ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

Detecting Malicious Web pages in Real Time

M. Anantha Raman [1], R. Anil Kumar [2], S. Gowri Shankar [3], P. Deivendran [4],

Department of Information Technology, Velammal Institute of Technology, Chennai, Tamil Nadu, India

ABSTRACT: Cloaking is a technique to show different content depending on conditions reflecting who is visiting the site. In this scenario, cloaking can be successfully used by malware writers to avoid being detected when the malware-detecting crawler visits a particular site. We have observed malware that checks if images have been successfully loaded before executing its attack. Accordingly, existing techniques to detect malicious websites are unlikely to work for such webpages. In this paper, we design and implement **malicious web page tracker technique**, a mechanism that distinguishes between malicious and benign mobile web pages. MWPT makes this determination based on static features of a webpage ranging from the number of frames to the presence of known fraudulent phone numbers. Finally, we build a browser extension using MWPT to protect users from malicious mobile websites in real-time.

KEYWORDS: MWPT Technique, authentication, access control

I. INTRODUCTION

Malicious Web pages are increasingly spread while we accessing the web. However, in spite of significant advances in processor power and bandwidth, the browsing experience on mobile devices is considerably different. These differences can largely be attributed to the dramatic reduction of screen size, which impacts the content, functionality and layout of mobile web pages. Content, functionality and layout have regularly been used to perform static analysis to determine maliciousness in the desktop space. Features such as the frequency of I frames and the number of redirections have traditionally served as strong indicators of malicious intent. Due to the significant changes made to accommodate mobile devices, such assertions may no longer be true. For example, whereas such behavior would be flagged as suspicious in the desktop setting, many popular benign mobile web pages require multiple redirections before users gain access to content.

Previous techniques also fail to consider mobile specific webpage elements such as calls to mobile APIs. For instance, links that spawn the phone's dialer (and the reputation of the number itself) can provide strong evidence of the intent of the page. New tools are therefore necessary to identify malicious pages in the mobile web. In this paper, we present MWPT , a fast and reliable static analysis technique to detect malicious mobile web-pages. MWPT uses static features of mobile web pages derived from their HTML and JavaScript content, URL and advanced mobile specific capabilities.

We first experimentally demonstrate that the distributions of identical static features when extracted from desktop and mobile web pages vary dramatically. We then collect over 350,000 mobile benign and malicious web pages over a period of three months. We then use a binomial classification technique to develop a model for MWPT to provide 90% accuracy and 89% true positive rate. MWPT's performance matches or exceeds that of existing static techniques used in the desktop space. MWPT also detects a number of malicious mobile web pages not precisely detected by existing techniques such as Virus Total and Google Safe Browsing.

Finally, we discuss the limitations of existing tools to detect mobile malicious web pages and build a browser extension based on MWPT that provides real time feedback to mobile browser users.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

II. RELATED WORK

Content-based and in-depth inspection
Techniques to Detect malicious websites:
Dynamic approaches using Virtual machines
and honey client systems provide deeper visibility into the Behavior of Web page. Therefore, such Systems have a very low false positive rate and are more accurate. However,
Downloading and executing ea h webpage
Impacts Performance and hinders Scalability of dynamic approaches. This Performance penalty can be avoided by Using static Approaches. Static approaches Rely on the structural and Lexical properties A webpage and do not execute the Content Of the webpage. One such technique of Malicious URLs is using statistical
Methods for URL Classification based on a URL's lexical and host-based Properties However, URL-based Techniques usually suffer from high false Positive rates. Using HTML and JavaScript Features extracted from a Webpage in Addition to URL classification helps address
This drawback and provides better results Static approaches Avoid performance penalty of dynamic Approaches. Additionally, using fast and Reliable static approaches to detect begin Web pages can Avoid expensive in-depth
Analysis of all webpages.

Differences between mobile and desktop

All these approaches for malicious Web page detection have focused on Websites built for desktop browsers in the Past. Mobile browsers have been shown to Differ From their desktop counterparts in Terms of security Although Differences in mobile and desktop websites Been observed before [19], it is Unclear How these differences impact security. Furthermore, the threats on mobile and Desktop websites are somewhat different Static analysis techniques using Features of desktop webpages have been
Primarily studied for drive-by-downloads
On desktop websites, whereas, the Biggest threat on the mobile web at present Is believed to be phishing Efforts in Mitigating phishing attacks on desktop Web sites include isolating browser Applications of different trust level Email filtering using content-based Features and blacklists The Best-known non-proprietary Content-based Approach to detect phishing webpages is Cantina Cantina suffers from Performance problems Due to the time lag Involved in querying the Google Search Engine. Moreover, Cantina does not workWell On webpages written in languages Other than English. Finally, existing Techniques do not account for new mobile Threats such as known fraud phone numbers That Attempt to trigger the dialer on the Phone. Consequently, whether existing Static analysis techniques to detect
Malicious desktop websites will work well
On mobile Websites is yet to be explored.

III. ARCHITECTRURE

Building a browser extension based on MWPT adds value for two reasons. First, the mobile specific design of MWPT enables detection of new threats previously unseen by existing services (e.g., pages including spam phone numbers). Second, building an extension allows immediate use of our technique. We discuss other potential avenues of adopting MWPT. We developed a browser extension using MWPT for Firefox mobile, which informs users about the maliciousness of the webpages they intend to visit. Our goal was to build an extension that runs in real-time. Therefore, instead of running the feature extraction process in a mobile browser, we outsourced the processing intensive functions to a backend server. Figure shows the architecture of the extension. User enters the URL he wants to visit in the extension toolbar. The extension then opens a socket and sends the URL and user agent information to MWPT's backend server over HTTPS. The server crawls the mobile URL and extracts static features from the webpage. This feature set is input to MWPT's trained model, which classifies the webpage as malicious or benign. The output is then sent back to the user's browser in real-time. If the URL is benign according to MWPT, the extension renders the intended webpage in the browser automatically. Otherwise, a warning message is shown to the user recommending

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

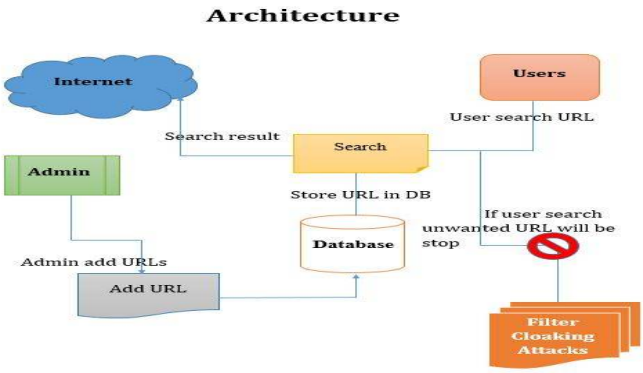
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

them not to visit the URL. Users of the extension will browse both mobile specific and desktop webpages since not all websites offer a mobile specific version. Recall that being a mobile specific technique, MWPT does not perform well on desktop webpages. Consequently, processing all pages of interest through MWPT might output incorrect results for



Desktop web pages. To address this problem, the backend server first detects whether the intended webpage is mobile specific using the same method explained in Section 4.2. The webpage is processed by MWPT only if it is mobile. The desktop web pages are analyzed using Google Safe Browsing. Note that any other existing technique for detecting desktop malicious webpages can be used instead of Google Safe Browsing. We performed manual analysis of 100 randomly selected URLs (90 benign and 10 malicious) from our test dataset and measured the performance of MWPT in real time. On an average, an output was rendered in 829 ms on average from the time the user entered a URL in MWPT's toolbar. We argue that the good performance is due to careful selection of quickly extractable features and lower complexity of mobile webpages as compared to desktop webpages. The maximum delay in result generation was seen in scraping the input webpage from its respective server. Caching already scraped webpages can reduce this delay, as we demonstrated experimentally, by an average of 85%. Figure 7 shows a screen shot of our browser extension at work. We plan to make the extension available publicly post publication

IV. ALGORITHMS USED

We describe the machine learning techniques we Considered to tackle the problem of classifying Mobile specific Web pages as malicious or We then discuss the strengths and Weaknesses of each classification technique, And the process for selecting the best model For MWPT. We build and evaluate our chosen Model for accuracy, false positive rate and true Positive rate. Finally, we compare MWPT to Existing techniques and empirically demonstrate The significance of MWPT features. We note That where automated analysis is possible,

V. CONCLUSION

In this way, we study the framework for detecting malicious webpages in real time. Therefore, existing techniques using static features of desktop webpages to detect malicious behavior for mobile specific pages. We designed and developed a fast and reliable static analysis technique that detects mobile malicious webpages and also detect phishing sites. Our application provides greater accuracy in classification, and detects a number of malicious webpages in the wild that are not detected by existing techniques such as Cantina. Finally, we build a browser extension that provides real-time feedback to users. We proposed an application for mobile platforms. We identified the weaknesses of the heuristics-based anti-phishing schemes that highly rely on the HTML source code of web pages. We conclude that our application detects new mobile specific threats such as websites hosting and takes the first step towards identifying new security challenges in the modern web.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

REFERENCES

[1] Chaitrali Amrutkar, Young Seuk Kim and Patrick Traynor, "Detecting Mobile Malicious Webpages in Real Time," IEEE Trans. Services Computing, IEEE, 2017 .
[2] Shuang Liang, Yong Ma and Yong Ma, "The Scheme of Detecting Encoded Malicious WebPages Based on Information Entropy", IEEE, 2016.
[3] Xi Xiao Ruibo Yan, and H. Yan Runguo Ye, "Detection and Prevention of Code Injection Attacks on HTML5-based Apps," IEEE,2016.
[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] Phish tank. http://www.phishtank.com/.
[7] Pin drop 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] Virus Total. 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] Dot mobi. Internet made mobile. Anywhere, any device. http://dotmobi.com/, 2013.