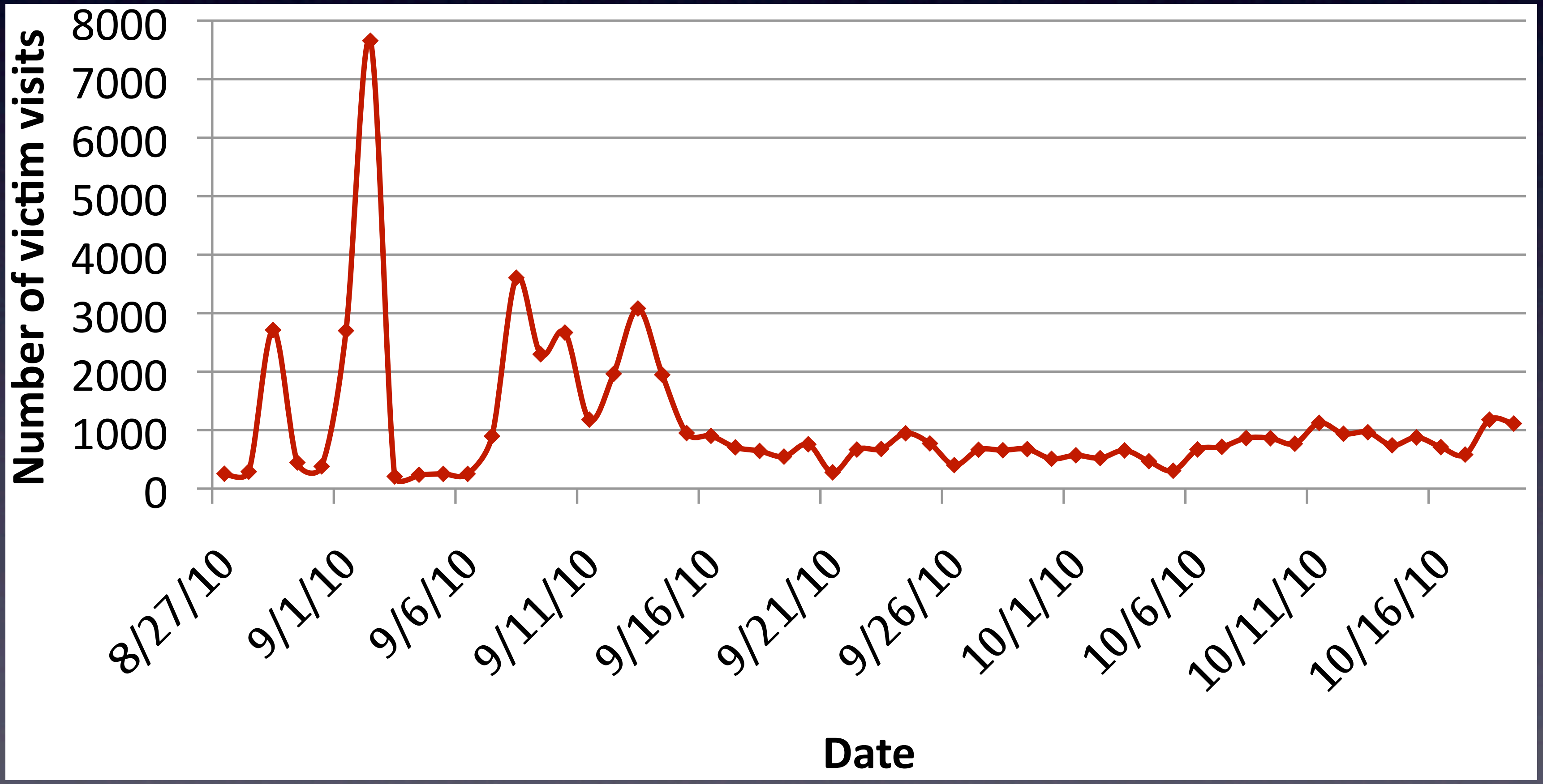
Redirection servers

- Three domains used for redirection
- Over 1,000 exploit URLs fetched



Redirection servers

- Three domains used for redirection
- Over 1,000 exploit URLs fetched



Almost 100,000 victims over 10 weeks

Evasive techniques

- Why can’t redirection behavior be easily detected?
  - Cloaking
  - Requiring user interaction
  - Redirection through javascript or flash

What are prominent features in search poisoning?

- Dense link structure
- Automatic generation of relevant pages
- Large number of pages with popular keywords
- Behavior of compromised sites
  - *before* - diverse content and behavior
  - *after* - similar content and behavior

What are prominent features in search poisoning?

- ~~Dense link structure~~
- ~~Automatic generation of relevant pages~~
- Large number of pages with popular keywords
- Behavior of compromised sites
  - *before* - diverse content and behavior
  - *after* - similar content and behavior

deSEO steps

1. History-based filtering

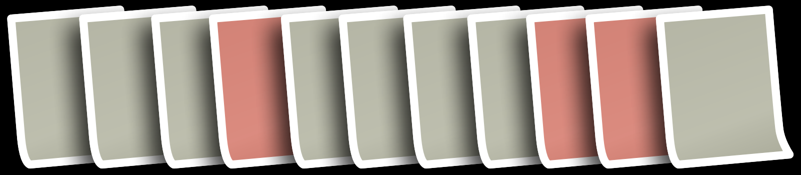
   select domains where many new pages are set up, different from older pages

2. Clustering suspicious domains

   using K-means++

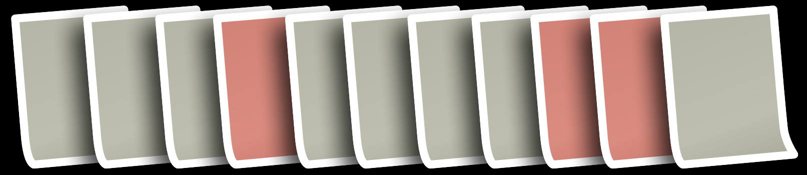
3. Group similarity analysis

   select groups where new pages are similar across domains

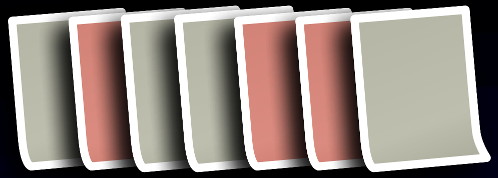


Sample web URLs with trendy keywords

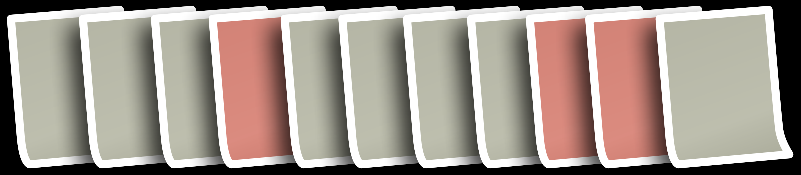
http://www.askania-fachmaerkte.de/images/news.php?page=justin+bieber+breaks+neck



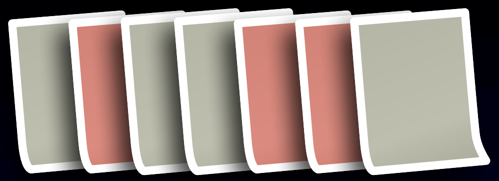
Sample web URLs with trendy keywords



History based detection



Sample web URLs with trendy keywords

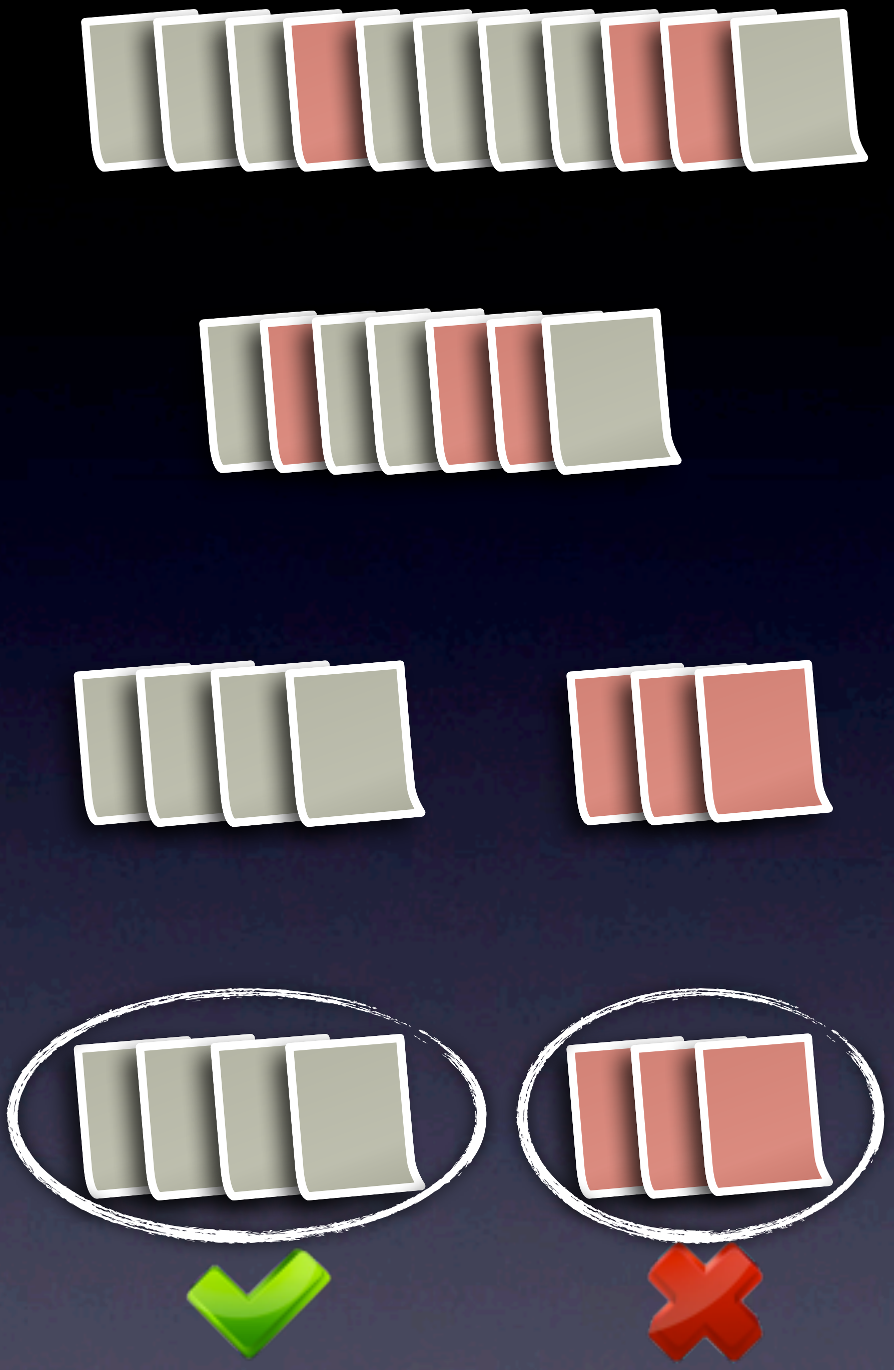


History based detection

Domain clustering

-lexical features of URLs

String features-	keyword separators, arguments, filename, path
Numerical features-	number of arguments, length of arguments, length of keywords
Bag of words-	set of keywords



Sample web URLs with trendy keywords

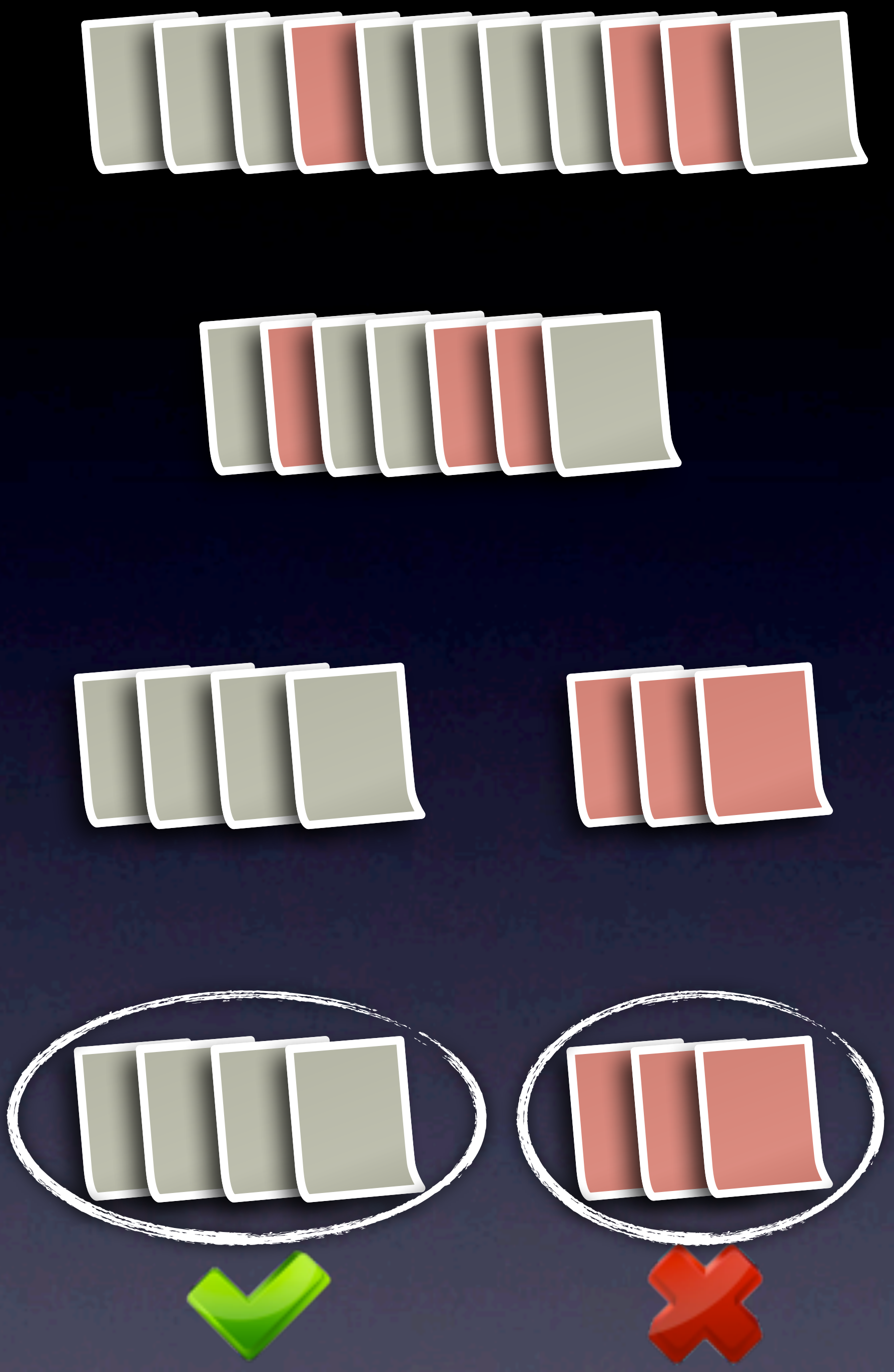
History based detection

Domain clustering

-lexical features of URLs

Group analysis

-web page feature similarity



Sample web URLs with trendy keywords

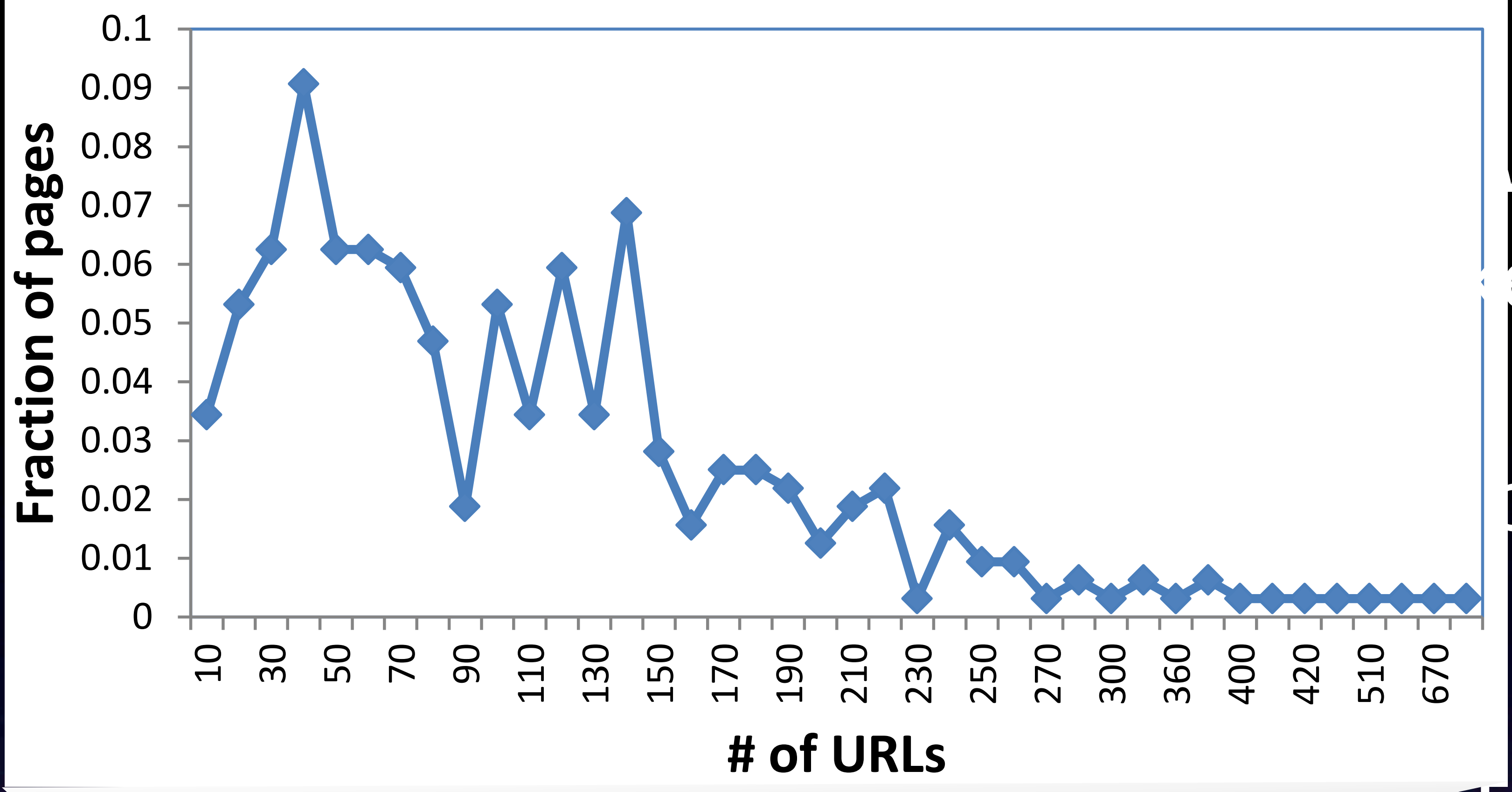
History based detection

Domain clustering

-lexical features of URLs

Group analysis

-web page feature similarity



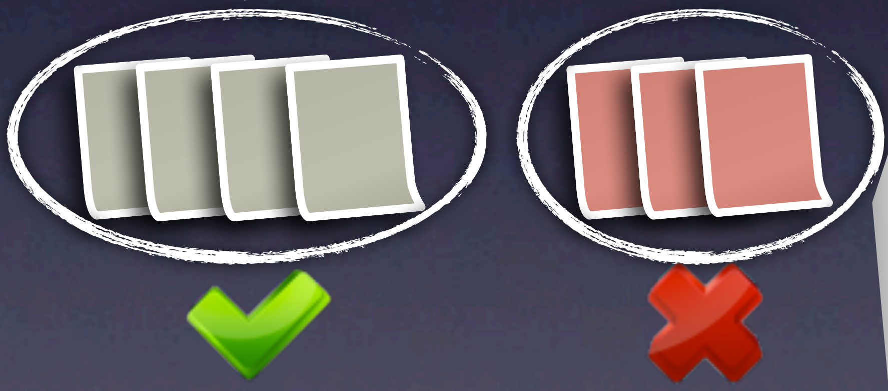
web URLs with

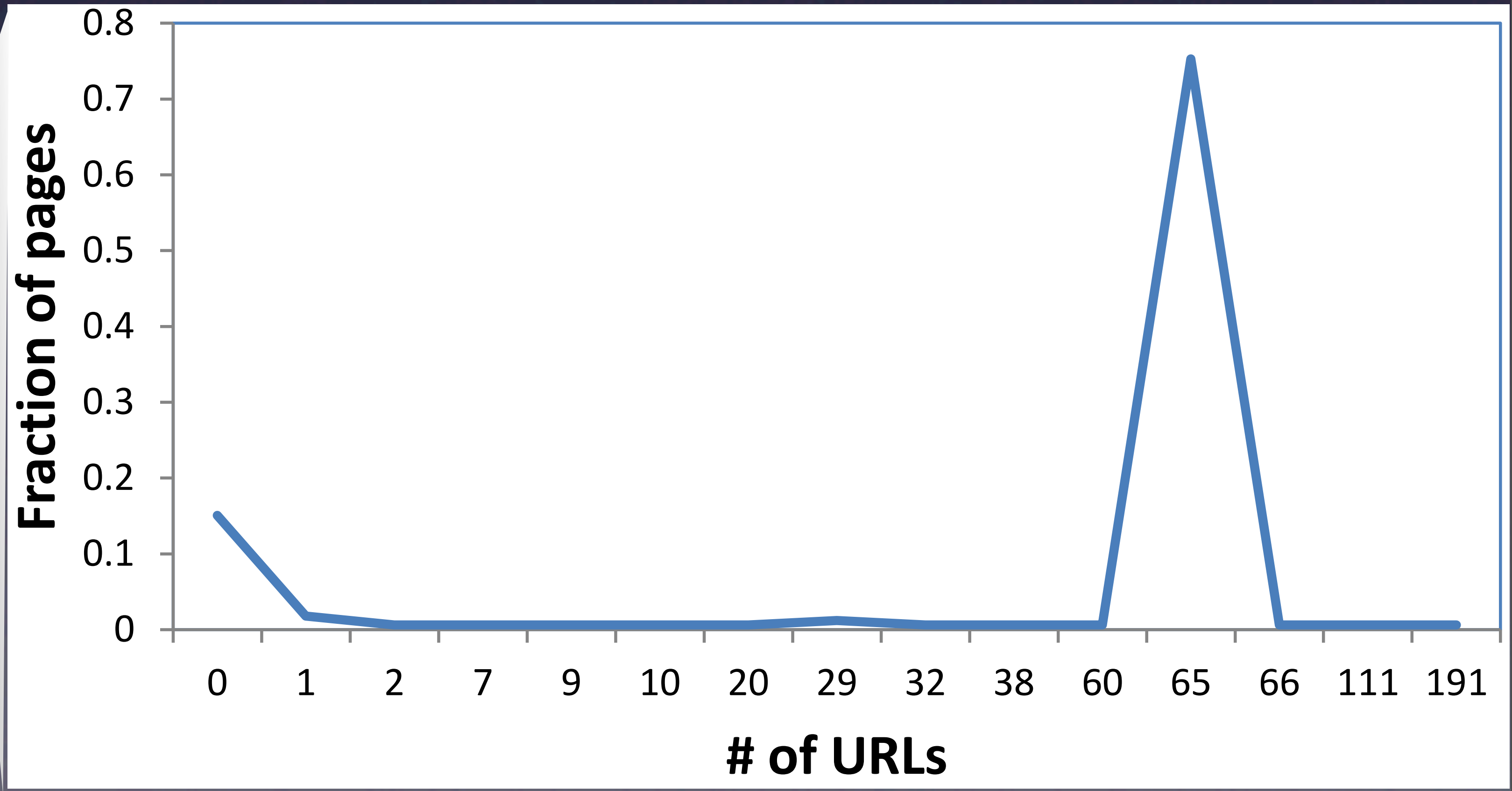
eywords

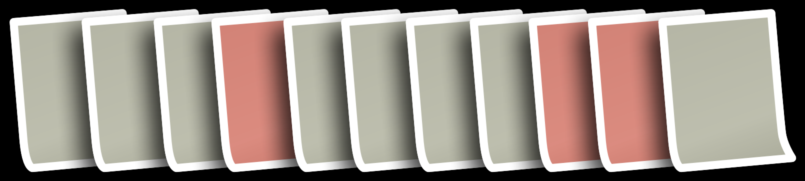
based detection

clustering

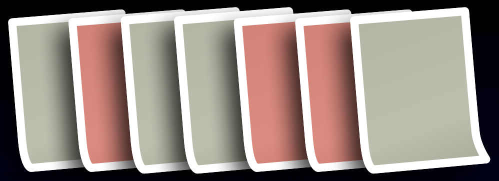
-lexical features of URLs







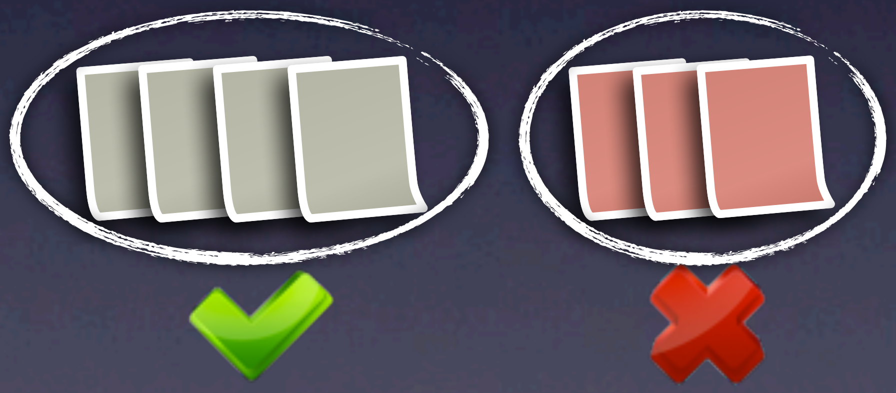
Sample web URLs with trendy keywords



History based detection

Domain clustering

-lexical features of URLs



Group analysis

-web page feature similarity

```
.*\/xmlrpc\.php\/\?showc=\w+(\+\w+)+$
```

Regular expressions

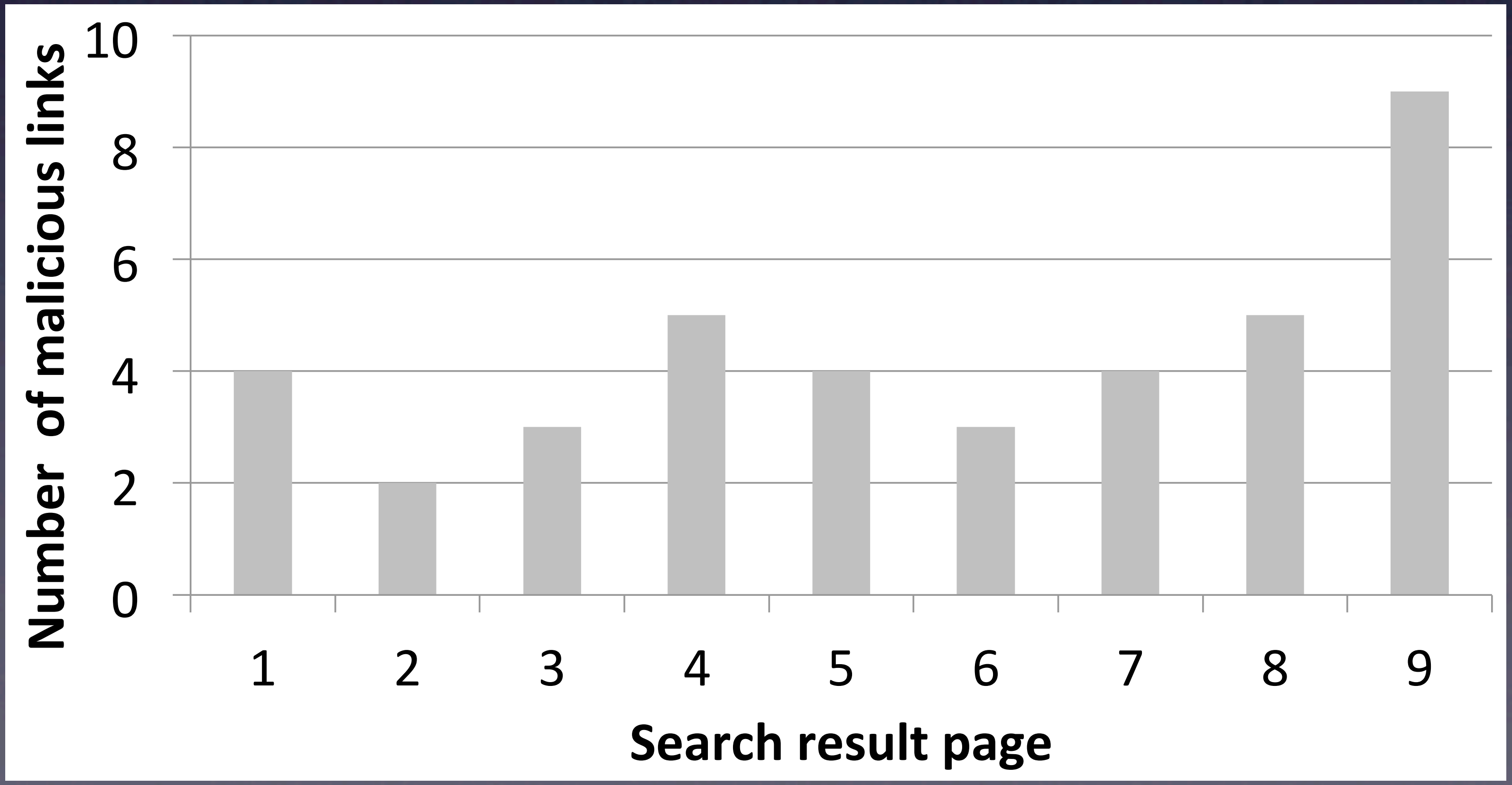
-to match URLs not in our sample

deSEO findings

- 11 malicious groups from sampled web graph in January 2011
  - 957 domains
  - 15,482 URLs
- Revealed a new search poisoning attack
  - compromised Wordpress installations
  - cloaking to avoid detection
  - different link topology

Applying to search results

- 120 keyword searches in Google and Bing
  - 163 malicious URLs detected in results
  - 43 search terms affected



Conclusion

- Malware and SEO are big problems
- Analyzed an ongoing scareware campaign
    - Identified thousands of compromised domains
- Identified prominent features in SEO attacks and used them to build deSEO
    - Promising results on a partial dataset from bing
    - Identified multiple live SEO attacks

Thank You

jjohn@cs.washington.edu