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

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 !!

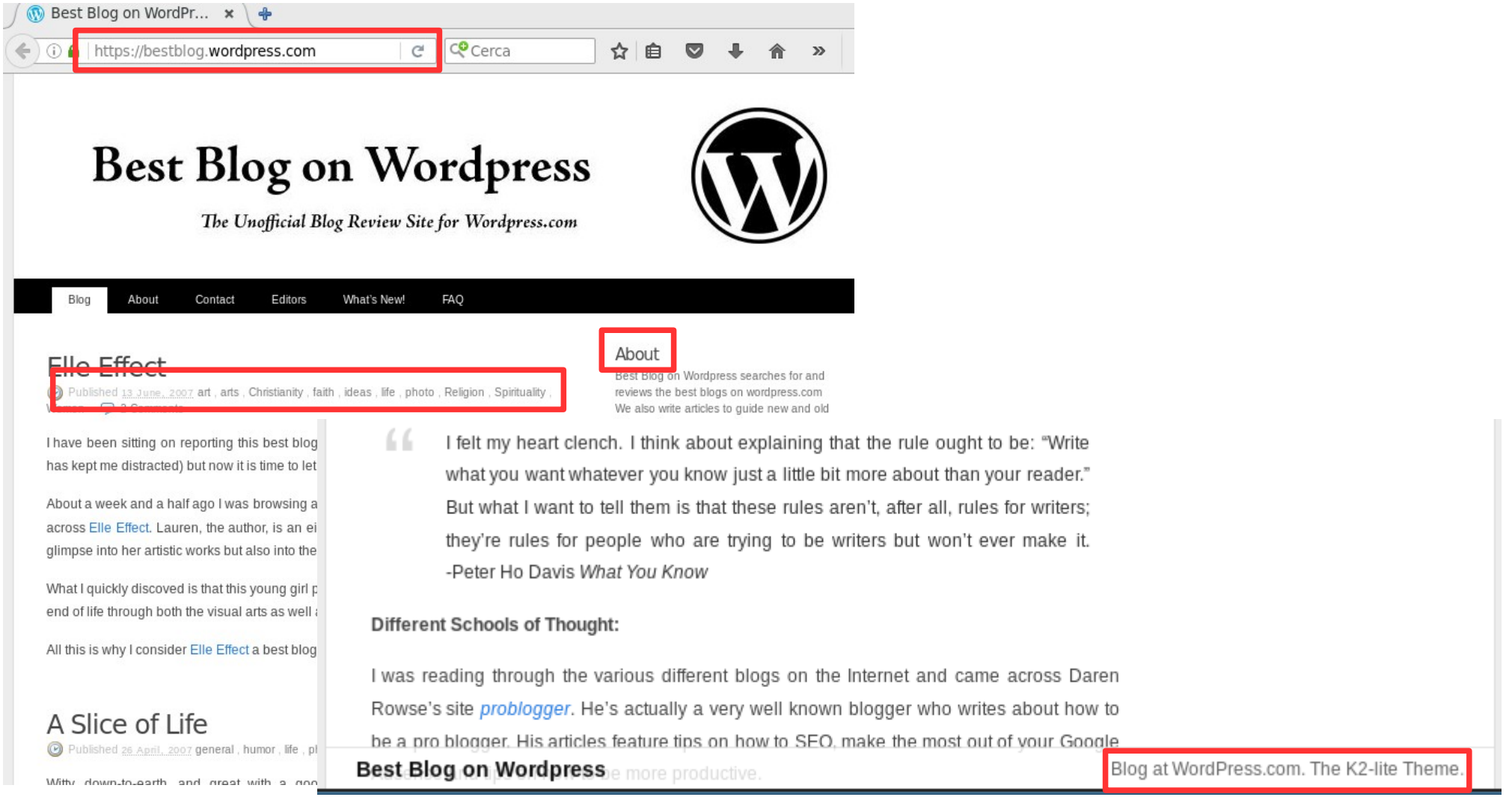
MOTIVATION

- 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

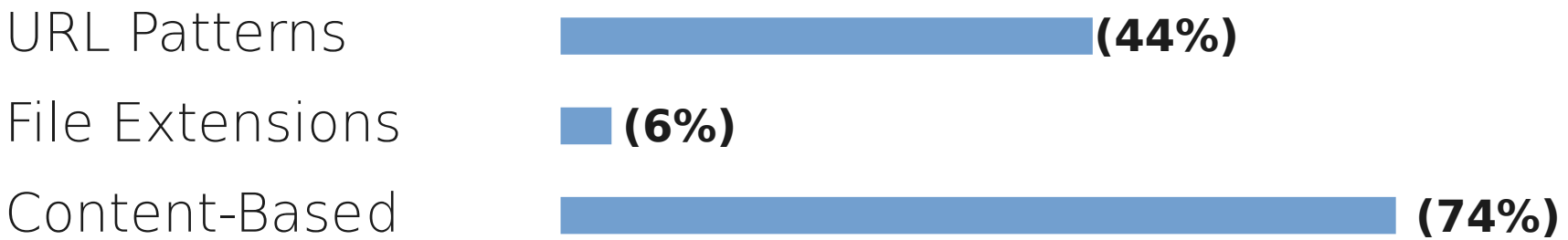
- Current practices
  - Understand which information is used by existing dorks
  - Design simple solutions to defeat those dorks
- Future threats
  - Test if attackers could move towards new styles of dorks
  - Design simple solutions to prevent it

GOOGLE DORKS



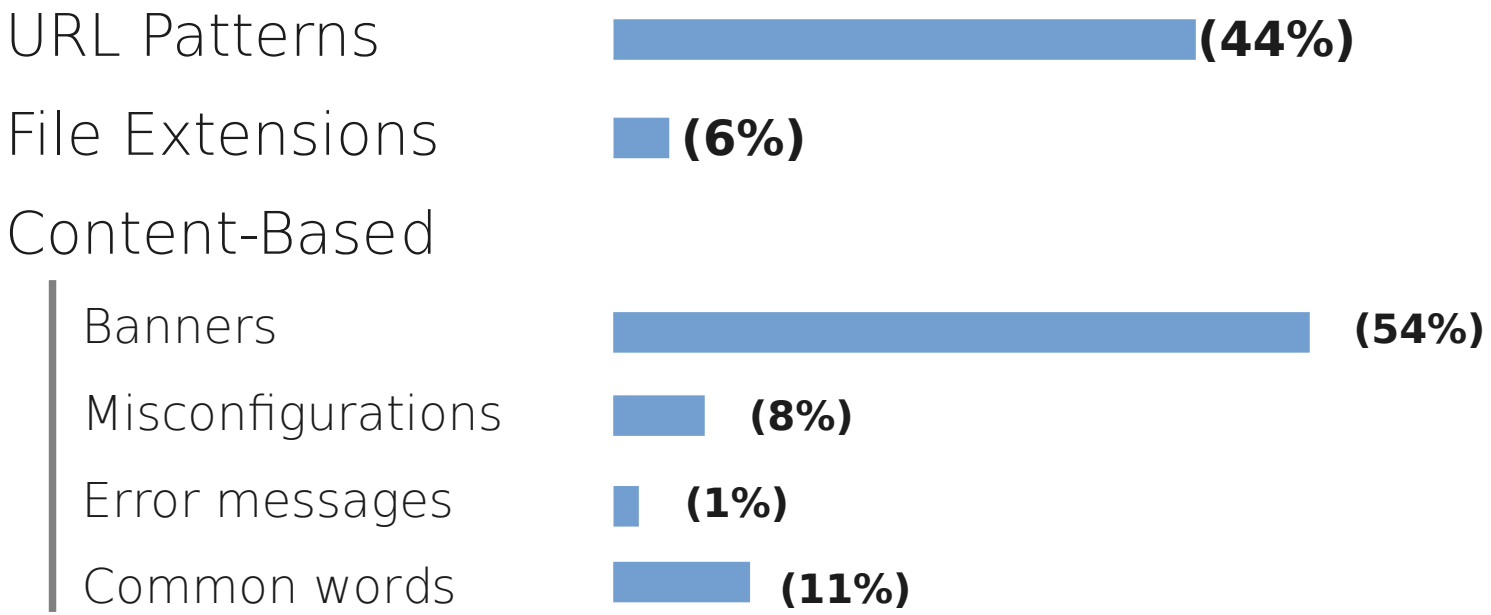
TAXONOMY

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

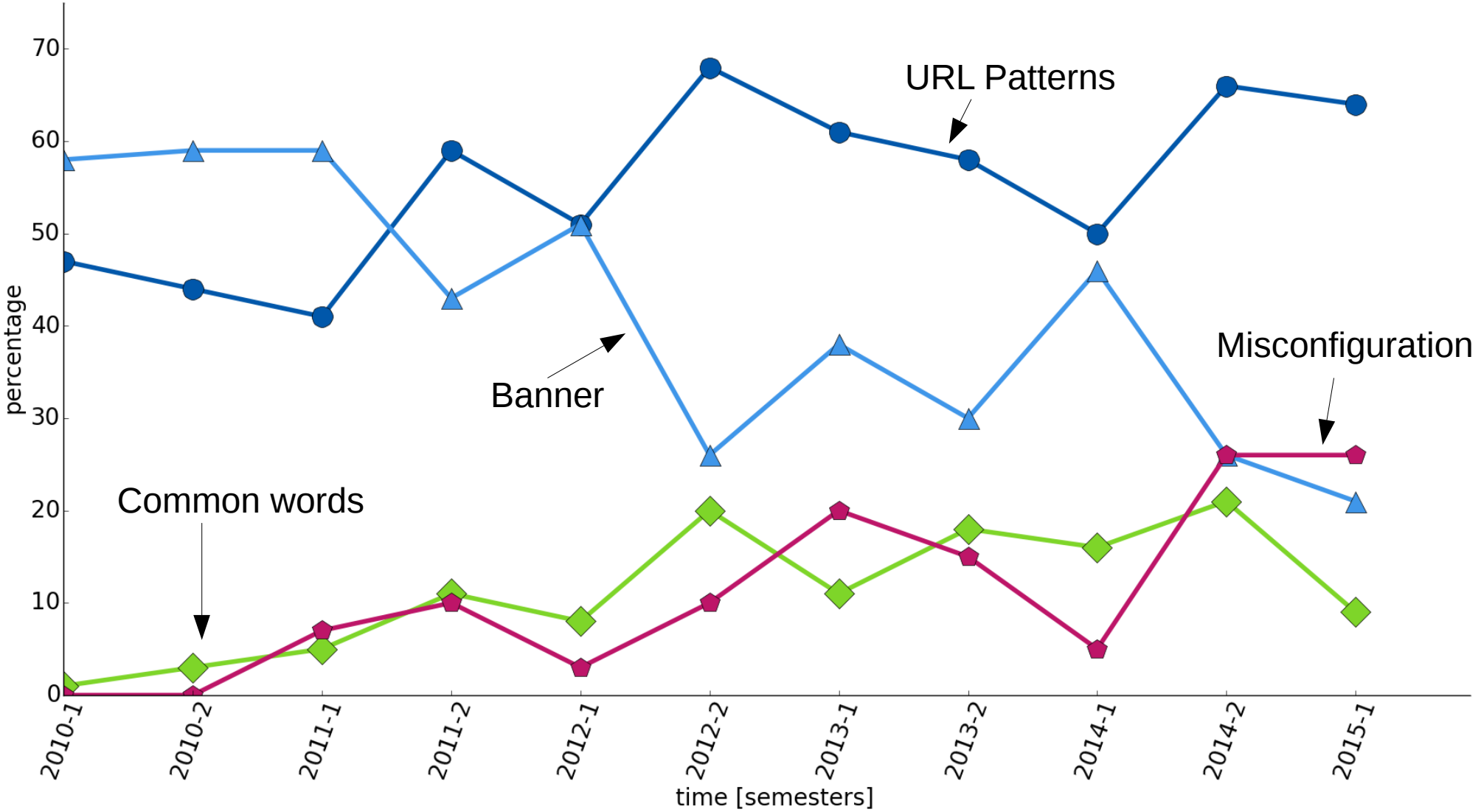


TAXONOMY

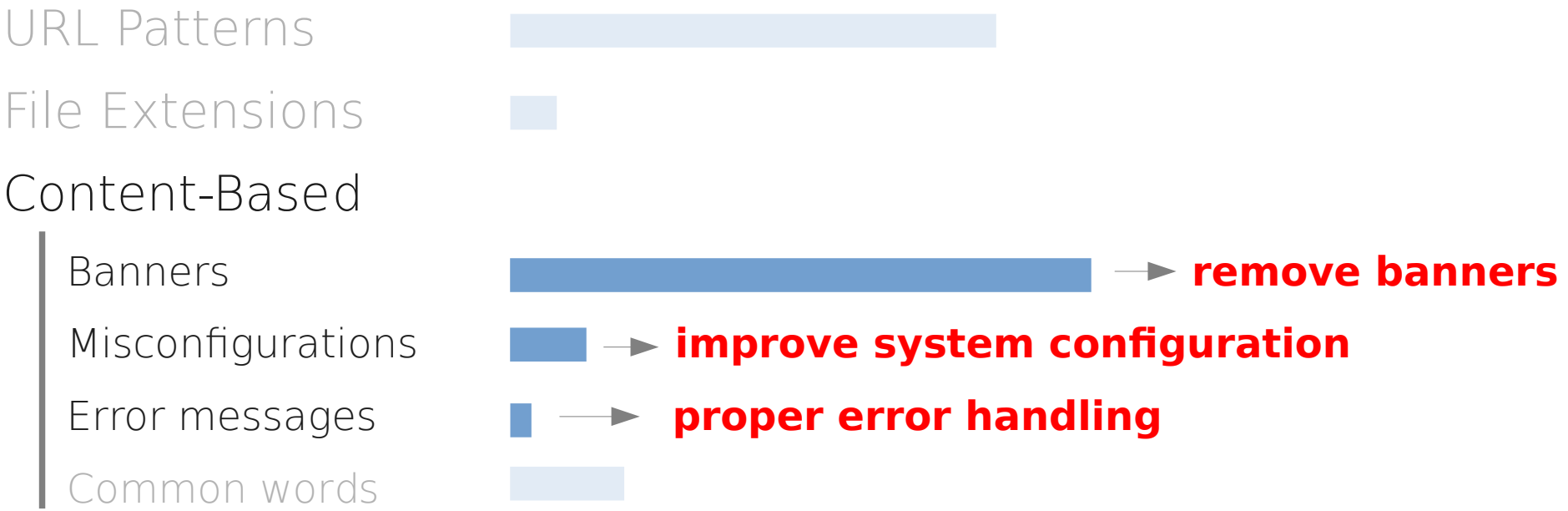
- The Exploit-DB database contains over 5143 dorks
- Automated/manual analysis



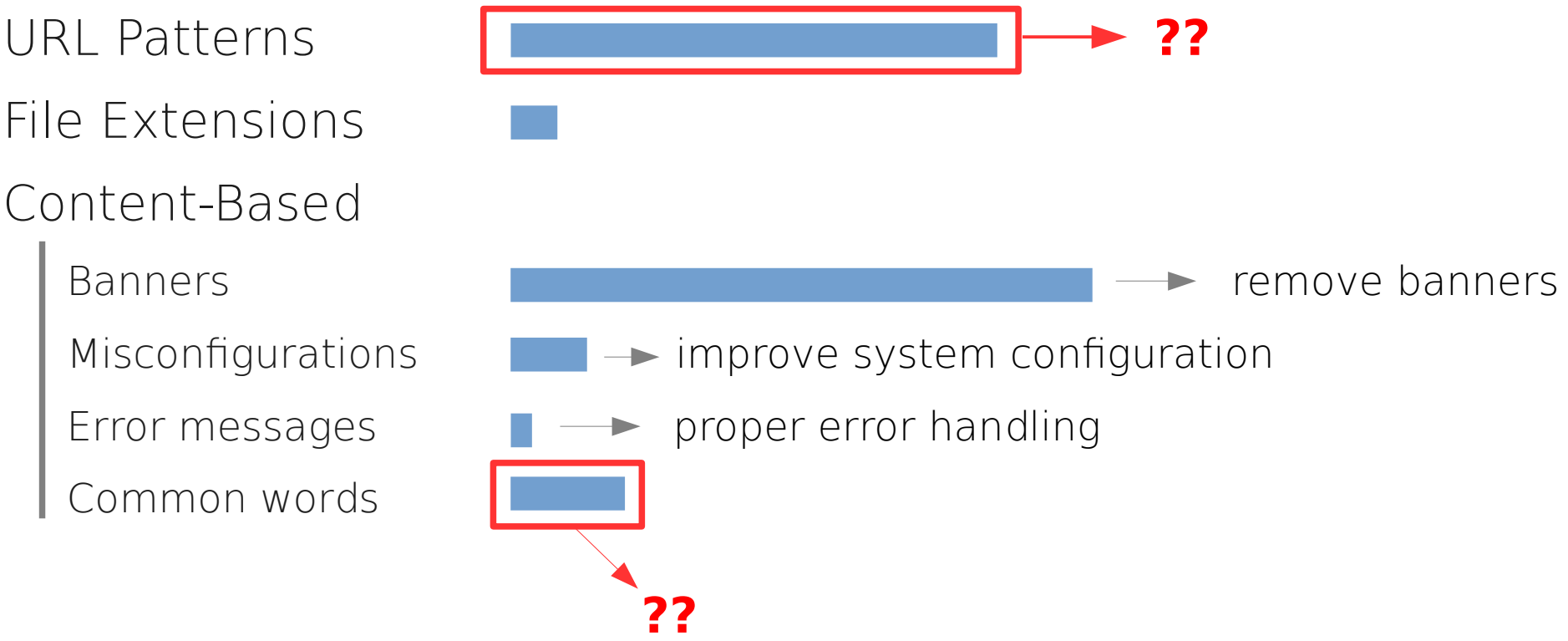
DORKS EVOLUTION BY CATEGORY



KNOWN DEFENSES



CONTRIBUTION



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

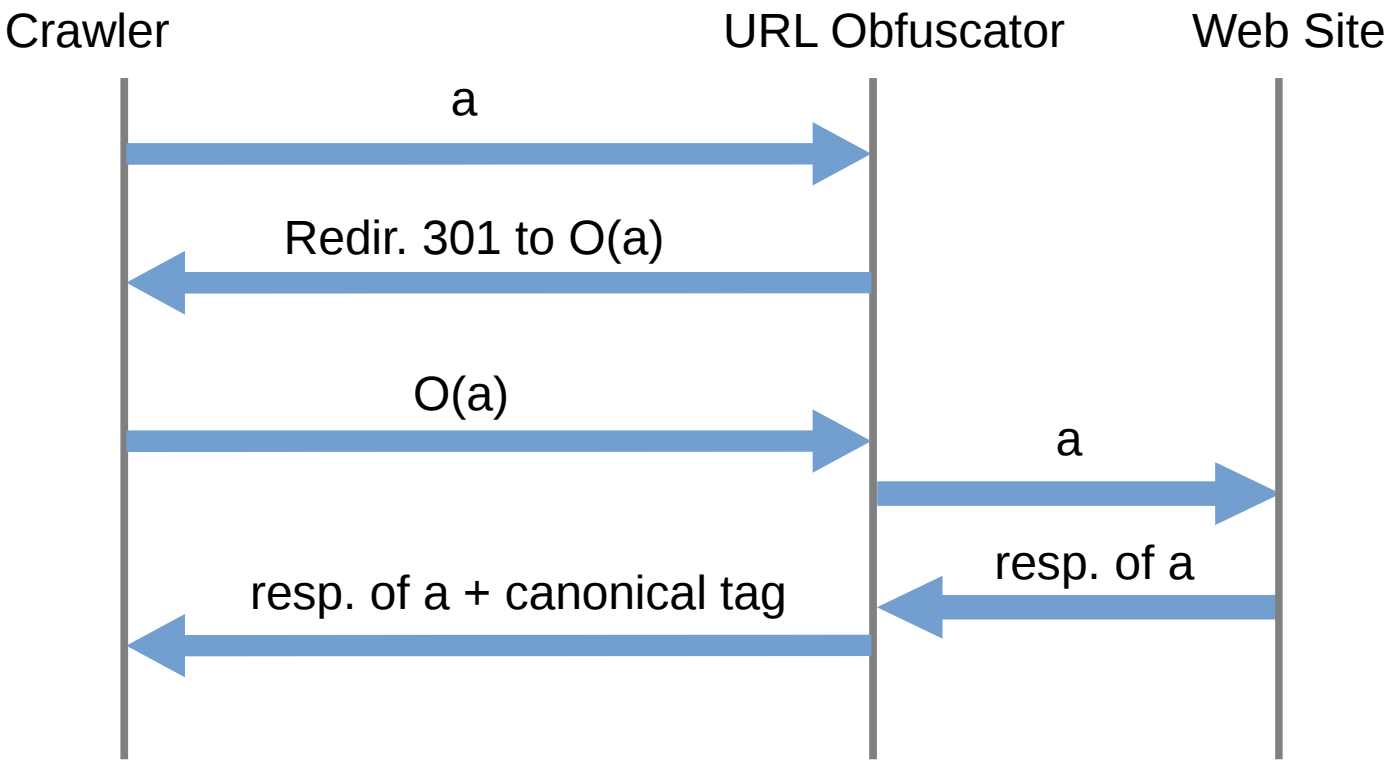
- XOR (part of) URLs with random seed kept in the server
  a = resource a
  O(a) = obfuscated resource a

- Redirect 301 to inform search engine that the page is moved

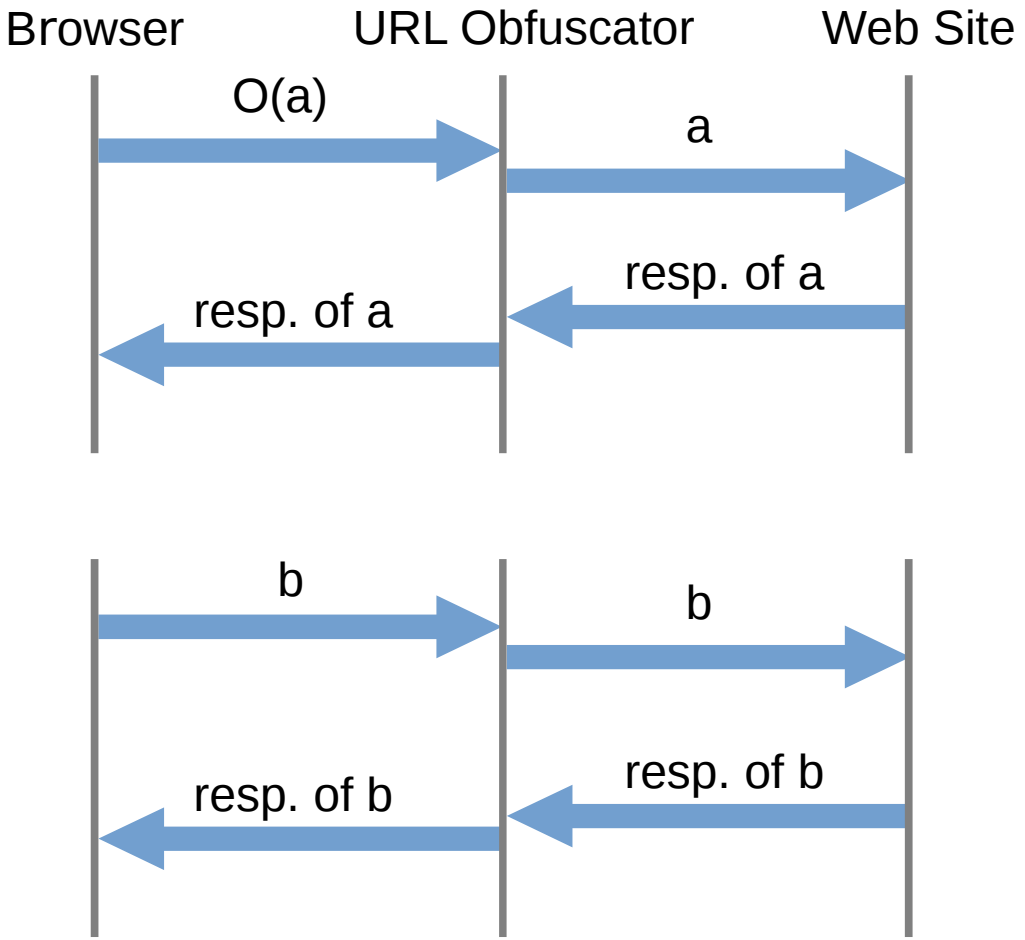
- Canonical URL Tag to delete plain URLs in the results

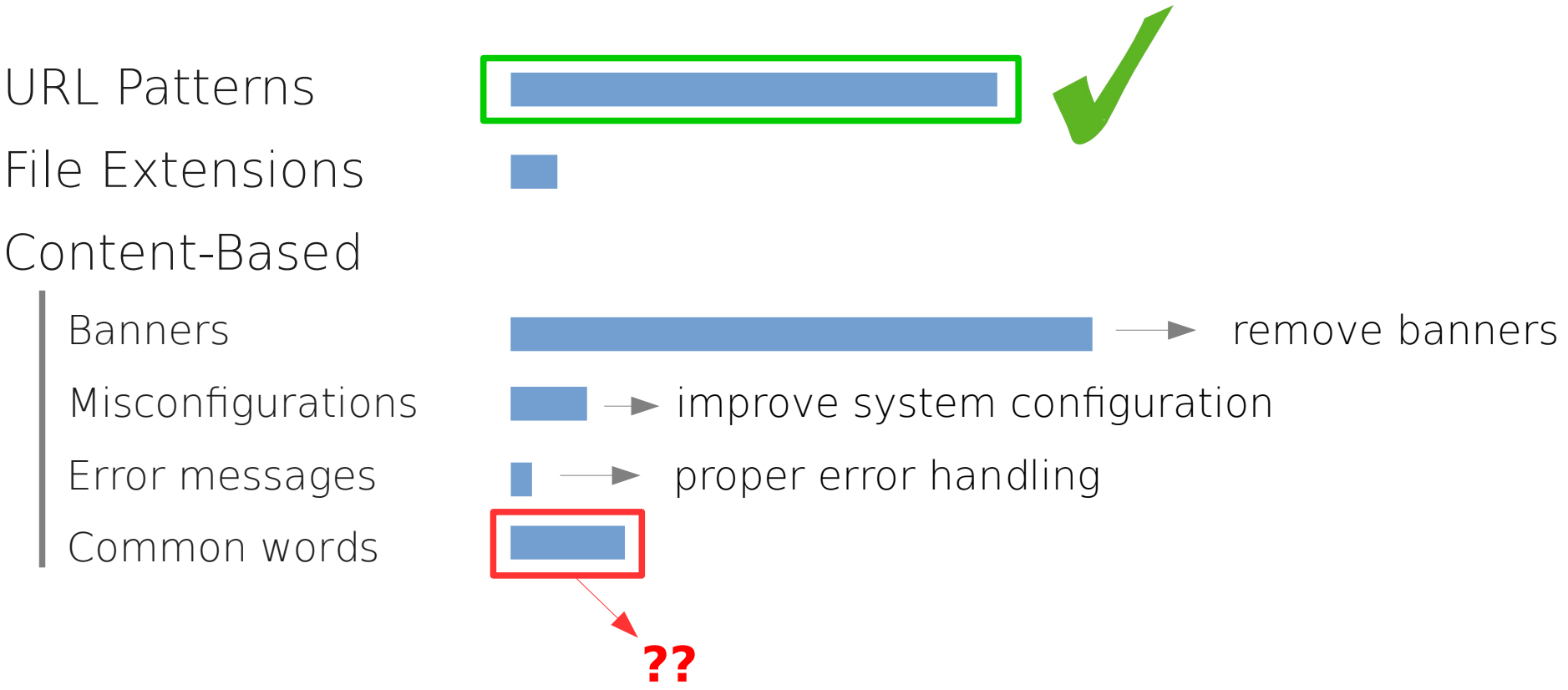
- Intercept and replace SiteMap

OBFUSCATION PROTOCOL - CRAWLERS



OBFUSCATION PROTOCOL - BROWSER

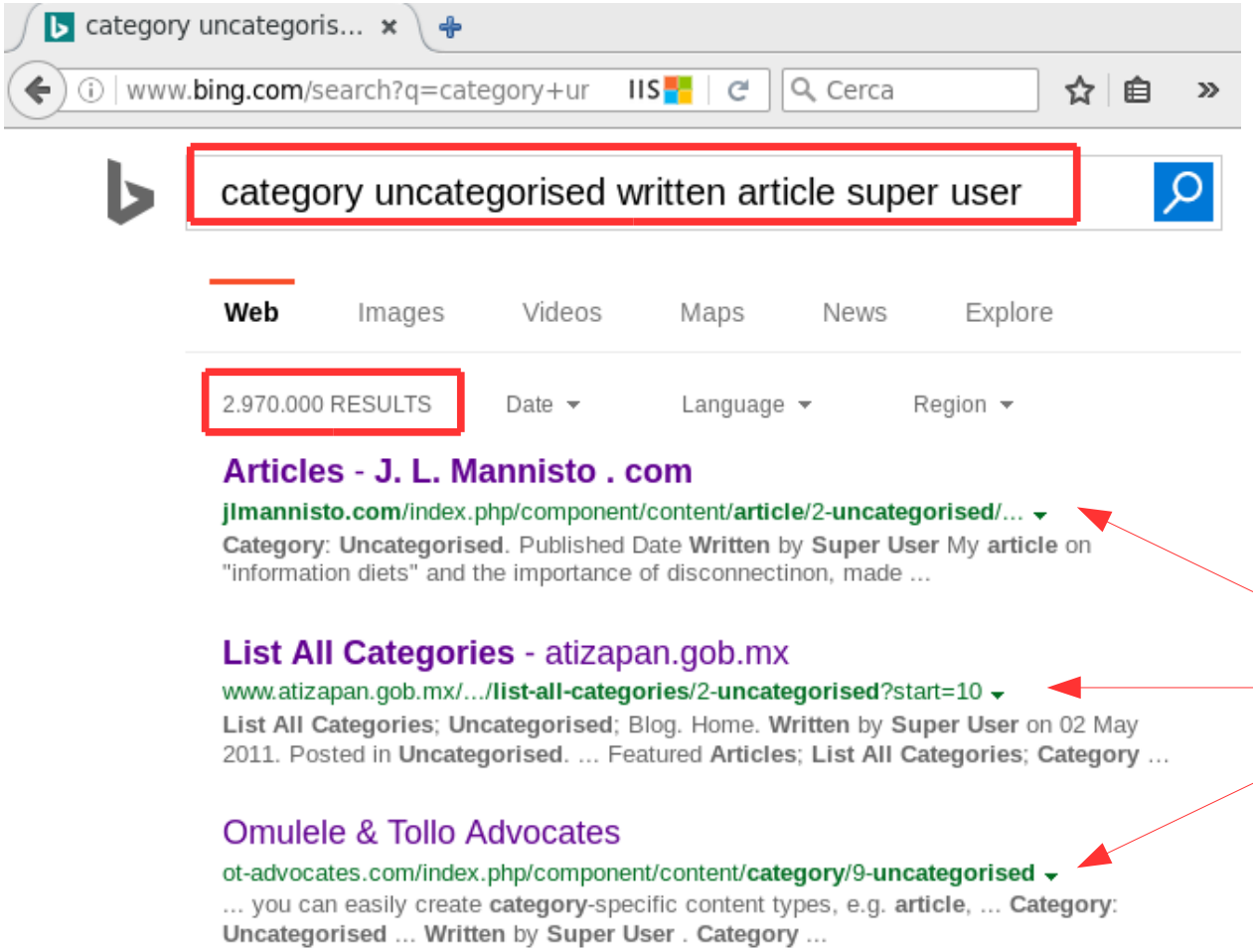



URL Patterns
File Extensions
Content-Based
Banners
remove banners
Misconfigurations
improve system configuration
Error messages
proper error handling
Common words
??

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



Joomla!

WORD-BASED DORKS: CREATION

Vanilla installation

Categories Buy Recent Register Submit Users Contact Registration List

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

“Category” + “Submit” + “....”

bing!

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

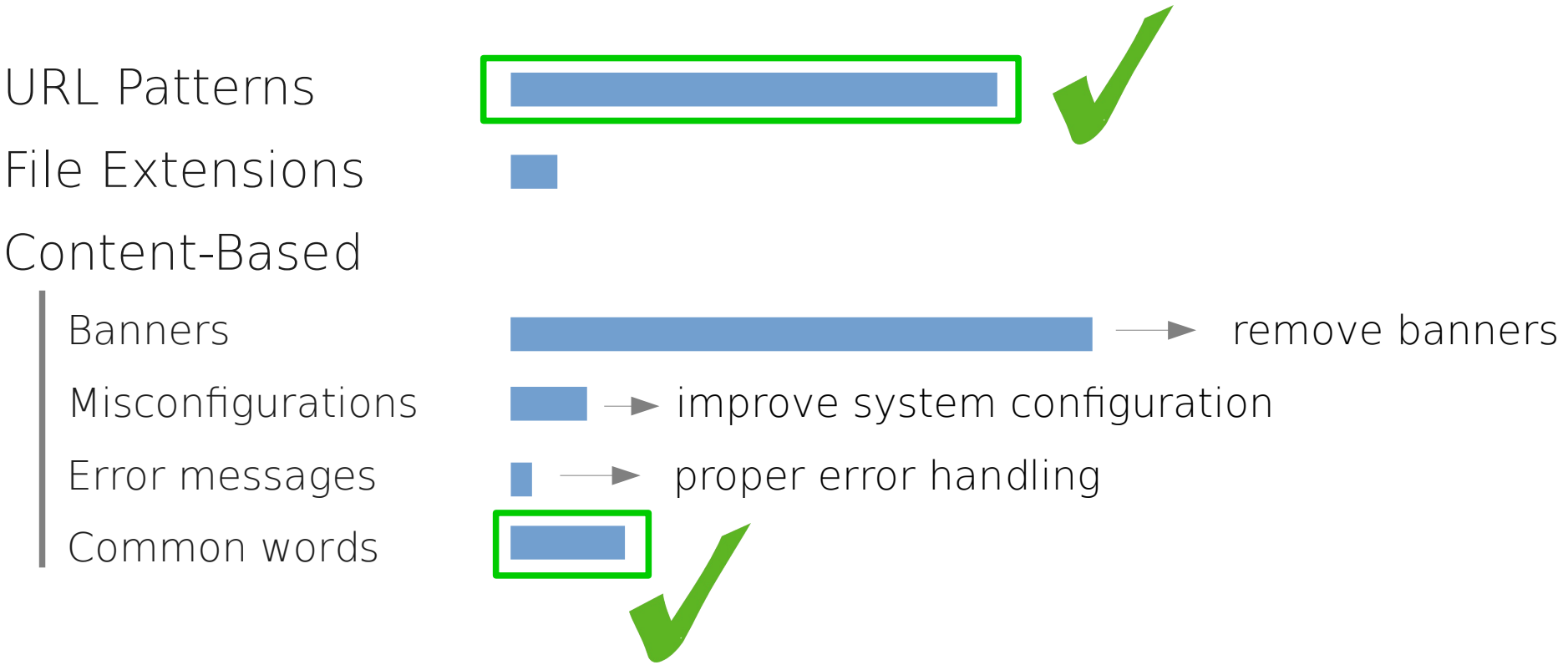
WORD-BASED DORKS: DEFENSES

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

Powered by WordPress

Power⁣ed b⁣y Wor⁣dPress

DORKS DEFENSES



CONCLUSION

1) Dork classification

2) URL Pattern Dork Defense

3) New type of Dork using common words

4) Defense against common word dorks