

Google Dorks: Analysis, Creation, and new Defenses

Flavio Toffalini, University of Verona, IT, flavio.toffalini@gmail.com

Maurizio Abbà, LastLine, UK, mabba@lastline.com

Damiano Carra, University of Verona, IT, damiano.carra@univr.it

Davide Balzarotti, Eurecom, FR, davide.balzarotti@eurecom.fr

GOOGLE DORKS



intext:db_password ext:env



inurl:wp-content -remove -exploit index-of



Web

Images

Videos

News

Shopping

More ▼

Search tools

About 980,000 results (0.43 seconds)

Index of /wp-content - Sojourn Music

www.sojournmusic.com/wp-content/ ▼

Index of /wp-content. Parent Directory | [plugins/](#) · [themes/](#) · [upgrade/](#) · [uploads/](#)

MOTIVATION

- Attackers use Dorks to quickly locate targets
 - After a new vulnerability is disclosed, one Google query is sufficient to identify a large amount of vulnerable installations
 - No time for sysadmins to apply patches !!

MOTIVATION

- Attackers use Dorks to quickly locate targets
 - After a new vulnerability is disclosed, one Google query is sufficient to identify a large amount of vulnerable installations
 - No time for sysadmins to apply patches !!
- If we could prevent dorks, attackers would need to resort to Internet scanning ... which is several orders of magnitude slower

GOALS

- **Current practices**

- Understand which information is used by existing dorks
- Design simple solutions to defeat those dorks

- **Future threats**

- Test if attackers could move towards new styles of dorks
- Design simple solutions to prevent it

GOOGLE DORKS

Best Blog on WordPr... x

https://bestblog.wordpress.com Cerca

Best Blog on Wordpress

The Unofficial Blog Review Site for Wordpress.com

Blog About Contact Editors What's New! FAQ

Elle Effect

Published 13 June, 2007 art, arts, Christianity, faith, ideas, life, photo, Religion, Spirituality, ...

I have been sitting on reporting this best blog has kept me distracted) but now it is time to let

About a week and a half ago I was browsing a across [Elle Effect](#). Lauren, the author, is an ei glimpse into her artistic works but also into the

What I quickly discovered is that this young girl p end of life through both the visual arts as well i

All this is why I consider [Elle Effect](#) a best blog

A Slice of Life

Published 26 April, 2007 general, humor, life, pl

With down-to-earth and great with a goo

About

Best Blog on Wordpress searches for and reviews the best blogs on wordpress.com We also write articles to guide new and old

I felt my heart clench. I think about explaining that the rule ought to be: "Write what you want whatever you know just a little bit more about than your reader." But what I want to tell them is that these rules aren't, after all, rules for writers; they're rules for people who are trying to be writers but won't ever make it. -Peter Ho Davis *What You Know*

Different Schools of Thought:

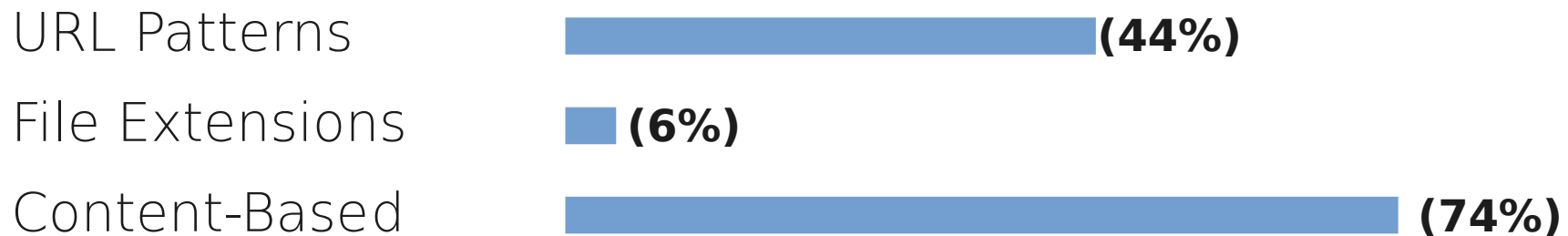
I was reading through the various different blogs on the Internet and came across Daren Rowse's site [problogger](#). He's actually a very well known blogger who writes about how to be a pro blogger. His articles feature tips on how to SEO, make the most out of your Google

Best Blog on Wordpress is more productive.

Blog at WordPress.com. The K2-lite Theme.

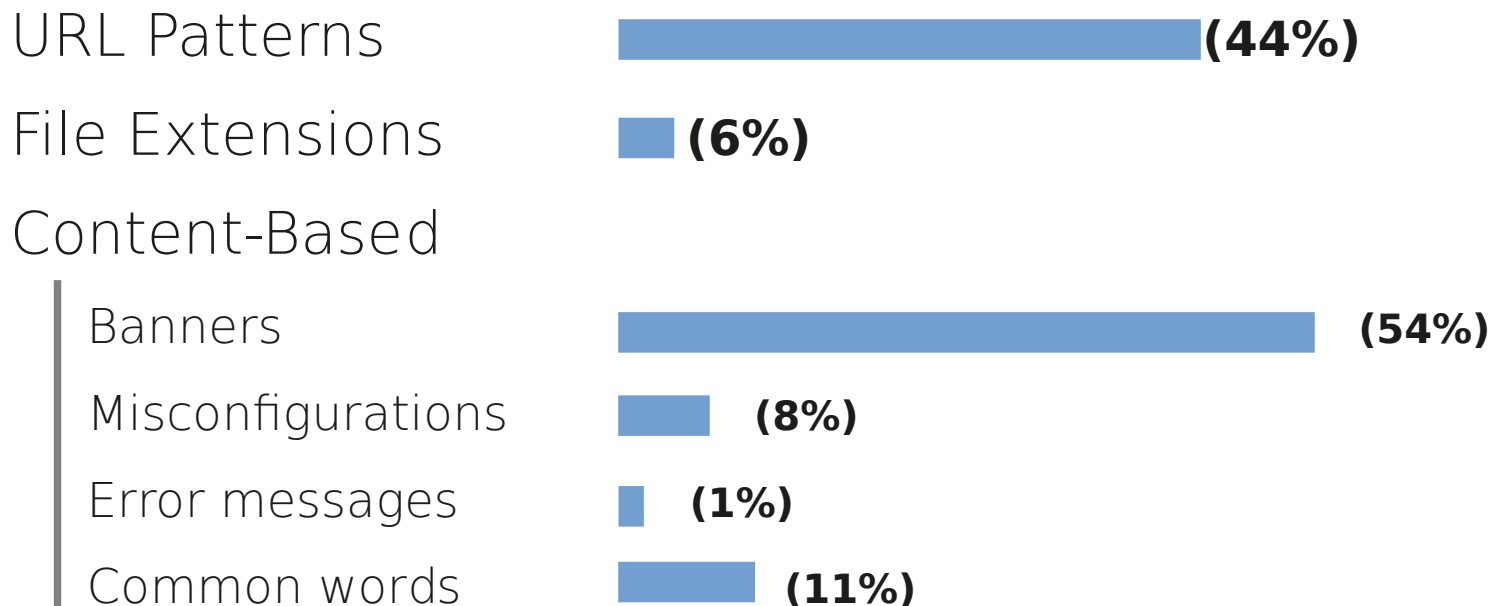
TAXONOMY

- The Exploit-DB database contains over 5143 dorks
- Automated/manual analysis

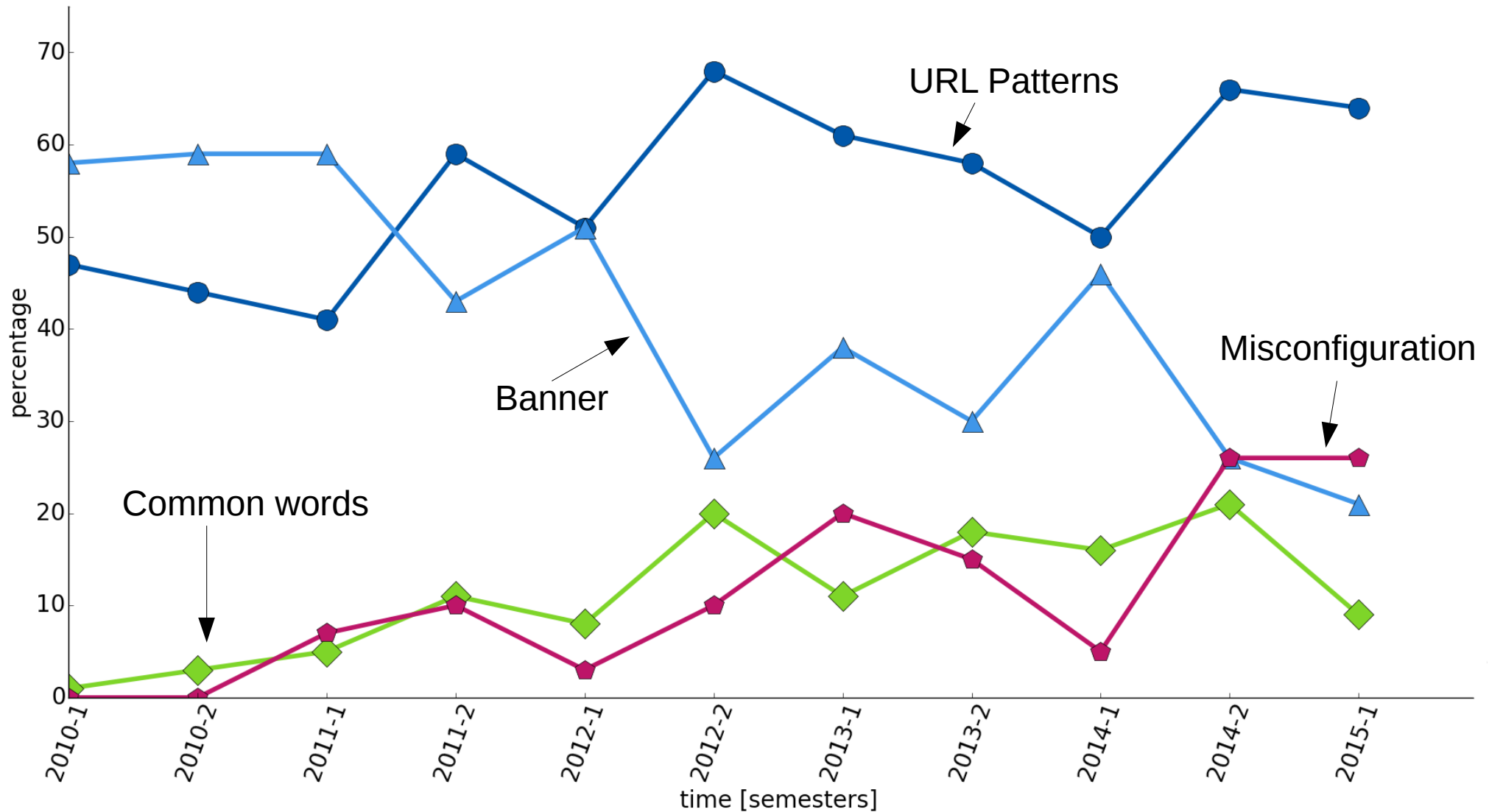


TAXONOMY

- The Exploit-DB database contains over 5143 dorks
- Automated/manual analysis



DORKS EVOLUTION BY CATEGORY



KNOWN DEFENSES

URL Patterns



File Extensions



Content-Based

Banners



→ **remove banners**

Misconfigurations



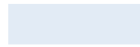
→ **improve system configuration**

Error messages



→ **proper error handling**

Common words



CONTRIBUTION

URL Patterns



File Extensions



Content-Based

Banners



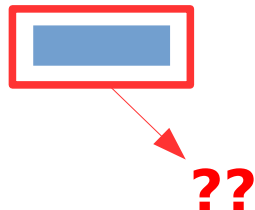
Misconfigurations



Error messages



Common words



URL-DORKS

- Force search engines to index “randomized” URLs
- Let the users navigate and share using cleartext URLs

`http://www.web-site.com/wp-content/dimva.html`



`http://www.web-site.com/HD12DAF35TR/dimva.html`

[\[PDF\] Flavio Toffalini – Curriculum Vitae](#)

www.flaviotoffalini.info/HEAD46421e575a58435d5744TRLR/./cv.pdf ▼

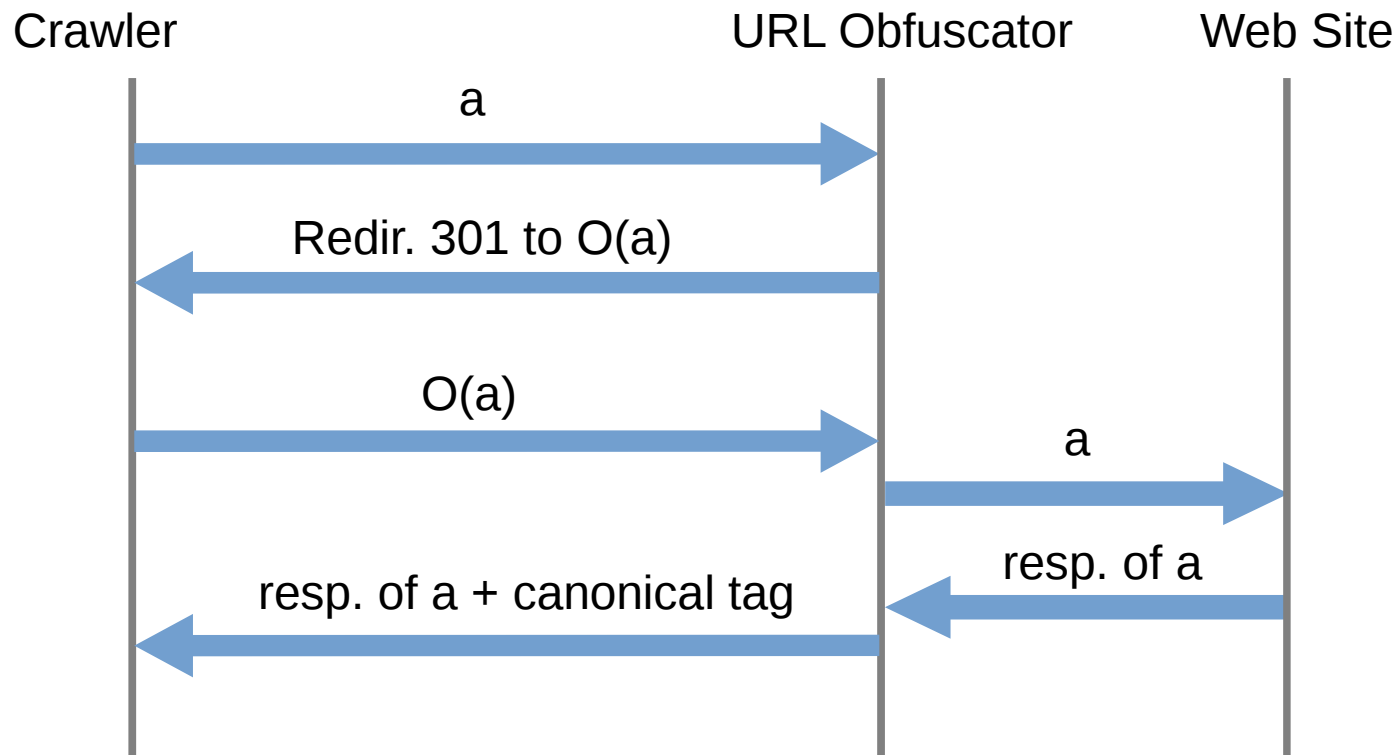
Flavio Toffalini. Curriculum Vitae. Education. Sept 2012. Present. Master degree in Computer Science and Engineering, University of Verona, Graduation.

You've visited this page 5 times. Last visit: 9/8/15

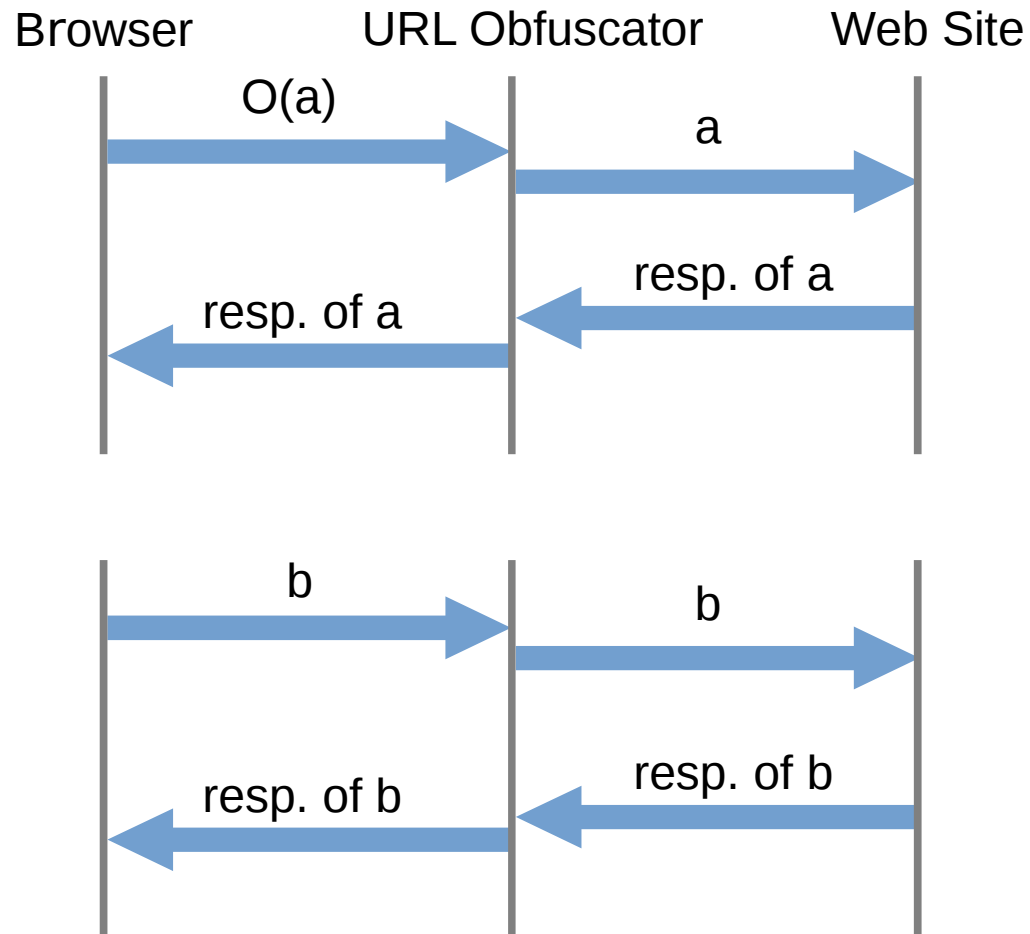
URL-DORKS

- **XOR** (part of) URLs with random seed kept in the server
a = resource a
O(a) = obfuscated resource a
- **Redirect 301** to inform search engine that the page is moved
- **Canonical URL Tag** to delete plain URLs in the results
- Intercept and replace **SiteMap**

OBFUSCATION PROTOCOL - CRAWLERS



OBFUSCATION PROTOCOL - BROWSER



URL Patterns

File Extensions

Content-Based

Banners

Misconfigurations

Error messages

Common words



remove banners



improve system configuration



proper error handling



??

WORD-BASED DORKS

- **Goal**
 - Using words left by CMSs to create a Google Dork
- **Greedy search algorithm to maximizes**
 - **Hit-rank**: percentage of web site made by a target technology
 - **Coverage**: number of entries extracted by the Dork

WORD-BASED DORKS: CREATION

category uncategoris... x +

www.bing.com/search?q=category+ur IIS Cerca

category uncategorised written article super user

Web Images Videos Maps News Explore

2.970.000 RESULTS Date Language Region

Articles - J. L. Mannisto . com
[jlmannisto.com/index.php/component/content/article/2-uncategorised/...](#)
Category: **Uncategorised**. Published Date **Written by Super User** My article on "information diets" and the importance of disconnection, made ...

List All Categories - atizapan.gob.mx
[www.atizapan.gob.mx/.../list-all-categories/2-uncategorised?start=10](#)
List All Categories; **Uncategorised**; Blog. Home. **Written by Super User** on 02 May 2011. Posted in **Uncategorised**. ... Featured Articles; List All Categories; Category ...

Omulele & Tollo Advocates
[ot-advocates.com/index.php/component/content/category/9-uncategorised](#)
... you can easily create **category**-specific content types, e.g. **article**, ... Category: **Uncategorised** ... **Written by Super User** . Category ...

Joomla!

WORD-BASED DORKS: CREATION



Vanilla
installation

Categories Buy
Recent
Register Submit
Users Contact Registration
List

Compute **hit rank**
& **coverage**

“Category” +
“Submit” +
“....”



WORD-BASED DORKS: CREATION

- Gradient Ascent algorithm
- How to add a new word?
 - At each step, we add the word that provides the highest hit rank between the ones that have a coverage above the median of all candidate words
(more details in the paper)

WORD-BASED DORKS:

	Common Words	Ground Truth	
WordPress	938/1000	967/1000	Hit rank
	47.1 M	83.6 M	Coverage
Joomla!	878/1000	887/1000	Hit rank
	7.24 M	3.73 M	Coverage
Drupal	827/1000	997/1000	Hit rank
	7.87 M	3.27 M	Coverage
Magento	871/1000	852/1000	Hit rank
	0.39 M	0.68 M	Coverage
OpenCart	891/1000	998/1000	Hit rank
	0.59 M	1.42 M	Coverage

WORD-BASED DORKS:

	Common Words	Ground Truth	
WordPress	938/1000	967/1000	Hit rank
	47.1 M	83.6 M	Coverage
Joomla!	878/1000	887/1000	Hit rank
	7.24 M	3.73 M	Coverage
Drupal	827/1000	997/1000	Hit rank
	7.87 M	3.27 M	Coverage
Magento	871/1000	852/1000	Hit rank
	0.39 M	0.68 M	Coverage
OpenCart	891/1000	998/1000	Hit rank
	0.59 M	1.42 M	Coverage

WORD-BASED DORKS:

	Common Words	Ground Truth	
WordPress	938/1000	967/1000	Hit rank
	47.1 M	83.6 M	Coverage
Joomla!	878/1000	887/1000	Hit rank
	7.24 M	3.73 M	Coverage
Drupal	827/1000	997/1000	Hit rank
	7.87 M	3.27 M	Coverage
Magento	871/1000	852/1000	Hit rank
	0.39 M	0.68 M	Coverage
OpenCart	891/1000	998/1000	Hit rank
	0.59 M	1.42 M	Coverage

WORD-BASED DORKS: DEFENSES

Idea: add invisible characters to break words and prevent them to be indexed.

Powered by WordPress



Power⁣ed b⁣y Wor⁣dPress

DORKS DEFENSES

URL Patterns



File Extensions



Content-Based

Banners



remove banners

Misconfigurations



improve system configuration

Error messages



proper error handling

Common words



CONCLUSION

- 1) Dork classification
- 2) URL Pattern Dork Defense
- 3) New type of Dork using common words
- 4) Defense against common word dorks