IJMTARC – VOLUME – VI – ISSUE – 24, OCT-DEC, 2018 ISSN: 2320-1363

RECOGNIZING MOBILE MALICIOUS WEB PAGES USING DATA MINING

B. JHANSI[1], B.S.N.V.SATYANARAYANA[2]

[1]PG Scholar, Dept of CSE, Srinivasa Institute of Engineering & Technology, Cheyyeru, Amalapuram-A.P, India

[2]Assistant Professor, Dept of CSE, Srinivasa Institute of Engineering & Technology, Amalapuram-A.P, India

Abstract- Mobile specific WebPages differ significantly from their desktop counterparts in content, layout and functionality. Accordingly, existing techniques to detect malicious websites are unlikely to work for such WebPages. In this paper, we design and implement kayo, a mechanism that distinguishes between malicious and benign mobile webpages. We makes this determination based on static features of a webpage ranging from the number of frames to the presence of known fraudulent phone numbers. First, we experimentally demonstrate the need for mobile specific techniques and then identify a range of new static features that highly correlate with mobile malicious webpages. We then apply to a dataset of over 350,000 known benign and malicious mobile WebPages and demonstrate 90% accuracy in classification. Moreover, we discover, characterize and report a number of WebPages missed by Google Safe Browsing and Virus Total. Finally, we build a browser extension using to protect users from malicious mobile websites in real-time. In doing so, we provide the first static analysis technique to detect malicious mobile web pages.

Keywords: security, safe browsing, web authentication, malicious pages.

1. INTRODUCTION

The most natural part of versatile registering innovation is the hand telephone. Around two decades back, a hand telephone was cumbersome and was utilized for voice correspondence. It was only an augmentation of the settled line communication that enabled clients to stay in contact with associates. Presently the hand telephone isn't utilized for voice correspondence, it is likewise used to send content and interactive media messages. Future cell phones won't just empower Internet get to; however will likewise bolster rapid information administrations.

Notwithstanding the hand telephone, different sorts of cell phones are currently accessible, for instance, individual computerized partners (PDAs) and pocket (PCs). Street warriors utilize cell phones to access breakthrough data from the corporate database. A cop at a wrongdoing scene may send a unique finger impression got there for coordinating with information in a focal database through a remote system, henceforth prompting quicker distinguishing proof and capture of potential suspects. The worldwide situating framework (GPS) is utilized in hunt and protects missions, for observing and safeguarding of natural life, and for vehicle burglary counteractive action. In spite of the fact that a significant number of us are unconscious of when portable registering innovation is being utilized, it has penetrated all parts of our lives.

What is portable registering? Essentially characterized, it is the utilization of a remote system foundation to give whenever, anyplace correspondences and access to data. There are numerous parts of portable figuring and, at times, extraordinary terms are utilized to allude to them. This part gives a diagram of what versatile registering brings to the table and how it enhances the nature of our lives. Later sections talk about the fundamental remote systems and advancements that make portable registering applications conceivable.

2. RELATEDWORK

Content-based and top to bottom investigation procedures to distinguish malignant sites: Dynamic methodologies utilizing virtual machines [45], [51] and nectar customer frameworks [32], [42], [46] give further deceivability into the conduct of a site page. In this manner, such frameworks have a low false positive rate and are more precise. Be that as it may, downloading and executing every site page impacts execution and thwarts adaptability of dynamic methodologies. This execution punishment can be kept away from by utilizing static methodologies.

IJMTARC – VOLUME – VI – ISSUE – 24, OCT-DEC, 2018 ISSN: 2320-1363

Static methodologies depend on the auxiliary and lexical properties of a site page and don't execute the substance of the site page. One such system of distinguishing noxious URLs is utilizing measurable strategies for URL characterization in light of a URL's lexical and host-based properties [22]. In any case, URL-based systems for the most part experience the ill effects of high false positive rates. Utilizing HTML and JavaScript highlights extricated from a page notwithstanding URL order helps address this disadvantage and gives better outcomes [20], [11], [15], [19]. Static methodologies evade execution punishment of dynamic methodologies. Moreover, utilizing quick and solid static ways to deal with identify benevolent WebPages can keep away from costly inside and out examination all things considered.

3. PROPOSED SYSTEM

3.1 METHODOLOGIES

Our goal is to plan and build up a strategy to recognize portable particular noxious website page continuously. We remove static highlights from a page and make.

3.2 Jhanu Feature Set

A site page has a few parts including HTML and JavaScript code, pictures, the URL, and the header. Versatile particular website pages likewise get to applications running on a client's gadget utilizing web APIs We separate auxiliary, lexical and quantitative properties of such parts to produce mouni's list of capabilities. We center on separating versatile pertinent highlights that take negligible extraction time. Our speculation is that such highlights are solid pointers of whether a website page has been worked for helping a client in their web perusing background or for malevolent purposes. Our list of capabilities comprises of 44 highlights, 11 of which are new and not beforehand recognized or utilized. We portray these new highlights in detail. A subset of highlights in have been utilized by different creators in static review of work area site pages in the past.5 However, take note of that these highlights in versatile pages and work area pages contrast in size and show fluctuating connection with the idea of the website page (i.e., malignant/favorable)., HTML and URL highlights. To the best of our insight, we are the first to utilize these portable particular highlights, and don't guarantee oddity on utilizing subsets of other already distinguished highlights.

3.3 Evaluation

Our dataset contained 349,137 considerate URLs and 5,231 malignant URLs. We separated our dataset into three subsets: preparing, cross-approval and test. We first arbitrarily rearranged the information and put aside 10% of the information as the test set. The staying 90% of information was utilized for preparing and 10-crease cross-approval. For every approval round we ascertained the exactness, the false positive rate and the genuine positive rate on the approval set. We additionally utilized `1-regularization to abstain from over fitting. We shifted the regularization parameter from 0 to 1,000 in the interims of 10 and picked the best parameter. We at that point plotted a ROC bend by taking the Comparison with existing static techniques:8 We have distinguished and utilized 11 new versatile applicable highlights beforehand not considered. We take note of that none of the current strategies represent versatile particular highlights considered The non-business static examination strategy nearest is Cantina [59]. It distinguishes phishing website pages progressively utilizing static highlights of jhanu's.

4. EXPERIMENTAL RESULTS

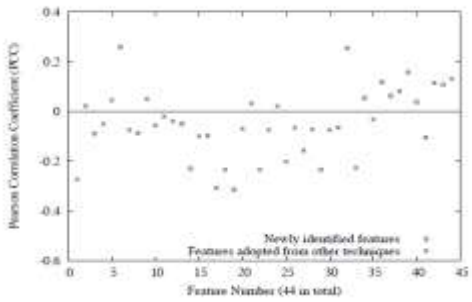
Results of a model trained on desktop webpages using desktop features studied in earlier techniques and then tested on mobile webpages. Ex2: Results of a model trained on mobile webpages by adding mobile specific features to the feature set and tested on mobile webpages. Ex1 shows that a model trained on desktop pages using features from related work performs poorly when applied to mobile webpages. However, when a model is trained with the same static features and additional mobile specific features exclusively on a mobile datatset, the results of testing on a mobile dataset improve significantly as seen in Ex2. We also compared kAYO's performance with existing static analysis tools that detect non-phishing attacks. The closest non commercial tool to kAYO based on the diversity of features and the scale of the evaluation set is Prophiler [20]. Prophiler detects drive-by-downloads on desktop webpages. We compare kAYO's performance with the performance numbers of existing static techniques described by Canali et al. [20]. Canali et al. performed an analysis of 15,000 webpages consisting of about 5,000 known webpages launching drive-bydownloads. The contenders of the comparison were then existing tools detecting malicious JavaScript [20], [17], [13], drive-by-downloads [20] and spam URLs [39]. Table 4 and Table 5 show the comparison of performance of kAYO with each of these techniques. kAYO provides the lowest false positive rate over an evaluation set twice as large as the one used by

IJMTARC – VOLUME – VI – ISSUE – 24, OCT-DEC, 2018 ISSN: 2320-1363

other techniques as shown in Table 4. Moreover, kAYO's feature extraction process is 10 times faster than the fastest existing technique [13] and classification process is 100 times faster than the fastest existing technique [19]. Finally, all the existing techniques are focused on desktop threats, whereas, kAYO focuses on mobile specific threats. Accordingly, had we been able to run these tools over our dataset, they would have performed more poorly.

Need for mobile specific techniques:

Because neither Cantina nor Prophiler were made available to us, we performed an experiment to demonstrate the need for new mobile specific models. Intuitively, due to the disparity in the same static features when measured on mobile and desktop webpages (as discussed in Section 3),



Tech-nique	Designed for		Tested on	False Pos rate	Evalua-tion set size
	Envio-nment	Threat			
[20]	Desktop	Drive by Downloads	Drive by download only	9.9	15000
[53]	Desktop	Malicious JavaScript	Drive by download only	13.7	15000
[39]	Desktop	Spam URLs	Drive by download only	14.8	15000
Union of [39], [53] [24], [37]	Desktop	Drive by, malicious JS, spam URLs	Drive by download only	17.1	15000
kAYO	Mobile	Existing mobile web threats	Existing mobile web threats	8.1	34914

5. CONCLUSION

Mobile webpages are significantly different than their desktop counterparts in content, functionality and layout. Therefore, existing techniques using static features of desktop webpages to detect malicious behavior do not work well for mobile specific pages. We designed and developed a fast and reliable static analysis technique called mouni that detects mobile malicious webpages. mouni makes these detections by measuring 44 mobile relevant features from webpages, out of which 11 are newly identified mobile specific features. mouni provides 90% accuracy in classification, and detects a number of malicious mobile webpages in the wild that are not detected by existing techniques such as Google Safe Browsing and Virus Total. Finally, we build a browser extension using mouni that provides real-time feedback to users. We conclude that mouni

IJMTARC – VOLUME – VI – ISSUE – 24, OCT-DEC, 2018 ISSN: 2320-1363

detects new mobile specific threats such as websites hosting known fraud numbers and takes the first step towards identifying new security challenges in the modern mobile web.

6. REFERENCES

[1] Gnu octave: high-level interpreted language. http://www.gnu.org/ software/octave/.
[2] hphosts, a community managed hosts file. http://hphosts.gt500.org/hosts.txt.
[3] Joewein.de LLC blacklists. http://www.joewein.net/dl/bl/dom-bl-base.txt.
[4]Lookout.https://play.google.com/store/apps/details?hl=en&id=com.lookout.
[5] Malware Domains List. http://mirror1.malwaredomains.com/files/domains.txt.
[6] Phishtank. http://www.phishtank.com/.
[7] Pindrop phone reputation service. http://pindropsecurity.com/Phone-fraud solutions/phone reputation service prs/.
[8] Scrapy — an open source web scraping framework for python. http://scrapy.org/.
[9] VirusTotal. https://www.virustotal.com/en/.
[10] Google developers: Safe Browsing API. https://developers.google.com/ Safe-browsing/, 2012.
[11] Alexa, the web information company. http://www.alexa.com/topsites,2013.
[12] dotmobi. internet made mobile. Anywhere, any device. http://dotmobi.com/, 2013.
[13] C. Amrutkar, K. Singh, A. Verma, and P. Traynor. VulnerableMe: Measuring Systemic weaknesses in mobile browser security. In Proceedings of the International Conference on Information Systems Security (ICISS), 2012.
[14] C. Amrutkar, P. Traynor, and P. C. van Oorschot. Measuring SSL indicators on mobile browsers: Extended life, or end of the road? In Proceedings of the Information Security Conference (ISC), 2012.
[15] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a dynamic reputation system for DNS. In Proceedings of the 19th USENIX
Conference on Security (SECURITY), 2010.
[16] V. A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M. T. Hunter, and P. Traynor. Pindr0p: using single-ended audio features to determine call Provenance. In Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS), 2010.
[17] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi. EXPOSURE: Finding malicious domains using passive DNS analysis. In Proceedings of the 18th Annual Network and Distributed System Security Symposium (NDSS), 2011.
[18] M. Boodaei. Mobile users three times more vulnerable to phishing attacks. http://www.trusteer.com/blog/mobile-users three times-More-vulnerable-to-phishing-attacks, 2011.
[19] M. Butkiewicz, Z. Wu, S. Li, P. Murali, V. Hristidis, H. V. Madhyastha, And V. Sekar. Enabling the transition to the mobile web with websieve. In Proceedings of the 14thWorkshop on Mobile Computing Systems and Applications.

About authors:

B. Jhansi:-B.Tech in C.S.E from affiliated to J.N.T.U. Kakinada. She is pursuing M.Tech in the stream of C.S.E in, Srinivasa Institute of Engineering& Technology an affiliated to J.N.T University Kakinada, A.P, Cheyyeru (v), Amalapuram.

Guide:

B.S.N.V.SATYANARAYANA:-Working as Assistant Professor, Srinivasa Institute of Engineering & Technology an affiliated to J.N.T University Kakinada, A.P, Cheyyeru (v), Amalapuram-53322. His Qualification is M.tech in C.S.E he has 7 years experience teaching in CSE and he has published 5+ papers on Various Streams.