EvilSeed: A Guided Approach to Finding Malicious Web Pages

L. Invernizzi\textsuperscript{1} \quad S. Benvenuti\textsuperscript{2} \quad M. Cova\textsuperscript{3,5} \quad P. Milani Comparetti\textsuperscript{4,5} \quad C. Kruegel\textsuperscript{1} \quad G. Vigna\textsuperscript{1}

\textsuperscript{1}UC Santa Barbara \quad \textsuperscript{2}University of Genova \quad \textsuperscript{3}University of Birmingham \quad \textsuperscript{4}Vienna University of Technology \quad \textsuperscript{5}Lastline, Inc.

IEEE Security & Privacy 2012
Finding malicious URLs

URL

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Landing and exploit pages

My cat's blog 😺
Landing and exploit pages

INJECTED PAGE

My cat's blog

redirection script

Landing and exploit pages

LANDING PAGE
(aka INJECTED PAGE)

EXPLOIT PAGE

redirection script

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Landing and exploit pages

blogs on cats

My cat's blog
My cat's blog, visit me!
Landing and exploit pages

LANDING PAGE
(aka INJECTED PAGE)

My cat's blog

exploitation script

EXPLOIT PAGE

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Landing and exploit pages

LANDING PAGE (aka INJECTED PAGE)

My cat's blog

redirection script

EXPLOIT PAGE

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Landing and exploit pages

LANDING PAGE
(aka INJECTED PAGE)

My cat's blog

redirection script

EXPLOIT PAGE

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Finding malicious URLs

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Finding malicious URLs

CRAWLER

PREFILTER

ORACLE

BENIGN

MALICIOUS

Finding malicious URLs is hard!

Wepawet

- Over 120 thousand URLs analyzed per day by the oracle.
- Available online: http://wepawet.cs.ucsb.edu

The problem

0,138% of the URLs reached with a random crawl are malicious
Our goal
Finding malicious URLs efficiently

WE HAVE
A set of pages that our oracle labeled malicious: our EVIL SEED

WE WANT
More landing pages
More exploit pages
What can a malicious URL tell us?

Similar content

Similar domain registration

Similar link topology

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Gadgets

Evil Seed → Datasets

GADGET

Google™ bing™ DNS

EvilSeed

Wepawet, Google Safebrowsing, FakeAv

EvilSeed

Google
bing
DNS queries
Domain registrations

[Links gadget]
[Content Dorks gadget]
[SEO gadget]
[Domain Registrations gadget]
[DNS gadget]

Malicious Seed URLs
Links gadget
Designed to locate *malware hubs*

Example query: `link:http://malicious-url.com`

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Links gadget

Sign Guestbook

Display 10 per page

There are 59674 messages

Sort by Last entry first ▼

buy cigarettes online
New York
privet prostitutki wluix pidari, buy cigarettes online usa, joyeoa,
Date: Friday, May 18, 2012

cheap prevacid
hCsOmToYLpJKUcBM
comment4, buy Cialis online, 8][, generic Zithromax, =PP, generic Proventil, 10108,
generic deltasone, gebwk, cheap prevacid, qudpq,
Date: Friday, May 18, 2012

Ebook The Zombie Survival Guide by Max Brooks download online
EYMBcBwsA2xM

http://www.formspring.me/ebookZombie ebook pda reader The Zombie Survival Guide by Max Brooks
Date: Friday, May 18, 2012

unrerceHoasse
Listing pharmacists and step is Uses Of Ultram Or Tramadol ran. <a href=
http://pharmacycipetraeus.com/products/pain_relief/tramadol >order generic tramadol online without a prescription</a> buy ultram overnight. <a href="http://pharmacycipetraeus.com/products/pain_relief/tramadol >order generic tramadol online</a> Tell your doctor and Askville is the most commonly reported in
Date: Friday, May 18, 2012

Content Dorks gadget

Creates signatures from the content of landing pages.

Two methods:

- **n-gram extraction**
- **term-extraction** (e.g., cnn.com yields: Eurozone recession, gay wedding, Facebook attack, graphic content)
Content Dorks gadget

"calendar about pregnancy"

About 189,000 results (0.35 seconds)

Buttons2
www.rhiossampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Chris Duffield home page
iptq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

mouth exact symbol - LineoneLabsUSA
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.

... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.
"calendar about pregnancy"

About 189,000 results (0.35 seconds)

Buttons2
www.rhiossampler.net/Buttons2.htm

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

Chris Duffield home page
iptq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

mouth exact symbol - LineoneLabsUSA
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.
... classes at Massachusetts College of Art, and the University of Massachusetts at ... childhood, including a week by week pregnancy calendar about pregnancy.
Content Dorks gadget

Buttons2
www.rhiosampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Chris Duffield home page
iptq.com/cd/

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

mouf11 exa1m sym6ol - L1neone1ab1sUSA
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.

... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Content Dorks gadget

"calendar about pregnancy"

About 189,000 results (0.35 seconds)

Buttons2
www.rhiosampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy.

Chris Duffield home page
iptq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy.

mouth exact symbol - LineoneLabsUSA
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy.

BigZanda Gallery: Surf Photo New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.

... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.
Content Dorks gadget

"calendar about pregnancy"

About 189,000 results (0.35 seconds)

**Buttons2**
[www.rhiossampler.net/Buttons2.htm](http://www.rhiossampler.net/Buttons2.htm)
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

**Chris Duffield home page**
[iptq.com/cd/](http://iptq.com/cd/)

**Bigzanda Gallery: Surf Photo-New England & Beyond**

**This site may harm your computer.**

... classes at Massachusetts College of Art, and the University of Massachusetts at ...
childbirth, including a week by week pregnancy calendar about pregnancy.

[EvilSeed](http://bit.ly/evilseed)
SEO gadget

User Agent: Google

www.seo.com

cheap hotels, free iPhones, iPad 4, free streaming movies, cheap Facebook stocks, wallpapers

User Agent: Firefox

www.seo.com

Cute kittens pictures

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Expansion strategies:

■ Find pages with similar content as Google sees it (e.g., query for title:"free iphones")
■ Find pages hosted on the same domain (e.g., query for site:seo.com)
■ Follow links

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

Domain Registrations gadget

We know that:

- a.com has been registered moments before b.com

We suspect that:

- http://b.com/exploit is also malicious
DNS Queries gadget

Visit web site  www.rotarynewalipore.in  aquarigger.com

Check mail  mx1.hotmail.com

System update  windowsupdate.com

time
DNS Queries gadget

domain is in our Evil Seed (hosts an exploit URL)

Visit web site
Check mail
System update

www.rotarynewalipore.in
mx1.hotmail.com
windowsupdate.com

aquarigger.com

time
DNS Queries gadget

Gadget discovers that this domain hosts a landing page that redirects to aquarigger.com

Visit web site: www.rotarynewalipore.in
Check mail: mx1.hotmail.com
System update: windowsupdate.com

time
Evaluation metrics

**Toxicity** = \( \frac{\text{URLs classified as malicious}}{\text{URLs submitted to the Oracle}} \)

**Seed Expansion** = \( \frac{\text{malicious URLs found by EvilSeed}}{\text{seed size}} \)

## Online evaluation: URLs

<table>
<thead>
<tr>
<th>Source</th>
<th>Seed</th>
<th>Analyzed</th>
<th>Malicious</th>
<th>Toxicity</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crawler w/ Prefilter</strong></td>
<td>437,251</td>
<td>604</td>
<td></td>
<td>0.14%</td>
<td></td>
</tr>
<tr>
<td><strong>EvilSeed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Links</td>
<td>604</td>
<td>71,272</td>
<td>1,097</td>
<td>1.53%</td>
<td>1.81</td>
</tr>
<tr>
<td>SEO</td>
<td>604</td>
<td>312</td>
<td>16</td>
<td>5.12%</td>
<td>0.02</td>
</tr>
<tr>
<td>Keywords</td>
<td>604</td>
<td>13,896</td>
<td>477</td>
<td>3.43%</td>
<td>0.78</td>
</tr>
<tr>
<td>Ngrams</td>
<td>604</td>
<td>140,660</td>
<td>1,446</td>
<td>1.02%</td>
<td>2.39</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>226,140</td>
<td>3,036</td>
<td></td>
<td>1.34%</td>
<td>5.02</td>
</tr>
<tr>
<td><strong>Web Search</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Strings</td>
<td>24,137</td>
<td>68</td>
<td></td>
<td>0.28%</td>
<td></td>
</tr>
<tr>
<td>Random Dictionary</td>
<td>27,242</td>
<td>107</td>
<td></td>
<td>0.39%</td>
<td></td>
</tr>
<tr>
<td>Trending Topics</td>
<td>8,051</td>
<td>27</td>
<td></td>
<td>0.33%</td>
<td></td>
</tr>
<tr>
<td>Manual Dorks</td>
<td>4,506</td>
<td>17</td>
<td></td>
<td>0.37%</td>
<td></td>
</tr>
</tbody>
</table>
### Online evaluation: domains

<table>
<thead>
<tr>
<th>Source</th>
<th>Seed</th>
<th>Analyzed</th>
<th>Malicious</th>
<th>Toxicity</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crawler w/ Prefilter</td>
<td>53,445</td>
<td>98</td>
<td></td>
<td>0.18%</td>
<td></td>
</tr>
<tr>
<td>EvilSeed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Links</td>
<td>98</td>
<td>7,664</td>
<td>107</td>
<td>1.39%</td>
<td>1.09</td>
</tr>
<tr>
<td>SEO</td>
<td>98</td>
<td>7</td>
<td>5</td>
<td>71.42%</td>
<td>0.07</td>
</tr>
<tr>
<td>Keywords</td>
<td>98</td>
<td>3,245</td>
<td>119</td>
<td>3.66%</td>
<td>1.22</td>
</tr>
<tr>
<td>Ngrams</td>
<td>98</td>
<td>33,510</td>
<td>263</td>
<td>0.78%</td>
<td>2.68</td>
</tr>
<tr>
<td>Total</td>
<td>44,426</td>
<td>494</td>
<td></td>
<td>1.12%</td>
<td>5.04</td>
</tr>
<tr>
<td>Web Search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Strings</td>
<td>4,227</td>
<td>16</td>
<td></td>
<td>0.37%</td>
<td></td>
</tr>
<tr>
<td>Random Dictionary</td>
<td>9,285</td>
<td>35</td>
<td></td>
<td>0.37%</td>
<td></td>
</tr>
<tr>
<td>Trending Topics</td>
<td>1,768</td>
<td>8</td>
<td></td>
<td>0.45%</td>
<td></td>
</tr>
<tr>
<td>Manual Dorks</td>
<td>3,032</td>
<td>13</td>
<td></td>
<td>0.42%</td>
<td></td>
</tr>
</tbody>
</table>
DNS evaluation

Data:
- 377,472,280 DNS resolutions
- 115 malicious seeds

Resulting in 3.5% toxicity, 1.48 seed expansion
"calendar about pregnancy"

About 189,000 results (0.35 seconds)

**Buttons2**
www.rhiossampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy **calendar about pregnancy**. Click for the ...

**Chris Duffield home page**
iptq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy **calendar about pregnancy**. Click for the ...

**mouth exact symbol - LineoneLabsUSA**
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy **calendar about pregnancy**. Click for the ...

**Bigzanda Gallery: Surf Photo-New England & Beyond**
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.
... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy **calendar about pregnancy**
EvilSeed for Search Engines

"calendar about pregnancy"

About 189,000 results (0.35 seconds)

Buttons2
www.rhiossampler.net/Buttons2.htm

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

Chris Duffield home page
iptq.com/cd/

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

mouth exact symbol - LineoneLabsUSA
lineonelabsusa.com/public_html/te_st.html

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ..

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html

This site may harm your computer.

... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.
EvilSeed for Search Engines

"calendar about pregnancy"

About 189,000 results (0.35 seconds)

Buttons2
www.rhiossampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Chris Duffield home page
intq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

mouth exact symbol - LineOneLabsUSA
lineonelabsusa.com/public_html/te_st.html
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.
... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.

L. Invernizzi, S. Benvenuti, M. Cova, P. Milani Comparetti, C. Kruegel, G. Vigna

EvilSeed for Search Engines

"calendar about pregnancy"
About 189,000 results (0.35 seconds)

Buttons2
www.rhiossampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the...

Chris Duffield home page
iptq.com/cd/
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the...
mouth exact symbol - LineoneLabsUSA
lineonelabsusa.com/public_html/te_st.html

The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the...

Bigzanda Gallery: Surf Photo New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html
This site may harm your computer.
... classes at Massachusetts College of Art, and the University of Massachusetts at ...
childbirth, including a week by week pregnancy calendar about pregnancy.
EvilSeed for Search Engines

"calendar about pregnancy"

Buttons2
www.rhiossampler.net/Buttons2.htm
The pregnancy guide can help you find information on pregnancy and childbirth, including a week by week pregnancy calendar about pregnancy. Click for the ...

Chris Duffield home page
iptq.com/cd/

Bigzanda Gallery: Surf Photo-New England & Beyond
www.daleratcliff.com/bigzanda/surf_photo/index.html

This site may harm your computer.

... classes at Massachusetts College of Art, and the University of Massachusetts at ... childbirth, including a week by week pregnancy calendar about pregnancy.
Conclusions

- Finding malicious URLs is important to protect the users
Conclusions

- Finding malicious urls is important to protect the users, but it’s hard
Conclusions

- Finding malicious urls is important to protect the users, but it’s hard.
- It’s critical to generate feeds with high toxicity (⇒ high efficiency).
Conclusions

- Finding malicious urls is important to protect the users, but it’s hard
- It’s critical to generate feeds with high toxicity (⇒ high efficiency)
- We designed EvilSeed, a guided search approach that is a ten-fold efficiency improvement over crawling
Conclusions

- Finding malicious urls is important to protect the users, but it’s hard
- It’s critical to generate feeds with high toxicity (⇒ high efficiency)
- We designed EvilSeed, a guided search approach that is a ten-fold efficiency improvement over crawling
- But crawling is needed nonetheless, to generate the evil seed

Thanks!


invernizzi@cs.ucsb.edu
Thanks!


invernizzi@cs.ucsb.edu
## SEO Evaluation

<table>
<thead>
<tr>
<th>URLs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloaking Seeds</td>
<td>248</td>
</tr>
<tr>
<td>Visited</td>
<td>1,219,090</td>
</tr>
<tr>
<td>Analyzed</td>
<td>12,063</td>
</tr>
<tr>
<td>Malicious</td>
<td>11,384</td>
</tr>
<tr>
<td><strong>Toxicity</strong></td>
<td><strong>94.37%</strong></td>
</tr>
<tr>
<td>Malicious/Visited</td>
<td>0.93%</td>
</tr>
<tr>
<td>Expansion</td>
<td>45.90</td>
</tr>
</tbody>
</table>

(Crawler’s: 0.14%)