ATTACKS ON NEWLY REGISTERED WEBSITES, A COMPARISON

Marko Niinimaki[1], Veli Pajula[2], John Lawrence[3], and Kitichai Chanyalikit[4]
[1,3,4]Webster University Thailand, 1 Empire Tower, Sathorn, Bangkok 10120, Thailand
[2]University Consortium of Seinajoki, Kampusranta 9C, 60320 Seinajoki, Finland

ABSTRACT

In this paper we present a case study of hacker/intrusion activities on newly registered websites. We study how much of the incoming traffic is potentially malicious and if different web designs attract different types of malicious traffic. To implement our study, we simultaneously register and activate two websites - with similar designs but different content - and a comparison website with no content. The sites run for two months on a platform of a commercial web-hosting provider. The sites are registered under a domain of network research consortium wirlab.net. The platform utilizes a standard Linux operating system with an Apache web server with no known vulnerabilities. All network traffic to the sites is recorded using the tcpdump application. Our analysis shows that more than 90% of all traffic to the websites is potentially malicious. Moreover, most of the intrusion attempts use the ssh (secure shell) protocol instead of http. Of the two non-empty web sites, the more adult oriented one attracted more intrusion attempts. Moreover, we compare the newly registered sites with an established site and notice differences in the web traffic.

KEYWORDS: website security, website intrusion

1. Introduction

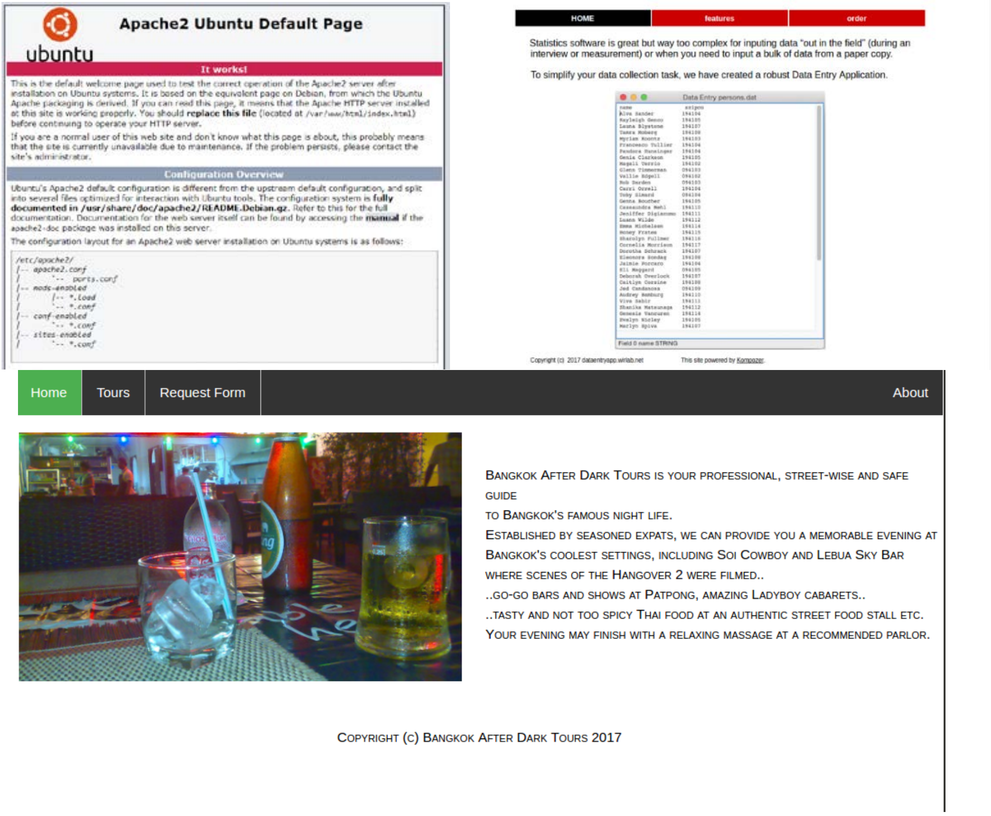
According to a recent estimate, there are one billion websites on the World Wide Web [1]. The same report states that Google deems over 50 million websites malicious: code on these sites was either trying to steal information or install malicious software. In 2012, the Sophos corporation [2] estimated that more than 30,000 websites are infected every day. High profile website defacements (replacing the original content by misleading or insulting messages) are likewise commonplace [3, 4]. Trying to get access to web servers (“website hacking”) may have been originally a hobby of activists and pranksters, but it is now seen as a more professional pursuit related to cybercrime [5, 6]. The motivation can be simply financial gain: a successful breach can turn a server computer into a crypto currency miner [7].

A lot of research concentrates on preventing and detecting web server intrusion attempts, [8-10] but, to our knowledge, there are less studies about profiles of web servers that get compromised. In this paper we present a comparison of two simple websites that are designed to appear as product/service advertisements. Both sites run in a standard Linux virtual machine leased from an inexpensive service provider. A few days before the web pages were uploaded, their domains were registered and the web server software (Apache) started. Additionally, we leased and registered a third web site that only shows the Apache default start page.

Data about network traffic to the sites was collected during a period of two months (in the future we will collect data from much period). We tracked all incoming TCP and UDP traffic using the tcpdump program. Neither of the websites was mentioned on any social media or via e-mail discussion to anyone, nor were they promoted by using search engine optimization on the pages. The site dataentryapp.wirlab.net describes a Windows/OSX application that can be used to record survey data. The site bangkoktour.wirlab.net describes a fictional tour guide company whose (only) service is to introduce tourists to Bangkok nightlife. The empty (Apache default start page) website is registered with the name pc123.wirlab.net. The front pages of each site are shown in Figure 1.

In this experiment, the research questions are (i) what is the proportion of malicious/hacking traffic in comparison with legitimate traffic to newly registered websites (ii) does the content of the website affect the size or proportion of malicious/hacking traffic and (iii) does the traffic that the sites receive differ from traffic that an older, more established website receives. The older site in question is wirlab.net's main web server, which has been active since 2001.

The rest of the paper is organized as follows. In Section 2, we describe the content and provisioning of the websites, and how data about network traffic is collected. Moreover, we study the traffic to the websites. Section 3 presents improvements to web server security based on our findings and Section 4 summarizes the results.



Apache2 Ubuntu Default Page

It works!

This is the default welcome page used to test the correct operation of the Apache2 server after installation on Ubuntu systems. It is based on the equivalent page on Debian, from which the Ubuntu Apache packaging is derived. If you can read this page, it means that the Apache HTTP server installed at this site is working properly. You should **replace this file** (located at /var/www/html/index.html) before continuing to operate your HTTP server.

If you are a normal user of this web site and don't know what this page is about, this probably means that the site is currently unavailable due to maintenance. If the problem persists, please contact the site's administrator.

Configuration Overview

Ubuntu's Apache2 default configuration is different from the upstream default configuration, and split into several files optimized for interaction with Ubuntu tools. The configuration system is **fully documented in /usr/share/doc/apache2/README.Debian.gz**. Refer to this for the full documentation. Documentation for the web server itself can be found by accessing the **manual** if the apache2-doc package was installed on this server.

The configuration layout for an Apache2 web server installation on Ubuntu systems is as follows:

```
/etc/apache2/
|-- apache2.conf
|       `--  ports.conf
|-- mods-enabled
|       |-- *.load
|       `-- *.conf
|-- conf-enabled
|       `-- *.conf
|-- sites-enabled
|       `-- *.conf
```

Statistics software is great but way too complex for inputing data "out in the field" (during an interview or measurement) or when you need to input a bulk of data from a paper copy.

To simplify your data collection task, we have created a robust Data Entry Application.

Data Entry persons.dat

```
name                 srlpos
Alva Sander          194104
Kayleigh Genco       194105
Launa Blystone       194107
Tamra Moberg         194108
Myriam Koontz        194103
Francesco Tullier    194104
Pandora Runsinger    194104
Genia Clarkson       194105
Magali Terrio        194102
Glenn Timmerman      094103
Vallie Edgell        094102
Rob Darden           094103
Carri Orzell         194104
Toby Simard          084104
Genna Boucher        194105
Cassaundra Mehl      194110
Jeniffer Digiacomo   194111
Luann Wilde          194112
Emma Michelsen       194114
Honey Frates         194115
Sharolyn Fullmer     194116
Cornelia Morrison    194117
Dorotha Schrack      194107
Eleonora Sondag      194108
Jaimie Forcaro       194104
Eli Meggard          094105
Deborah Overlock     194107
Caitlyn Corzine      194108
Jed Candanosa        094109
Audrey Hamburg       194110
Viva Sabir           194111
Shanika Matsunaga    194112
Genesis Vanzuren     194114
Evelyn Kisley        194105
Marlyn Spiva         194107
```

Field 0 name STRING

Copyright (c) 2017 dataentryapp.wirlab.net   This site powered by Kompozer.

BANGKOK AFTER DARK TOURS IS YOUR PROFESSIONAL, STREET-WISE AND SAFE GUIDE
TO BANGKOK'S FAMOUS NIGHT LIFE.
ESTABLISHED BY SEASONED EXPATS, WE CAN PROVIDE YOU A MEMORABLE EVENING AT
BANGKOK'S COOLEST SETTINGS, INCLUDING SOI COWBOY AND LEBUA SKY BAR
WHERE SCENES OF THE HANGOVER 2 WERE FILMED..
..GO-GO BARS AND SHOWS AT PATPONG, AMAZING LADYBOY CABARETS..
..TASTY AND NOT TOO SPICY THAI FOOD AT AN AUTHENTIC STREET FOOD STALL ETC.
YOUR EVENING MAY FINISH WITH A RELAXING MASSAGE AT A RECOMMENDED PARLOR.

COPYRIGHT (C) BANGKOK AFTER DARK TOURS 2017

Figure 1 Pc123.wirlab.net (top left), dataentryapp.wirlab.net (top right), and bangkoktour.wirlab.net.

2. The Websites, Data Collection and Analysis

Our data collection period was Feb 14 – April 11 2017. At all three sites we were able to see incoming traffic to a wide variety of ports, but much of the traffic was related to the virtual machine hosting environment. The web server software was activated on Feb 16th. The first http requests arrived within one hour after starting the web server software, and even before the registration of the domain names were propagated by the domain name system (DNS). However, other suspicious traffic started at both sites as soon as the virtual machines were activated. We consider traffic suspicious (potentially malicious) if it consists of requests to log in to the virtual machine (by ssh) or requests that test HTTP vulnerabilities. It is, of course, possible that this kind of traffic originates from outsourced security consultants, but we could not confirm this with the service provider.

Table 1 shows the breakdown of incoming IPv4 traffic by destination port at the sites. The amount of IPv6 traffic was insignificant. We see that for all the virtual machines, ssh traffic has a large volume. Another interesting feature was the number of incoming connections from different IP addresses.

Table 1 Incoming Traffic by Packets

	dataentryapp	bangkoktour	pc123
Incoming packets	1.8M	3.3M	3.3M
Incoming tcp packets	1.56M	3M	3.2M
ssh	1.4 M	2.9M	1.56M
http	17700	23300	1266
telnet	36000	33800	0
remote desktop	13600	2700	0
Incoming udp packets	270 000	460 000	7132
dns	256 000	447000	802
sip	6000	6000	0
ssdp	1800	2000	0
data/other	6000	2000	6330
Unique IP’s	60000	53600	21000

Table 2 shows the number of initiated ssh connections, and the number of IP addresses where the connections originated. The number of initiated connections corresponds to “key exchange init” packets in the ssh traffic. We can see that the bangkoktour site attracted twice the amount of ssh connection attempts compared with the dataentryapp site. However, connection attempts to bangkoktour originated from just 691 IP addresses, whereas attempts to dataentryapp came from 1352 addresses. 142 of these addresses attempted ssh connections to both sites. Surprisingly, these IP addresses are not listed at openbl.org, a website that collects information about intrusion attempts. Figure 2 shows the distribution of the attacks by date. The pc123 site received ssh connection attempts from a range of IP addresses, but the attempts were not as persistent as with the other sites. For all the sites, most of the IP's that tried to connect by ssh were located in China PR (IP geolocation for the IP addresses was provided by www.infobyip.com). However, for the dataentryapp site, the top originator site was in the Czech Republic.

Table 2 Incoming Ssh Connection Attempts

	dataentryapp	bangkoktour	pc123
Number of attempts	151456	358400	126787
Unique IP's	1352	691	3105

Compared to ssh connection attempts, there were few apparent intrusion attempts by http. We can quite safely assume that a human user accesses a website using a standard web browser that loads the images from the site. Since our sites had images, we should find a request of loading them in the web server's logs. However, there were none. Thus, it looks like all the accesses were via computer programs. Some of these were web indexing programs like Googlebot, but most requests were trying to find vulnerabilities in an obvious way by issuing requests like "GET /current_config/passwd HTTP/1.1". Dataentryapp and bangkoktour were accessed from 650 and 611 IP addresses, respectively. 115 of these were the same. Interestingly, only 10 of the addresses that tested web server vulnerabilities were the same as those that tried to intrude by ssh. Dataentryapp received many more connections (probable intrusion attempts) by the remote desktop protocol. Moreover, the most active web server vulnerability probes came from North America and the Netherlands. Details of http probes are shown in Table 3. This table contains, additionally, a comparison with http traffic that www.wirlab.net received during the same time.

Table 3 Features of HTTP Traffic

Dataentryapp	Num. requests	Robots	Probably malicious	Probably human
Request	2568	60	1456	0
Unique IP's	650	36	70	0
Bangkoktour				
Requests	2801	57	1897	0
Unique IP's	611	40	126	0
Pc123				
Requests	1316	53	1263	0
Unique IP's	223	40	183	0
www.wirlab.net				
Requests	26337	3458	7878	400
Unique IP's	8697	668	6704	27

Comparing the http traffic of the newly registered sites with traffic of www.wirlab.net (registered since 2001) reveals some more characteristics. Www.wirlab.net received almost 10 times the amount of http requests compared to either of the new sites. About 30% of the traffic seems malicious. However, the newly registered sites received bursts of vulnerability probes from the same IP addresses (ca 15 to 20 probes per address). Www.wirlab.net received such requests only at the rate of 1.2 probes per address.

3. Towards Improved Web Server Security

Though brief, our data collection and analysis indicate that web servers are under constant probes and intrusion attempts. However, we can see that the intrusion attempts are not very sophisticated: the intruders' favorite methods seem to be brute force or dictionary based ssh break-in attempts (see e.g. [11]) and scripts that try to exploit web server vulnerabilities that our web server does not have.

Fortunately, it is possible to protect one's web server against such intrusion attempts by well-known "host hardening" [6] methods: up-to-date security patches, strong passwords, firewalls and other intrusion detection systems; and backups.

- Security patches: most Linux/Apache based web servers run a popular Linux distribution like Ubuntu, that can be configured to apply security updates automatically. To our knowledge, the most recent critical vulnerability was heartbleed [12] that affected the ssh service. The servers used in the case study were not vulnerable to it.
- Strong passwords: in our virtual machines, the only login method was using ssh. The ssh user account had a long, randomly generated password.
- Firewalls and other intrusion detection systems: the popular software-based firewall Snort [9] can be configured to detect and stop requests like "GET /current_config/passwd HTTP/1.1" discussed in Section 3.
- Backups: if the website is compromised, the most obvious recovery method is to re-create it completely by using a new virtual machine and the most recent backup of the web pages.

4. Summary, Conclusions and Future Work

In this case study, the research questions were (i) what is the proportion of malicious/hacking traffic in comparison with legitimate traffic to newly registered web sites (ii) does the content of the website affect the size or proportion of malicious/hacking traffic and (iii) does the incoming http traffic of newly registered web sites differ from that of older sites.

We registered three web sites, pc123.wirlab.net (that shows only Apache default page), dataentryapp.wirlab.net and bangkoktour.wirlab.net using a low-cost virtual machine provider. In all the newly registered sites, potential intrusion attempts by the ssh protocol started almost as soon as the virtual machines were set up, and intrusion attempts by http occurred within 1h – 12h of turning on the web server software. Most of the incoming traffic was intrusion attempts: trying to crack ssh credentials or trying to gain access by web server-related vulnerabilities. We estimate that the proportion of malicious/hacking traffic was about 90%, but the exact number is difficult to establish since some of the traffic could be standard probes and updates by the virtual machine provider.

The dataentryapp.wirlab.net and bangkoktour.wirlab.net websites contained a form that a customer can fill in and a working e-mail address. The form and the e-mail address could have attracted interest from website scanners since implementing forms often provides opportunities to find vulnerabilities, and e-mails can be harvested and sold for marketing. However, the forms were not submitted with false data, nor did we receive any mail from the e-mail addresses.

Bangkoktour.wirlab.net received much more ssh intrusion attempts than dataentryapp and pc123, but the amount of intrusion attempts by http was fairly similar at both dataentryapp and bangkoktour. It is interesting to note that the attempts most often did not originate from the same IP addresses. Moreover, the source of http intrusion attempts was often in the US and the Netherlands, whereas the ssh attempts came most often from China.

We see interesting differences when we compare http traffic of these newly registered sites with http traffic to a more established site, www.wirlab.net. The newly registered sites received a lot of malicious requests from a small set of IP addresses (15 to 20 requests per address), whereas www.wirlab.net received malicious requests from a wider set (1.2 requests).

Since our sample of web sites was very limited, this case study only serves as a proof of concept for a much larger, and longer-term data collection[a]. The case study was limited in other aspects, too: the websites (dataentryapp and bangkoktour) had only one script, whereas hackers are probably more attracted by e-commerce and content management system (CMS) websites that rely heavily on scripts. To our knowledge, research on security of CMS systems has concentrated in finding vulnerabilities in CMS's rather than in empirical analysis [13]. Thus, in the future, our research will concentrate in (i) long term changes in traffic patterns and (ii) analysis of targeted attacks to e-commerce platforms.

This research has been partially supported by Webster University Thailand grant “Big Data Integration” 19th June 2017.

[a] Our reference site, www.wirlab.net has web server logs since 2005. Our reference CMS site (running WordPress) has been active for 4 years. Both of these are “real” web sites serving an active user community.

References

[1] Sucuri Security, Website hacked trend report Q1-2016, 2017. [Internet]. Available from https://sucuri.net/reports/Sucuri-Website-Hacked-Report-2016Q1.pdf

[2] Sophos, Security Threat Report, 2012. [Internet]. Available from https://www.sophos.com/medialibrary/pdfs/other/sophossecuritythreatreport2012.pdf

[3] Keizer, G. "Hackers" deface UN site, Computerworld, 12 Aug 2007.

[4] FBI. ISIL defacements exploiting WordPress vulnerabilities, 7 April 2015. [Internet]. Available from https://www.ic3.gov/media/2015/150407-1.aspx

[5] Hui K. Do hackers seek variety? An empirical analysis of website defacements. Singapore Management University, School of Information Systems; 2012.

[6] Panko, R. Corporate Computer Security (4th ed.), New York: Pearson, 2015.

[7] Tahir, R et. al. Mining on someone else's dime: Mitigating covert mining operations in clouds and enterprises. Proc. International Symposium on Research in Attacks, Intrusions, and Defenses; 2017 Sept 18-20; Atlanta, GA. Springer, 2018.

[8] Provos, N. A Virtual Honeypot Framework. Proc. USENIX Security Symposium; 2004 Aug 9-13; San Diego, CA. 2004.

[9] Roesh M. Snort: Lightweight intrusion detection for networks. Proc LISA'99: 13th Systems Administration Conference; 1999 Nov 7-12; Washington, USA. 1999

[10] Nauta, KR, Lieble, F. Offline network intrusion detection: Mining tcpdump data to identify suspicious activity. Proc. AFCEA Federal Database Colloquium; 1999 Sep 18-21; San Diego, CA. 1999.

[11] Valli, C, Rabadia, P, Woodward, A. A profile of prolonged, persistent SSH attack on a Kippo based honeynet; Proc. Annual Conference on Digital Forensics, Security and Law; 2015 May 19-21; Daytona Beach, FL. 2015.

[12] Durumeric, Z et al. The matter of heartbleed. Proc. 2014 Conference on Internet Measurement; 2014 Nov 5-7; Vancouver, Canada. ACM, 2014.

[13] Patel, SK, Rathod, VR., Prajapati, JB. Comparative analysis of web security in open source content management systems. Proc. International Conference on Intelligent Systems and Signal Processing (ISSP); 2013 Mar 1-2; Vallabh Vidyanagar, Anand, Gujarat, India. IEEE, 2013.

Authors' Profiles

Marko Niinimaki, Dr. earned his M.A. degree in Philosophy and Ph.D. in Computer Science. Niinimaki has worked at CERN, various universities, and private companies. He joined Webster University Thailand in 2017 and lectures about databases, computer security and data structures. Email: niinimakim@webster.ac.th

Veli Pajula is a computer security specialist and has maintained the computer infrastructure of UCS since the 1990's. Email: veli.m.pajula@uta.fi

John Lawrence earned his M.S. degree in Education. He has applied ICT, including online CMS development, to authentic and project-based learning environments in North America and Thailand, both at the high school and university levels. He joined Webster University in 2015 and lectures on communication in education. Email: lawrencej@webster.ac.th

Kitichai Chanyalikit earned his MS ITC degree in Bangkok. He joined Webster University in 2017 where his specialty is computer networks. Email: chanyalik@webster.ac.th