Poster : SHRT – New method of URL shortening including relative word of target URL

Soojin Yoon Jeongeun Park Changkuk Choi Seungjoo Kim†

CIST(Center for Information Security Technologies), Korea University
145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea
{idtptkd, planet_in, necojin, skim71}@korea.ac.kr

1. INTRODUCTION

URL shortening services provide short URLs instead of long target URLs. These services redirect a shortly shortened URL to a target URL. As the usage of shortened URLs is convenient, shortened URLs have become common. However, most of shortened URLs are not relative to their target URLs, so shortened URLs are abused for phishing attack. Phishing is a social engineering technique to steal users' privacy data or financial account credentials.

In this paper we propose SHRT, a new method of URL shortening. SHRT inserts a relative word of a target URL into its shortened URL. SHRT's shortened URLs have relativeness of the target URL, so SHRT's shortened URLs are less likely to be used for phishing.

2. RELATED WORK

2.1 Shortened URL

A shortened URL is a short URL which is redirected to a target URL. Users click a shortened URL, then its service site redirects to its target URL[1][2]. A shortened URL is composed of a service site and a unique number. The unique number is base 64, and its characters are 0~9, a~z, and A~Z. As example, in a shortened URL 'shrt.name/PeOH0f', 'shrt.name' is a service site and 'PeOH0f' is a unique number.

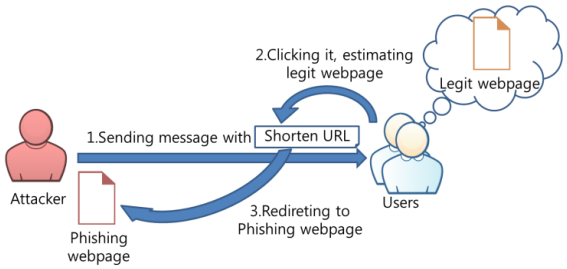

Figure 1. The phishing attack using a shortened URL

2.2 Pishing attack using a shortened URL

Users cannot guess target URLs, because a shortened URL is not relative to its target URL. As Figure 1, an attacker abuses Shortened URLs' weakness for phishing[3][4]. The attacker sends a message with a shortened URL. In most cases, the message says that its shortened URL is for a legit webpage. Users who get the message click it, estimating the legit webpage. But it is redirected to a phishing webpage and users' data can be stolen.

Avoiding abusing of shortened URLs for phishing, bit.ly provides several site domains of shortened URLs like on.fb.me, 1.usa.gov and matches target URLs to the proper site domain. In addition, E.Vishria et al proposed a patent of creating shortened URLs using contents of its webpage[5], and S.L.Hancock's patent uses shortcodes instead of geographic identifiers, URL addresses and etc. in the embedded system[6].

3. SHRT

SHRT is a new method of creating shortened URLs. We got an idea from a writing system of Arabic. Though the writing system of Arabic has no vowel characters, Arabs understand their writing without any problems. We expect that people can guess a site name when even it has no vowels.

3.1 Creating method

First, extract a site name of a target URL. The site name is the lowest of domains of URL, so it excludes protocol, host, SLD, TLD, path and etc.. Then, make a relative word of a target URL. You can make the relative word from the site name of target URL by dropping vowels, numbers and hyphen. For instance, the target URL is 'http://www.korea.ac.kr/do/Index.do', then its site name is 'korea'. Its relative word is 'kr'. Vowels of 'korea' are omitted. When the site name has not any consonant, then we can apply additional rules to creating the relative word.

And check the registration of the relative word. If it is registered on a shortened URL service, then check the registration of the target URL. If the target URL is registered, print a SHRT URL registered. If the relative word is not registered or the target URL is not registered, then count the number of URLs including relative word. These URLs are registered on the shortened URL service. After counting, make a unique number including the relative word and the counting number. For example, a relative word 'kr' is registered 4 times, then next unique number consists of 'kr' and '4'. In this paper, the index of counting starts from 0. If we set that creating unique number rule is 'the relative word' + '_' + 'the count', then the unique number is 'kr_4' for example.

SHRT inserts a relative word into a SHRT's shortened URL. Then, the target URL 'http://www.korea.ac.kr/do/Index.do' has a SHRT's shortened URL 'shrt.name/kr_4'.

We give another example of shortening. After 'http://www.korea.ac.kr/do/Index.do' was registered, if a new URL 'http://cist.korea.ac.kr/about/abouts1_01.asp' is requested to be shortened, then the new URL has same relative word 'kr' of previous target URL, but its count is 5. So, the new target URL

Symposium on Usable Privacy and Security (SOUPS) 2013, July 24–26, 2013, Newcastle, UK.

† Corresponding author, skim71@korea.ac.kr

'http://www.korea.ac.kr/do/Index.do' has a SHRT's shortened URL 'shrt.name/kr_5'. Though they have same relative word, they have distinct shortened URLs.

3.2 Safety from phishing

SHRT is the method for anti-phishing. For anti-phishing, the possibility of having same relative word must be low. We consider two aspects. One is the possibility of duplication of a relative word theoretically. The other is the case that both of phishing sites and target sites have same relative word.

The possibility of duplication of a relative word is the case of the relative word from all possible site names. All possible site names are about 6.4×10^{98}. The case of the certain relative word depends on the length of the relative word. The number of possible site names of the certain relative word is $\sum_{i=0}^{63-n} 16^i \times_{i+n} C_n$. n is the length of the relative word. The highest possibility of collision of the relative word is 9.18×10^{-23}, when the length of the relative word is 2 characters. And the weighted average possibility of collision is 2.40×10^{-38}.

To show the other case that both of phishing sites and target sites have same relative word, we utilized 100 phishing sites from phishtank(phishtank.com). Phishtank is a free community site where anyone can submit, verify, track and share phishing data. 100 phishing sites were reported at May 10th, we dropped phishing sites with an IP address instead of a domain. We compared the relative word of target sites and it of phishing sites. The result shows that there is no phishing site which has the relative word of target site. Even some phishing sites confuse users with domains like 'http://www.facebook.com.facebook-com-profile-0001564835456.tk'. SHRT extracts the site name 'facebook-com-profile-0001564835456' and converts it to a relative word 'fcbkcmprfl' which is distinct from a relative word 'fcbk' of 'facebook.com', so SHRT prevents tricks of domains.

As the probability of collision of the relative word is low in both real and theoretical viewpoint, SHRT is safe from phishing.

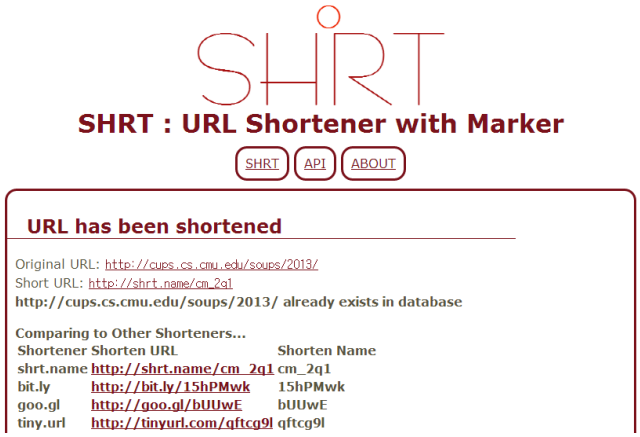

Figure 2. The snapshot of shrt.name

3.3 Efficiency

To compare efficiency of SHRT with other URL shortening services, we made a test site 'SHRT'(http://shrt.name). The test site 'SHRT' has an implemented module of creating method of SHRT and a hash index module. The hash index module inserts a hash for searching DB into a shortened URL's unique number.

We gathered 100 sites from 'randomwebsite.com' for measuring a length of 'SHRT''s URLs. The result shows that an average length of 'SHRT''s unique numbers is 10.0 characters. We tested 14 URL shortening services, and found an average length of unique numbers of other shortened URLs is 6.2 characters.

Though the average length of 'SHRT''s unique numbers is longer than it of others, users just memorize 3 characters for recalling a SHRT's shortened URL. If users already have known a target site's domain, then they can guess a relative word. So, for memorizing the SHRT's shortened URL, it costs only 3 characters.

4. CONCLUSION

In this paper, we proposed a new method of creating shortened URLs, SHRT. It gives users a hint to guess a target URL. We are sure it is safe from phishing attack, because the shortened URL of a phishing site looks different from the shortened URL of a target site. Even though SHRT's URLs is longer than other shortened URLs, memorizing SHRT's URLs is easier.

In the future, we will try to prove usability of SHRT. We plan to design an experiment in Online Social Network. If users of SHRT perceive that SHRT's URLs can help users easy to guess target URLs, SHRT will be proved to be useful.

5. ACKNOWLEDGMENT

This work was supported by the IT R&D program (10043959, Development of EAL 4 level military fusion security solution for protecting against unauthorized accesses and ensuring a trusted execution environment in mobile devices) of KEIT/MOTIE, Korea.

6. REFERENCES

[1] N Mdgiddo, P Alto, KS McCurley. 2005. Efficient Retrieval of Uniform Resource Locators. Patent No. US 6,957,224 B1

[2] D Antoniades, I Polakis, G Kontaxis, E Athanasopoulos, S Ioannidis, E P. Markatos, T Karagiannis. 2011. We.b: the Web of Short Urls. In Proceedings of the 20th international conference on World wide web (The Hyderabad, The India, March 28 – April 01, 2011). WWW '11. ACM Press, New York, NY, 715-724. DOI= http://dl.acm.org/citation.cfm?id=1963505

[3] DK McGrath, M Gupta. 2008. Behind Phishing: an examination of phisher modi operandi. In Proceeding of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats (The San Francisco, The California, April 15, 2008). LEET '08.

[4] C Grier, K Thomas, V Paxson, M Zhang. @spam: the Underground on 140 Characters or Less. In Proceedings of the 17th ACM conference on Computer and communications security (The Chicago, The USA, October 04 – 08, 2010). CCS '10. ACM Press, New York, NY, 27-37. DOI=http://dl.acm.org/citation.cfm?id=1866311

[5] E. Vishria, S. Carlos, T. Howes, L. Altos, and R. Churehill. 2011. Integrated Adaptive URL-Shortening Functionality. Patent No. US 2011/0264992 A1

[6] S. L. Hancock. 2011. Systems and methods for creating and using imbedded shortcodes and shortened physical and internet addresses. U.S. Patent 2011/0244882 A1