bit.ly/can-do-better

Neha Gupta, Anupama Aggarwal, Ponnurangam Kumaraguru
Indraprastha Institute of Information Technology, Delhi (IIIT-D)
{neha1209, anupamaa, pk}@iiitd.ac.in

ABSTRACT

Bitly, launched in year 2008 is one of the most popular URL shortening and curating services on the web. Due to content length restriction on various social media platforms, many users with malicious intent shorten and embed URLs to misguide the audience. In this work, we explore 763,160 malicious bitly links and their associated attributes to underline weak security system and policies used by bitly.

1. INTRODUCTION

Usage of URL shortening services nowadays have become a trend in Online Social Media (OSM). Such services does not only reduce the content length but also help obfuscate the actual URL behind a shortened link. Spammers take advantage of this obfuscation to make quick money by posting malicious links on OSM. An article reveals that Facebook spammers makes close to 200 million dollar just by posting shortened links to lure users [1]. Security researchers from Symantec also found that spammers abused URL shortening services to spread work from home scam [2].

Bitly, launched in year 2008 is one of the most popular URL shortening and curating services on the web. It gained major traction when Twitter started to use it as a default URL shortening service in year 2009 before the launch of its own service, *t.co* in year 2011. [1] Bitly allows its users to create an account and shorten the links. Each link shortened by a user has a unique *global hash* (an aggregated identifier corresponding to a link). Such shortened links, known as *bitmarks* can then be saved, tracked, and shared. Users are also allowed to connect any number of Facebook / Twitter accounts with their bitly accounts. With 80 million new links shortened on bitly each day and 8 billion clicks each month, spammers have also been exploiting the service to a great extent. [2] A news article reports the spread of phishing attacks on Twitter using malicious bitly links [3]. For protection against spam, bitly claims to use real-time spam detection services like Google safe-browsing and SURBL, and flags 2-3 millions of the shortened links as spam each week. Bitly neither deletes a flagged suspicious link nor suspends the associated user, but displays a warning page whenever such a link is clicked. [3]

Looking at these measures adopted by bitly but continued existence of spam, some questions arises: (i) What domains is such spam coming from? (ii) Is bitly using the claimed spam detection services effectively? (iii) Does a warning page alone help curtail the overall problem of spam? (iv) How quick is bitly in identifying suspicious accounts?

In this work, we performed a detailed analysis on a dataset of 763,160 links directly obtained from bitly to address the above questions. To the best of our knowledge, this is the first attempt to inspect only data labeled as spam by bitly to identify security loopholes contributing to such a data.

2. METHODOLOGY

We requested and received a dataset of 763,160 bitly links with warning pages in October 2013 from bitly. This dataset comprised of the global hash, associated long URL, and number of warning pages displayed for the global hash. Using this dataset as our seed input and bitly API, we collected link analytics including (1) *number of clicks*, (2) *encoder* details (bitly users whose shortened links correspond to the global hash), (3) *link history* (past 100 or complete link history, whichever is less) of encoders, (4) *network history* (associated Twitter / Facebook handles) of encoders, (5) *referrer* details (details of all platforms where the link was posted), etc. Figure 1 presents the complete dataset considered for this study. Out of 763,160 links provided, we could collect all this information for 144,851 (18.98%) links till December (our data collection is still on). We call this our *link-dataset*.

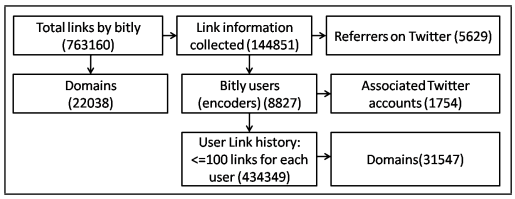


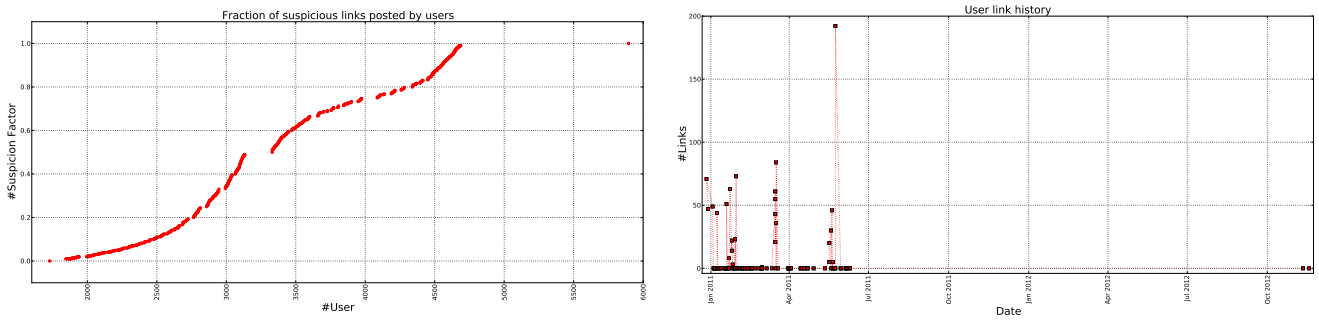
Figure 1: Dataset

3. RESULTS

We started our analysis with the identification of top domains for which multiple suspicious URLs were shortened using bitly. For this, we extracted domain information from all long URLs in the link-dataset. From 763,160 links, we obtained 22,038 unique domains out of which 665 domains had at least 100 links giving warning pages, and 146 domains had a frequency of at least 1000. Table 1(a) lists the top 10 domains and their frequencies in our dataset.

To address our second question, whether bitly uses the claimed detection services effectively, we extracted all the encoders corresponding to our link-dataset and their link history. Using this, we obtained a total of 434,349 links from the link history of 8,827 encoders. We then extracted the domains for all links in the link history and checked each domain against the blacklist used by SURBL. For this, we used the surblclient package implemented in python [4]. Simultaneously, we also made get requests to check how many links

[1] https://support.twitter.com/articles/109623-about-twitter-s-link-service-http-t-co
[2] http://www.enterprise.bitly.com/about-us/
[3] http://blog.bitly.com/post/138381844/spam-and-malware-protection
[4] https://pypi.python.org/pypi/surblclient/



(a) (b)

Figure 2: (a) Cumulative distribution on number of bitly users posting suspicious links. (b) Timeline for bitly user *bamsesang*.

in the link history were redirected to a bitly warning page. To our surprise, we found 136 unique domains blacklisted by SURBL but still undetected by bitly. These 136 domains contributed to 781 links in the link history. Top 10 of these domains based on their occurrence is shown in Table 1(b). This clearly highlights that even though bitly claims to use SURBL as one of the spam detection techniques, but all blacklisted domains by SURBL are not caught.

On the analysis of all links in encoder's link history, we obtained 89,043 links redirecting to a warning page by bitly, giving us more suspicious URLs. With this knowledge, we looked at the fraction of suspicious links shortened by different encoders. For this, we computed and assigned a *Suspicion Factor* for each encoder as :-

$$Suspicion\ Factor = \frac{\#Links\ redirecting\ to\ bitly\ warning\ page}{\#total\ links\ collected}$$

Figure 2(a) shows the cumulative distribution on number of bitly users based on their Suspicion Factor. The graph shows that 5,895 users had a Suspicion Factor less than or equal to 1, and 4,688 users had a Suspicion Factor less than or equal to 0.99. This means that 1,207 (5,895 - 4,688) out of 8,827 encoders (13.67%) had a Suspicion Factor of 1, indicating that they shortened only suspicious links. Also 18.27% encoders had 80% of their shortened URLs as malicious (Suspicion Factor = 0.8). This clearly highlights the malicious intent of these encoders on creating their bitly accounts.

Considering the existence of large number of encoders with such motives, a simple warning page does not defeat their purpose completely. We made a blog entry on our initial data analysis, in response to which we were informed that bitly does not suspend user accounts but forbids suspicious users from creating any new links [4]. To investigate, we looked at the timestamp of bitly link creation over entire link history for all encoders with a non-zero Suspicion Factor. To our surprise, we identified 265 bitly accounts with a lag between posting first and last suspicious link of at least 5 months. For encoders with Suspicion Factor as 1, maximum lag was found to be 23 months for user *bamsesang*, followed by 21 and 17 months for user *tmadalia* and *iplayonlinegames* respectively. Figure 2(b) presents the link history timeline for encoder *bamsesang* versus number of clicks on each link. The user kept shortening suspicious links for close to 2 years and whether he is detected / undetected in still unknown. This evidently underlines the security loopholes in the system used by bitly for detecting spam / malware. Account suspension can thus be a better approach to enhance security and trustworthiness of the service.

Domain	Frequency
dlinkddns.com	83019
rsscb.com	41251
cbtrends.com	40076
cbfeed.com	35069
systranet.com	12616
rambler.ru	10264
internostrum.com	10242
pp.ua	10230
weightpage222.com	9493
sina.com.cn	8697

(a)

Domain	Frequency
timesfancy.in	236
2010-film.ru	56
consultdisplayextract.info	50
linkz.it	46
1uralsk.ru	44
4roof.ru	43
ipod4u.ru	22
violinsite.ru	17
nobrain.dk	13
throwpillowbidcom.com	12

(b)

Table 1: (a) Top 10 domains in link-dataset. (b) Top 10 domains blacklisted by SURBL but undetected by bitly.

4. IMPLICATIONS

With this study, we aim to highlight the lack of effective implementation and security concerns on bitly. We present a small but in depth analysis on encoder's link history to bring to notice the existence of users who take advantage of bitly's no suspension policy. Extreme delay in suspicious user detection has also been identified. In our future work, we would like to analyze the associated Twitter and Facebook accounts of the suspicious encoders. We also plan to examine the referrers of our link-dataset (e.g. public tweets and Facebook posts containing these bitly links) to determine non-connected suspicious Twitter / Facebook profiles disseminating malicious content.

5. ACKNOWLEDGEMENTS

We would like to express our sincere thanks to bitly and particularly Brian David Eoff (senior data scientist) and Mark Josephson (CEO) for sharing the data with us.

6. REFERENCES

[1] *http://www.theguardian.com/technology/2013/aug/28/-facebook-spam-202-million-italian-research*

[2] *http://www.pcworld.com/article/2012800/spammers-abuse-gov-url-shortener-service-in-workathome-scams.html*

[3] *http://news.softpedia.com/news/Twitter-Phishing-Scam-This-Profile-Is-Spreading-Nasty-Blogs-Around-About-You-318618.shtml*

[4] *http://precog.iiitd.edu.in/blog/2013/12/bitly-could-do-better/*