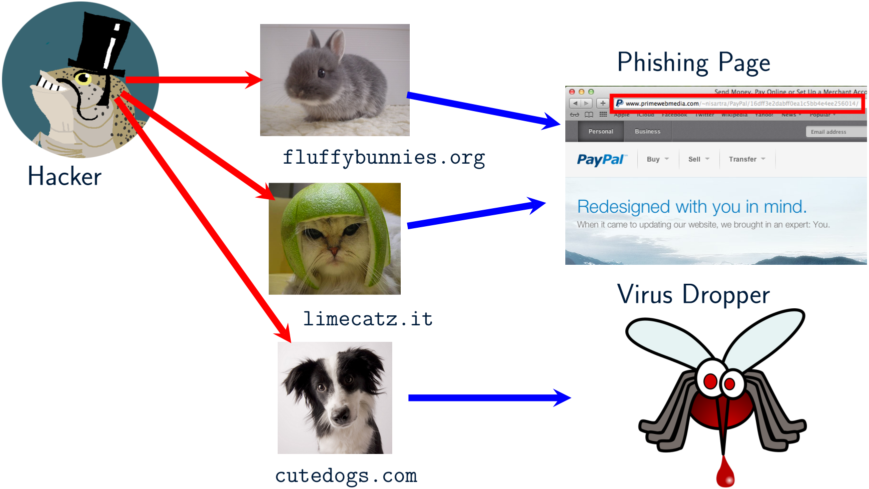
Identifying Risk Factors for Webserver Compromise

Marie Vasek & Tyler Moore

SMU

Financial Cryptography
March 5, 2014



Hacker

fluffybunnies.org

limecatz.it

cutedogs.com

Phishing Page

Send Money, Pay Online or Set Up a Merchant Acco

www.primewebmedia.com/~nisartra/PayPal/16dff3e2dabff0ea1c5bb4e4ee256014/

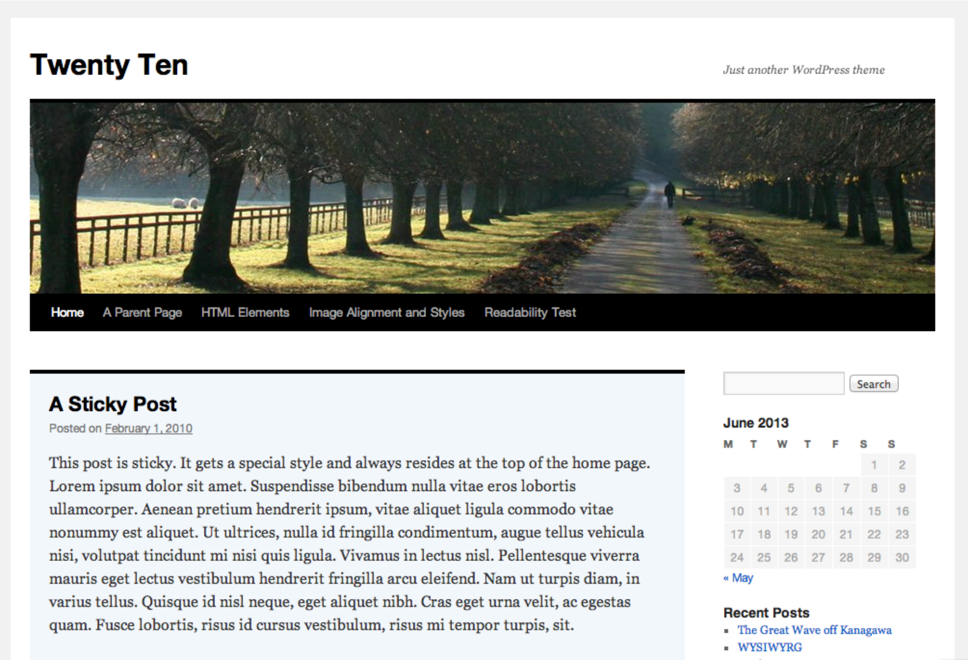
Redesigned with you in mind.

When it came to updating our website, we brought in an expert: You.

Virus Dropper

POISON HELP!
1-800-222-1222

Why?



Twenty Ten

Just another WordPress theme

A Sticky Post

Posted on February 1, 2010

This post is sticky. It gets a special style and always resides at the top of the home page. Lorem ipsum dolor sit amet. Suspendisse bibendum nulla vitae eros lobortis ullamcorper. Aenean pretium hendrerit ipsum, vitae aliquet ligula commodo vitae nonummy est aliquet. Ut ultrices, nulla id fringilla condimentum, augue tellus vehicula nisi, volutpat tincidunt mi nisi quis ligula. Vivamus in lectus nisl. Pellentesque viverra mauris eget lectus vestibulum hendrerit fringilla arcu eleifend. Nam ut turpis diam, in varius tellus. Quisque id nisl neque, eget aliquet nibh. Cras eget urna velit, ac egestas quam. Fusce lobortis, risus id cursus vestibulum, risus mi tempor turpis, sit.

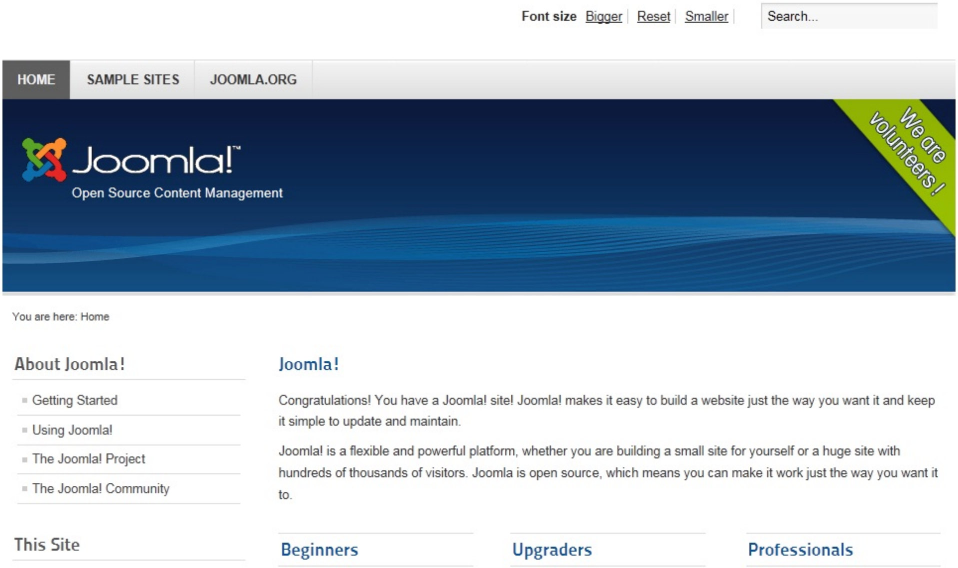
June 2013

M	T	W	T	F	S	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

« May

Recent Posts

- The Great Wave off Kanagawa
- WYSIWYRG



Joomla!

Congratulations! You have a Joomla! site! Joomla! makes it easy to build a website just the way you want it and keep it simple to update and maintain.

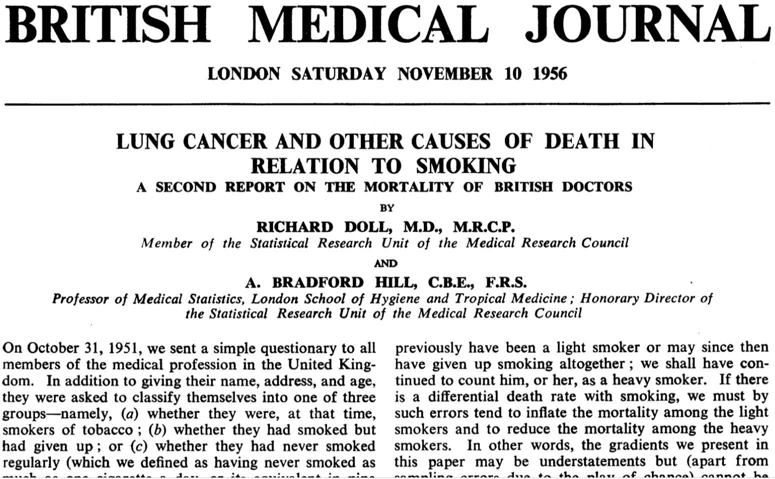
Joomla! is a flexible and powerful platform, whether you are building a small site for yourself or a huge site with hundreds of thousands of visitors. Joomla is open source, which means you can make it work just the way you want it to.

Beginners

Upgraders

Professionals

Identifying risk factors in epidemiology



BRITISH MEDICAL JOURNAL

LONDON SATURDAY NOVEMBER 10 1956

LUNG CANCER AND OTHER CAUSES OF DEATH IN RELATION TO SMOKING

A SECOND REPORT ON THE MORTALITY OF BRITISH DOCTORS

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

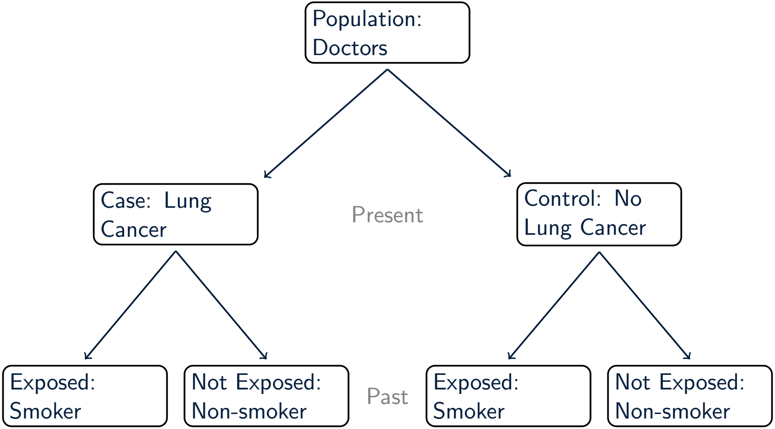
A. BRADFORD HILL, C.B.E., F.R.S.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

On October 31, 1951, we sent a simple questionary to all members of the medical profession in the United Kingdom. In addition to giving their name, address, and age, they were asked to classify themselves into one of three groups—namely, (*a*) whether they were, at that time, smokers of tobacco; (*b*) whether they had smoked but had given up; or (*c*) whether they had never smoked regularly (which we defined as having never smoked as much as one cigarette a day, or its equivalent in pipe

previously have been a light smoker or may since then have given up smoking altogether; we shall have continued to count him, or her, as a heavy smoker. If there is a differential death rate with smoking, we must by such errors tend to inflate the mortality among the light smokers and to reduce the mortality among the heavy smokers. In other words, the gradients we present in this paper may be understatements but (apart from sampling errors due to the play of chance) cannot be

Case-control study design: smoking and lung cancer



Population: Doctors

Case: Lung Cancer

Present

Control: No Lung Cancer

Exposed: Smoker

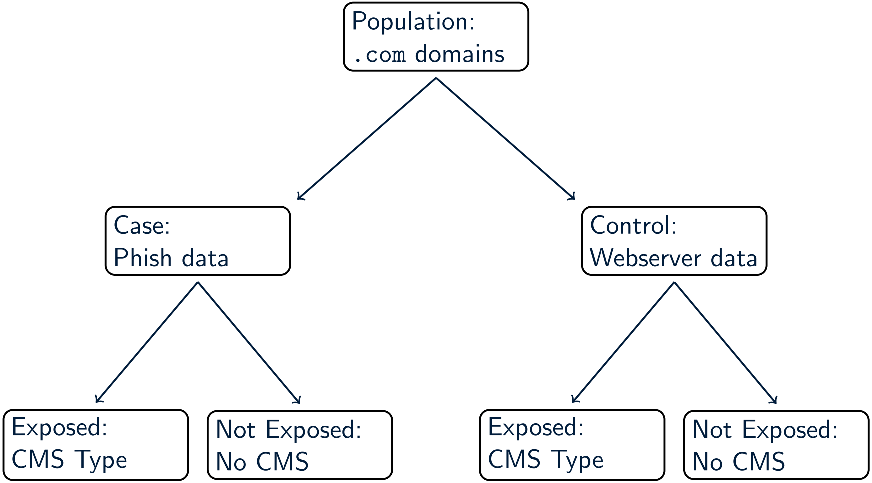
Not Exposed: Non-smoker

Past

Exposed: Smoker

Not Exposed: Non-smoker

Case-control Study Design: webservers and phishing



Population:
.com domains

Case:
Phish data

Control:
Webserver data

Exposed:
CMS Type

Not Exposed:
No CMS

Exposed:
CMS Type

Not Exposed:
No CMS

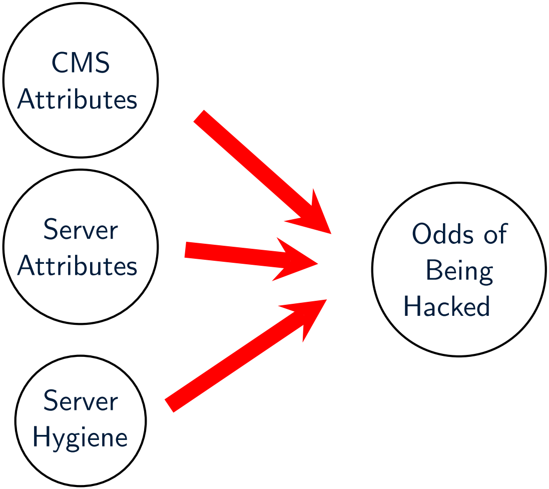
Research Hypotheses

Indicators

CMS type
# Exploits for CMS
CMS Market Share

Server Type
Country

Version Visible
Shared Hosting
HTTPONLY



CMS Attributes

Server Attributes

Server Hygiene

Odds of Being Hacked

Research Hypotheses

H0: Running a CMS is a positive risk factor for compromise.

H0b: *(corollary)* Some CMS types are risk factors for compromise.

H1: Some server types are risk factors for compromise.

H2: CMS market share is a positive risk factor for webserver compromise.

H2b: *(corollary)* Outdated software with limited market penetration is a negative risk factor for compromise.

H2c: *(corollary)* The number of exploits available for a type of software is a positive risk factor for compromise.

H3: Actively hiding detailed software version information is a negative risk factor for compromise.

H4: Running a webserver on a shared hosting platform is a positive risk factor for compromise.

Data Collection Overview

Case Datasets

Phishing Dataset: 2 months' worth of data from

- PhishTank
- APWG
- 2 takedown companies

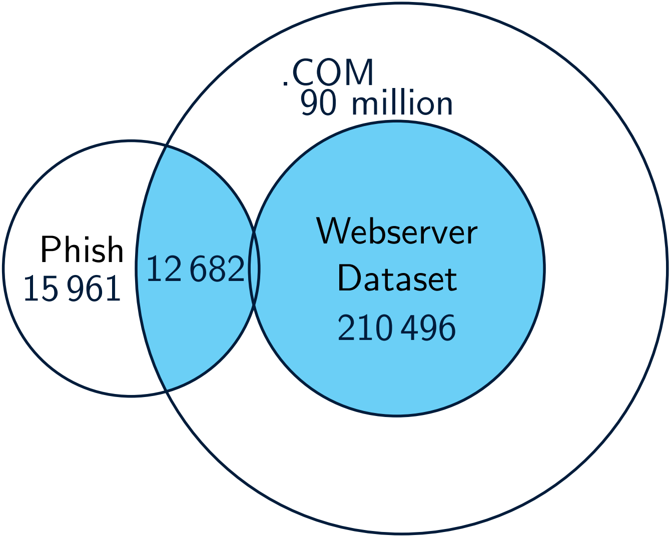
Search-Redirection Dataset: 14 months' worth of data from

- data collected by Leontiadis, Moore, and Christin

Control Dataset

Webserver Dataset: Random sample of .COM zone file

Data Collection Overview



Identifying Content Management Systems

- Attempted to identify all CMSes with at least 1% market share.
- CMSes successfully identified include:
  - WordPress
  - Joomla
  - Drupal
  - Zen Cart
  - Blogger
  - TYPO3
  - Homestead
- CMSes not successfully identified:
  - vBulletin (3.5%)
  - DataLife Engine (1.5%)
  - PHP Link Directory (1.6%)
  - Discuz! (1.3%)
  - phpBB (1.2%)
  - Bitrix (1.0%)

Identifying CMS from HTML

```
<meta name=‘‘generator’’ content=‘‘WordPress 3.0.3’’ />
```

```
<meta content=‘‘SimplePress v.4.7’’ name=‘‘generator’’/>
```

```
<link rel=‘stylesheet’ id=‘cptchStylesheet-css’
href=‘http://fluffybunnies.org/blag/wp-content/plugins
/captcha/css/style.css?ver=3.5.2’ type=‘text/css’
media=‘all’ />
```

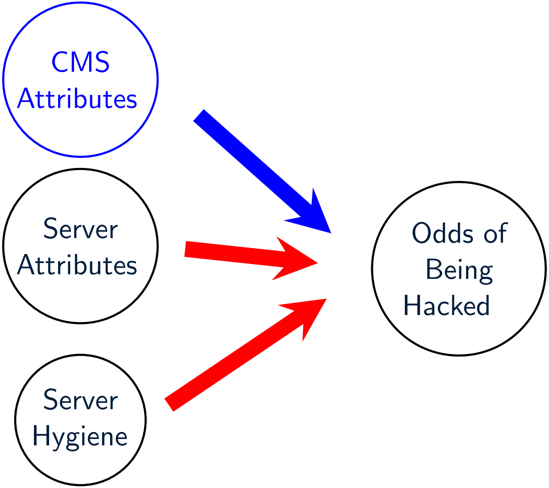
Hypotheses

Indicators

CMS type
# Exploits for CMS
CMS Market Share

Server Type
Country

Version Visible
Shared Hosting
HTTPONLY



CMS Attributes

Server Attributes

Server Hygiene

Odds of Being Hacked

Does Content Management System Matter?

Odds compared to no CMS

	Phishing	Search Redirection Attack
WordPress	4.41	17.08
Joomla	7.05	23.82
Drupal	0.78	6.56
Zen Cart	4.80	2.35
Blogger	0.28	1.08
TYPO3	0.14	4.20
Homestead	0.04	0.16

- WordPress and Joomla have higher odds of being hacked than servers running no CMS.
- Less customizable / less popular CMSes have lower odds of being hacked than servers running no CMS.

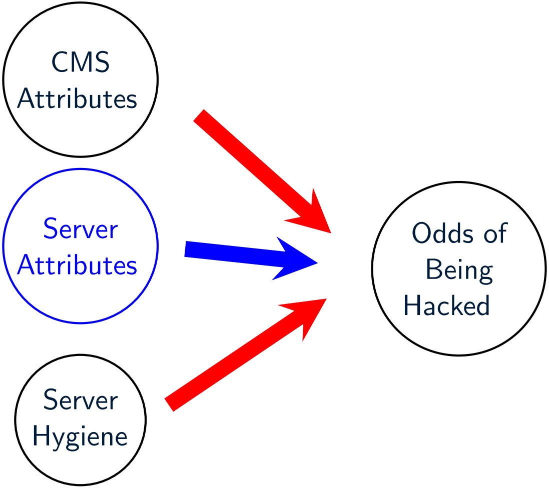
Hypotheses

Indicators

CMS type
# Exploits for CMS
CMS Market Share

Server Type
Country

Version Visible
Shared Hosting
HTTPONLY



CMS Attributes

Server Attributes

Server Hygiene

Odds of Being Hacked

Does Server Software Matter?

Odds compared to Microsoft IIS

	Phishing	Search Redirection Attack
Apache	5.44	14.13
Nginx	2.24	8.63
Yahoo	0.62	1.56
Google	0.63	1.75

- Apache and Nginx have higher odds of being hacked.
- These are also more likely to host sites running CMSes like WordPress.
  - We'll later run a regression to control for this effect.

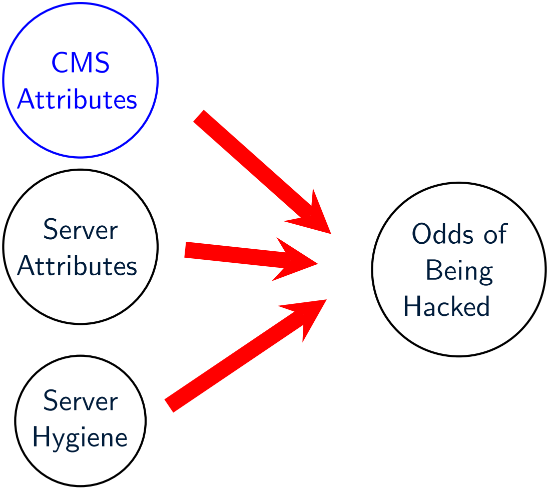
Hypotheses

Indicators

CMS type
# Exploits for CMS
CMS Market Share

Server Type
Country

Version Visible
Shared Hosting
HTTPONLY



CMS Attributes

Server Attributes

Server Hygiene

Odds of Being Hacked

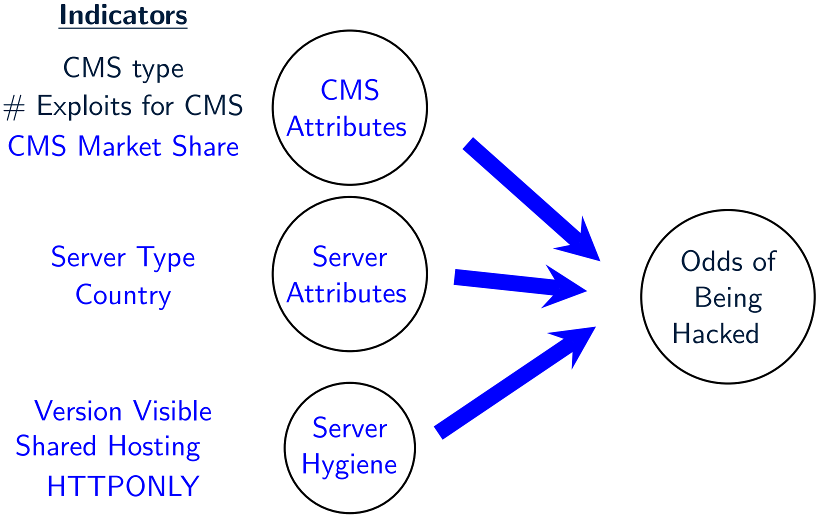
Does CMS Market Share Drive the Discovery of Exploits?

- 52 CMS platforms
- Exploits listed in ExploitDB.
- # servers = market share · # .COM domains · response rate

	coef.	95% conf. int.	Significance
Intercept	**3.05**	(2.33, 3.76)	$p < 0.00001$
lg(# Servers)	**0.68**	(0.39, 0.98)	$p = 0.00003$
Model fit: $R^2 = 0.29$			

- Since these are so correlated, we'll use one in our regression: market share.

Why Compromise Rates Vary



Indicators

CMS type
# Exploits for CMS
CMS Market Share

CMS Attributes

Server Type
Country

Server Attributes

Version Visible
Shared Hosting
HTTPONLY

Server Hygiene

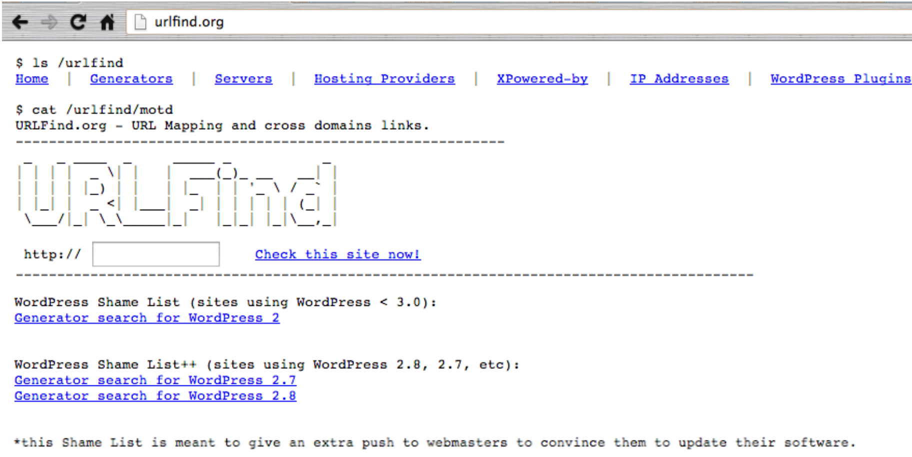
Odds of Being Hacked

Why Compromise Rates Vary

	CMS		No CMS	
	Phish	Cloak	Phish	Cloak
lg (# Servers)	1.09	1.02		
HTTPONLY	1.12	0.43	0.42	1.14
No Server Vsn	0.87	1.07	1.05	1.37
Shared Host	2.20	0.23	1.35	0.29

- Controlled for server and country in this regression.
- The higher the market share, the more likely to be hacked.
- Being on a shared host makes one more likely to be hacked to serve phishing pages, and less likely to be hacked to search redirect.
- Other hygiene variables a wash.

Conventional Wisdom: Name and Shame



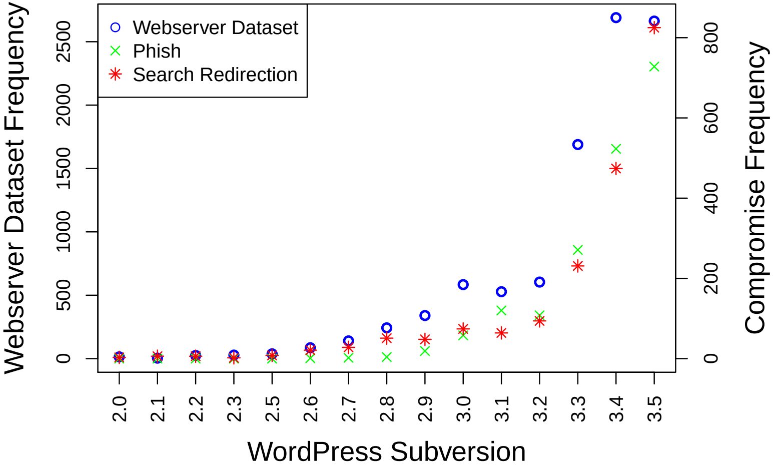
```
$ ls /urlfind
Home  |  Generators  |  Servers  |  Hosting Providers  |  XPowered-by  |  IP Addresses  |  WordPress Plugins

$ cat /urlfind/motd
URLFind.org - URL Mapping and cross domains links.
----------------------------------------------------------
 _   _ ____  _     _____ _           _ 
| | | |  _ \| |   |  ___(_)_ __   __| |
| | | | |_) | |   | |_  | | '_ \ / _` |
| |_| |  _ <| |___|  _| | | | | | (_| |
 \___/|_| \_\_____|_|   |_|_| |_|\__,_|

 http://                 Check this site now!
-----------------------------------------------------------------------------------------
```

WordPress Shame List (sites using WordPress < 3.0):
Generator search for WordPress 2

WordPress Shame List++ (sites using WordPress 2.8, 2.7, etc):
Generator search for WordPress 2.7
Generator search for WordPress 2.8

*this Shame List is meant to give an extra push to webmasters to convince them to update their software.

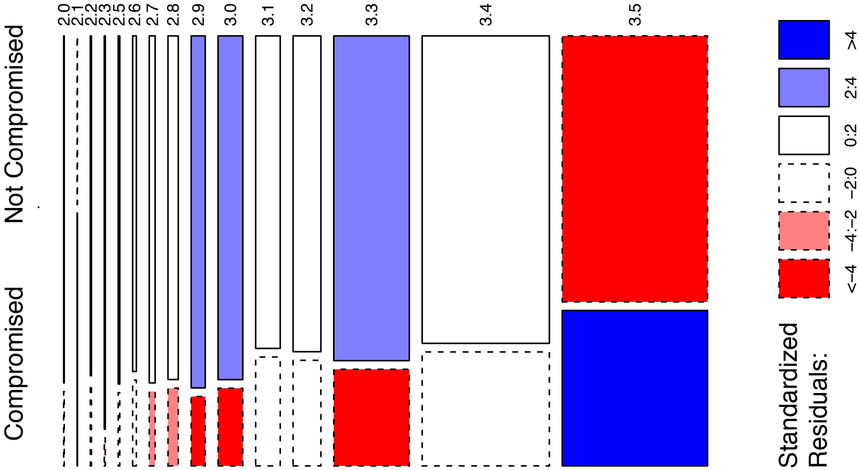
Compromise by WordPress Version



Compromise by WordPress Version

Outdated installations less at risk

Up-to-date installations more at risk



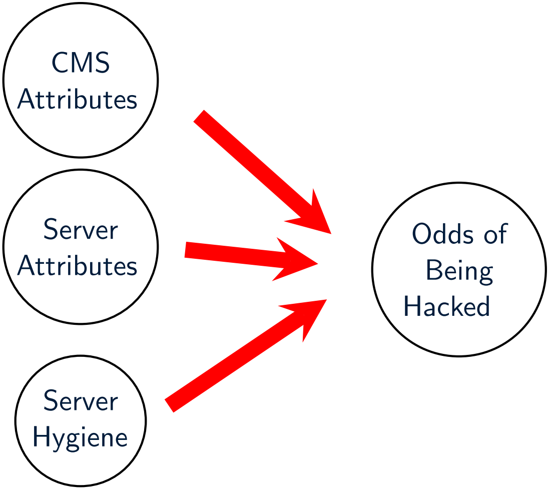
Conclusion: Revisiting the Model

Indicators

CMS type
# Exploits for CMS
CMS Market Share

Server Type
Country

Version Visible
Shared Hosting
HTTPONLY



Conclusion

- Case-control studies are useful tools for measuring cybersecurity
- Certain CMSes (notably Joomla and WordPress) more likely to be compromised.
- Woefully outdated CMSes less likely to be compromised!
- Key driving factor for CMS compromise is popularity.
- Our approach challenges traditional notions of security
  - We care about **secure outcomes** not configurations

Future Work

- Tracking indicators over time
- Additional sources of compromise data
- Expand TLD selection

Any questions?

This research is partially funded by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) Broad Agency Announcement 11.02, the Government of Australia and SPAWAR Systems Center Pacific via contract number N66001-13-C-0131