2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Full Cycle Analysis of a Large-scale Botnet Attack on Twitter

Christoph Besel, Juan Echeverria*, Shi Zhou
Department of Computer Science
University College London (UCL)
London, United Kingdom
(*) j.guzman.11@ucl.ac.uk

***Abstract*—This work presents an in-depth forensic analysis of a large-scale spam attack launched by one of the largest Twitter botnets reported in academic literature. The Bursty botnet contains over 500,000; many of which have not been suspended. The bots have generated over 2.8 million spam tweets, with 2.2 million mentions directly targeting over 1.3 million distinct Twitter users. We reveal that the botnet used a network of URL shortening services and redirections to obfuscate the real landing pages. We show that users clicked on these URLs shortly after they were published and in large numbers. We even discovered the botmaster who was behind the whole operation, including creation of the Bursty botnet and registration of the several landing pages, which happen to be phishing websites. Furthermore, we found that this botmaster is still active selling Twitter bot related services. Our work reconstructs the complete course of the spam attacks, from planning to execution. This work provides in depth analysis and insight into the operation of cybercriminals on Twitter, and the cyberspace infrastructure and black-markets that they rely on. Finally, we address how the state-of-the-art bot classifiers are unable differentiate the Bursty bots from normal users, highlighting the need and importance of individual botnet analysis.**

I. Introduction

Online Social Networks (OSNs) have become an integral part of life across the world. They've not only changed the way we communicate with our friends and family, but also how we inform ourselves and how we consume news or entertaining content.

Social bots (short for 'software robots') are accounts on OSNs that produce content automatically and are operated by computer programs. A social botnet can be defined [1] as a group of bots under the control of a single botmaster.

Some social bots might be benign or helpful, but many are designed to inflict harm. They spread malicious content (spam, scams, malware) , manipulate digital influence (e.g. fake followers) and the social media discourse by faking trending topics, launching orchestrated misinformation campaigns (e.g. astroturfing attacks), and polluting the Twitter streaming API [2]. It has been reported, that social bots have influenced election campaigns, manipulated public opinion [1] and spread 'fake news'[2]. A recent review article [3] concludes: 'today's social bots are sophisticated [...] their presence can endanger online ecosystems as well as our society.' Therefore, research into the purpose and inner workings of social botnets, beyond their detection, is essential.

Here we present an in-depth analysis of the spam attack launched by one of the largest Twitter botnets, with over 500,000 accounts, reported in academic literature to date. We analyze its objective (phishing), the different campaigns it created, the performance of its strategy, and the botmaster behind it.

Finally, we show that this botnet, even with its rudimentary implementation and methods, is both successful in attracting user clicks and in evading bot classifiers that are publicly available.

[1]https://www.newscientist.com/article/2094629

[2]https://www.technologyreview.com/s/608561/

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

II. Related Work

A. Twitter Bots

1) Bot Detection: To organise related work, [3] proposed a taxonomy dividing the different approaches in the literature into three classes: Social network (graph) approaches, including a series of bot detection methods and their evaluations [4], [5]. Crowdsourcing approaches that rely on human intelligence [6]–[8]. Machine learning methods, that are based on assumed features of bot accounts [3], [9]–[11]. In addition, there is a class of 'hybrid approaches' including [12]–[14].

When real botnet datasets are found, retrieved and analysed they often conflict with previous assumptions about bot/Sybil accounts [13], [15]–[17]. This strongly suggests that botnet analysis is needed in addition to general bot detection. Most bot datasets analysed are either comparatively small or contain bots from a mix of several botnets, with a few exceptions [15], [17].

2) Analyses of botnets: Given the rich literature on detection systems for Sybil or bot accounts, there is a surprisingly small number of studies on the actual analysis of botnets. Among the few in-depth analyses of botnets [13], [16], [18], some were of anecdotal nature, others based on very small datasets. Most authors agree on the importance of studying botnets [16] in order to develop effective detection measures, ground truth remains difficult to obtain. It is evident that most bots will be part of a botnet, so studying general bot classifiers will certainly miss on key insights that are only attainable from individual botnet analysis.

B. Twitter spam

Historically, miscreants have been quick to adapt to Twitter and other new channels of communication [19]. Here are some of the techniques commonly used by spammers

1) Strategies of Twitter Spammers: Spammers use features exclusive to Twitter to increase their audience of potential victims [20], [21]. **a) User Timeline.** Infiltrating a user's timeline by getting a user to follow a spam account, the spam account's tweets will show on the user's timeline. **b) Direct Messages.** Direct messages are private communication between two users. These messages cannot be collected/analysed easily.

c) Hashtags Using hashtags (over 70 % of spam tweets in [20]), spammers try to initiate or infiltrate trending topics, read by a large audience. **d) Retweets and Mentions.** On Twitter, mentions are used to reference another user in a tweet. Tweets of other users can be shared through retweets. In contrast to strategies (a) and (b), a spammer does not need to be followed to retweet or mention another user. Thomas et. al. [10] reported that 58 % of users click on spam links in tweets they are mentioned in [10], partly explaining the popularity of this strategy.

2) Spam Accounts: The majority of spam accounts are supported by a growing underground market [22]. 56 % of the spam accounts become active immediately after registration, which indicated that spammers create accounts on demand [10]. Over 40 % have 0 followers while 89 % have less than 10 followers. In terms of tweet frequency [10] shows that there are two types of spam strategies: 34 % are short-lived accounts that flood as many tweets as possible, the rest are long-lived accounts with a low daily tweet count.

3) Success of Twitter Spam: It is estimated that the Click-Through-Rate of Tweet spam is higher compared to traditional email spam [23] . Grier et. al. [20] found, that some Twitter spam URLs receive large numbers of visitors, over 100,000 for a single URL, and 1.6 million for 6,000 URLs.

4) Underground Infrastructure : At the heart of Twitter spam are thousands of fraudulent accounts either compromised or created specifically for spam. Creating high numbers of accounts in bulk requires: (a) access to a diverse pool of IP addresses, (b) fraudulent email credentials to verify accounts and (c) CAPTCHA solving services. All readily available on web store fronts, blackhat forums or freelance labour sites [22].

Spammers also need infrastructure to host the landing pages of their campaigns and a large number of unique URLs to advertise them (to circumvent URL blacklisting). URL shortening services convert a long URL to a significantly shorter one that points to the same page. They are popular with spammers [10], [20], [22], [24] while also making it harder to collect the contents of the webpages.

5) Advertised Campaigns: Coordinated Action: By clustering accounts that posted URLs redirecting to the same landing page, [20] identified several spam campaigns. Even though they found that the majority of accounts did not collude with other accounts, some campaigns were advertised by multiple spam accounts.

6) Counter Measures: Grier et al. [20] analysed the time until posted spam URLs were flagged by three popular blacklisting services (Google Safebrowsing, URIBL and Joewein) and found that they were to slow to protect most victims. Twitter does not retroactively blacklist links, allowing malicious URLs to persist.

It is also important to note their limitations: the datasets used in the analysis might be biased as they rely on either blacklisting services or the undocumented Twitter suspension algorithm (black box). There is a need for further work based on different detection methods to complement the diverse picture of spam on Twitter.

III. The Bursty botnet dataset

Echeverria and Zhou recently reported a large botnet, called the 'Bursty botnet' [17], which consists of more than half a million bots showing the following properties.

a) User IDs between 5×10^8 and 5.35×10^8.
b) They only tweet in the first hour of registration.
c) They only tweet from the source 'Mobile Web'.
d) They mostly tweet a URL or/and a mention of another user.

We collected all the Bursty bots and their tweets in [17]. Table I shows selected properties of the dataset.

Figure 1 shows the Bursty botnet was active from 23 February 2012 to 24 March 2012 covering a period of one month. In this time the bots produced over 2.8 million tweets, the majority of which were posted during the first two weeks of the botnet's active period.

The Bursty bots, by definition, only tweeted within the first hour after registration. More than 80% of tweets were actually posted within the first 2 minutes. This strategy requires the creation of new accounts to maintain activity and is directly linked to the growth of the botnet. The red curve shows the size of the botnet (number of accounts) over time and resembles a step function that grows in large steps reflecting the botnet's peaks/bursts of activity.

IV. URLs posted by the Bursty botnet

In total, 2,823,743 unique URLs were extracted from the tweets posted by the Bursty Bots. Table I lists the 3 most frequent domain names, which together account for 74% of all URLs in the dataset. *Tinyurl.com* and *Bit.ly* are popular URL shortening services often abused by spammers. The

Number of bots	528,000	Bots with no friend	>99%
Number of tweets	2,881,370	Bots with no follower	>99%
Bots with URLs	>97 %	Bots with no retwees	>99%
Bots with mentions	>97 %	Bots with no reply	>99%
Tweets with URLs	>97 %	Bots with no hashtag	>99%
Tweets with mentions	>64 %	Tweets with no location tag	>99%

Most tweeted domain names	URL counts
tinyurl.com	1,179,369
google.com	562,557
bit.ly	327,985

Table I: Properties of the Bursty botnet on Twitter

google.com URLs are used as an open-redirect sending unsuspecting users to malicious sites, which is just another method for hiding the real destination.

Figure 2 plots the number of tweets containing URLs shortened with the three most common URL shortening services in the dataset. Temporal clustering is clearly visible: In the first ten days of the botnet's activity, almost exclusively tinyurl.com has been used to shorten the posted URLs. After that, a large cluster of URLs exploited google.com to redirect users to another landing page. In the last two weeks of the botnet's recorded activity, mainly bit.ly links were posted. It is likely that the low activity window in the middle of the plot is caused by Twitter suspending a cluster of the Bursty bots.

A. *tinyurl.com*

Of the over 1.1m tinyurl.com links posted by the botnet, more than half (57 %) were reported to be spam and consequently, the redirection was stopped by tinyurl.com. This is a strong indication that the Bursty botnet was indeed used to spread spam.

Notably, all of the remaining 0.5m URLs redirected to only two distinct landing pages: 503,672 direct to facebook-goodies.com and 1,937 to ggew.info. tinyurl.com does not provide any additional form of analytics such as click statistics.

B. *google.com*

The more than half a million google.com URLs tweeted by the Bursty botnet might be surprising at first, but a closer look at the URLs reveals that they exploit a less known feature of google.com that allows redirecting users to arbitrary web pages. Hence, those URLs essentially resemble the functionality of a URL shortener. As google.com enjoys an excellent reputation among users and web services, it is unlikely for web pages with this (sub)domain to be blocked by a blacklist of a malware- or spam-filter. As the URL starts with google.com, users might not suspect any malicious content.

In contrast to shortened tinyURL and bit.ly URLs, the google.com URLs (not to be confused with Google's own URL shortener goo.gl) tweeted by the Bursty bots are more complex, but they share the same basic structure, for example:

http://www.google.com/url?sa=t&rct=j&q=ggew.info&source=web&cd=1&ved=0CB8QFjAA&url=http%2F%2F**ggew.info**%2F&ei=HelQT6qNKdTO4QSa-6HyDQ\&usg=AFQjCNH-w263vKoOBPNOe8JRlkxWUKOzfA&9ilnpfrg=M1au9mv7m7

This URL requests the path `/url/` on the host google.com and provides an additional query with a set of key-value pairs (preceded by '?'), which are not publicly documented. However, the parameter 'q' provides another domain name (in this case 'ggew.info'), whereas the value of 'url' is an encoded URL.: *http://ggew.info/&ei=HelQT6qNKdTO4QSa-6HyDQ*.

Such URLs are carefully structured to utilise google.com's open redirect feature, which was/is used by Google internally to redirect users to the desired destination when they click on a search result link. It is not meant to be used from outside. To avoid abuse, Google usually displays a warning message to inform users about the redirect.

Users can then decide to follow the link or leave the page. However, if the 'usg' parameter is provided with the correct value (a hash function performed on the page itself), this message is *not* shown and the user is *automatically* forwarded to the page specified by whoever constructed the link without any further notice.

Apart from 22 links, 561,989 google.com URLs in the dataset redirected users to a single landing page: ggew.info, which was also found in the links shortened with tinyurl.com. This does not only show that bots tweeting the google.com URLs belong to the same botnet, but also that the bots tweeting tinyurl.com and google.com URLs are linked.

C. *bit.ly*

The Bursty bots tweeted about 330,000 bit.ly URLs. In contrast to the other shortening services, bit.ly also provides click statistics for each shortened link through their publicly available API. We were able to retrieve the final landing pages and click statistics for 97% of the bit.ly links.

Table II shows the most frequent landing pages and the corresponding number of bots and tweets that contained a URL to the landing pages. The top 3 landing pages account for over 97 % of bots and 83 % of all tweets respectively. Bit.ly was used to shorten more than 10k google.com redirection pages, which mainly point to two final destinations: ggvc.info and

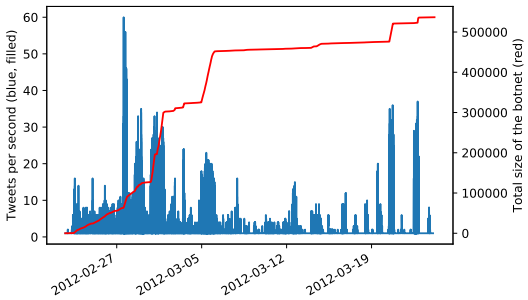

Figure 1: Tweeting activity of the Bursty botnet over time

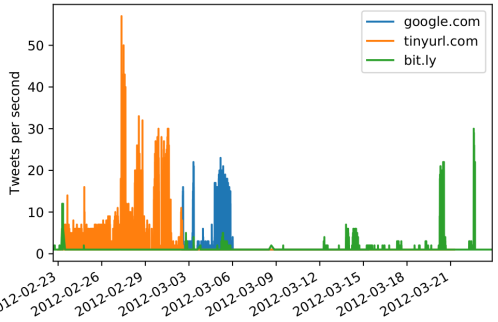

Figure 2: Usage of URL shortening services over time

Landing pages	Numbers of Bots	Tweets	Clicks
carucioare-copii.biz	44,153	174,979	13,442
gglw.info	12,751	50,442	20,104
google.com (redirecions)	10,167	40,137	-
ggqw.info	9,925	39,248	146,627
ggvc.info	227	889	107,040
salesrcs.com	315	9168	n/a
turnagainresources.com	153	3714	n/a

Table II: Most frequent landing pages of **bit.ly** URLs

ggqw.info. Table II also shows the registered clicks of the top landing pages.

Figure 3 shows the number of tweets containing bit.ly links (blue) and the registered clicks (red) on them during March 2012 based on bit.ly's click statistics. The number of tweets and clicks are clustered in a small number of distinct sharp peaks. A clear correlation between the number of bit.ly tweets and clicks is visible. This suggests that many tweeted links are clicked immediately after they were posted and became visible to potential victims.

Figure 4 shows the distribution of time differences between when a URL is tweeted and it receives its first click. The plot shows that almost all clicks happen within two days (∼2800 minutes) after a url is tweeted. It shows that most blacklisting solutions, with reported delays of up to 12 days [20], would not able to protect users from this malicious spam on Twitter.

Figure 5 shows the click distribution for the most-clicked landing pages over time. The extremely marked peak for 'carucioare-copii.biz' is particularly striking. The sudden interruption of clicks might indicate a form of throttling or blocking from either Twitter or bit.ly.

V. Spam Attack by the Bursty botnet

A. Tweet Mentions of the Bots

Over 99% of the Bursty bots mention less than 4 other users. A total of 1,313,008 other users were mentioned by the botnet. Over 85 % of the mentioned users were targeted exactly once, and about 12% were mentioned twice. This means that each of the Bursty bots targeted a small number of up to 4 seemingly random selected users in a tweet with a distinct, shortened URL. This is a strong indication that the popular mention spam strategy , as described in Section II-B1, was employed by the botnet to target over a million users.

B. Network of the Spam Attack

Based on data collected through the Wayback Machine[3], the Web of Trust and the WHOIS records[4], the network structure of the landing pages promoted by the Bursty botnet could be reconstructed as shown in Figure 6. The vast majority (over 2.2m) of URLs in the dataset promoted only two distinct spam campaigns: 'facebook-goodies.com' and 'daily-freebies.org'. All URLs were at least shortened through one public URL shortening service (or abusing google.com for redirects). Many of the shortened domains were redirected through another layer of domains, and finally forwarded users to these two landing pages. Apart from one exception, these middle layer domains follow a similar naming pattern starting with 'gg' followed by random letters. The additional layer of redirects seems to act as a 'pawn sacrifice'. That is, in case the resolved URLs are blacklisted, they can be easily replaced by another 'ggxx.info' domain redirecting to the same landing page. They also make the network of redirects more complicated and harder to resolve.

It is particularly striking that all of the custom redirect domains as well as the two final landing pages were registered and are owned by the same person. Even though this could be a fake identity or the contact details of a front man, it clearly ties the different landing pages and campaigns together. It is a very strong argument for the Bursty bots belonging to the same botnet that was created and operated by the same botmaster.

The final landing pages were deleted/suspended soon after the botnet's period of activity, and there are no useful snapshots of the final landing pages available on the Wayback Machine either. However, API querying of the Web of Trust returns poor reputation scores (2 out of 100) and that both websites have been flagged for spam. However, none of the websites is listed in the most common URL blacklists.

But the abandoned Facebook page[5] for 'facebook-goodies.com', specifically in the time of early 2012, revealed

[3]Wayback Machine (https://web.archive.org/) is a digital archive of the Web.
[4]WHOIS provides registration information on domain names.
[5]https://www.facebook.com/FB-Goodies-336592676352892/

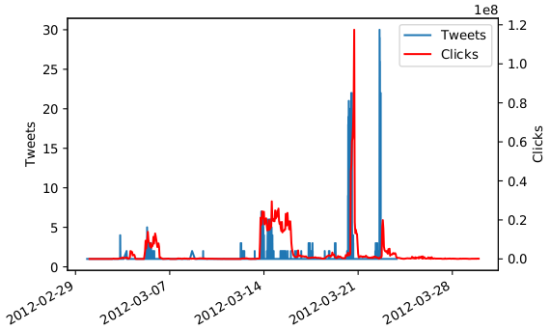

Figure 3: Tweets and registered clicks over time

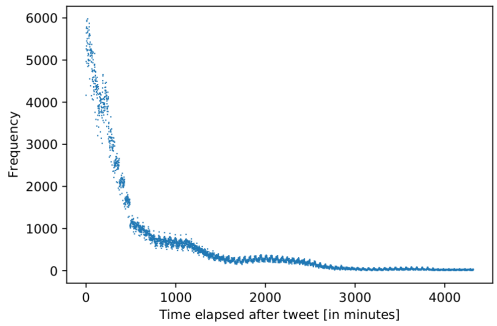

Figure 4: Distribution of time elapsed between tweet and click

Domain	Created at	Registrar	Registrant
carucioarecopii.biz	2011-08-26	Godaddy	Alexandru F.
daily-freebies.org	2011-12-20	Godaddy	Alexandru F.
facebook-goodies.com	2011-12-29	Godaddy	Alexandru F.
ggvc.info	2011-12-29	Godaddy	Alexandru F.
ggew.info m	2012-02-21	Unknown	Alexandru F.
gglw.info	2012-02-21	Unknown	Alexandru F.
ggqw.info m	2012-02-21	Unknown	Alexandru F.

Table III: Domain name WHOIS records of the Bursty botnet spam attack network.

reports[6] that some users received mention spam claiming they had won a gift card of a well-known brand. After clicking on the link they were asked to provide their personal data in return for the allegedly won gift card. This is a typical so-called survey scam, which has a comparatively high conversion rate. The stolen personal information is sold on the black market and/or used for identity theft. It should be noted that the facebook page has not been suspended/removed either, even while openly linking to blacklisted URLs.

C. Botmaster of the Bursty botnet

As [3] put it: 'If social bots are the puppets, additional efforts will have to be directed at finding their 'masters." Based on the analysis of the promoted spam campaigns, there is striking evidence leading to the alleged botmaster, who created and controlled this large botnet on Twitter and the cyber criminal ecosystem he was operating in.

Table III shows that according to the WHOIS records, all domain names used for both the final landing page and redirects, were registered by the same person. Alexandru F. (his full name is anonymised) registered over 440 other domains (including a number of very similar domain names) using the same email address, a valid Bucharestian postal address and a Russian telephone number.

Further research revealed that this threat actor has already come to attention to other security researchers for spam. He operates a proxy service, that offers access to a global IP pool of hundreds of thousands of compromised hosts, which explicitly advertises allowing customers to create Twitter accounts in bulk, without being throttled or blocked by Twitter. Alexandru F. is also active on 'blackhatworld.com', a forum that is known [22] to be a marketplace for spammers and other cybercimimals. Organised spam campaigns, bot(net) accounts and fake followers are among the services promoted on this website. In his over 1,600 posts, Alexandru F. not only promoted a proxy service and requested to buy 'installs' (compromised hosts to extend the IP pool), but also offered an automatic CAPTCHA solving service. As previous research by Thomas et. al. [22] showed, these are all vital tools to create and operate a large botnet on Twitter. All this evidence supports the hypothesis, that the domain owner is also the botmaster, and not just a front man or customer of the botnet, although the identity or parts of it could still be fake.

[6]https://nakedsecurity.sophos.com/2012/01/03/walmart-gift-card-survey-spam-twitter/

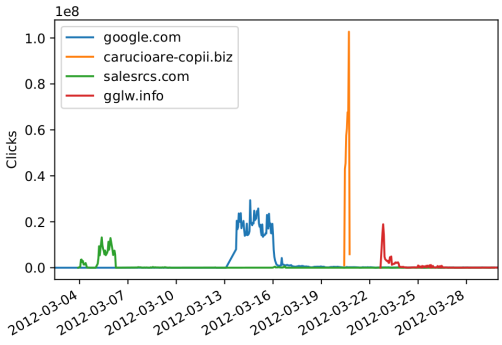

Figure 5: Clicks on landing pages over time.

D. The Hit-and-Run Attack Pattern

Based on all the information described above, it is possible to reconstruct how the attacker proceeded. Table IV outlines his actions and gives a rough time line of how this major spam campaign on Twitter unfolded.

The complete attack from registering the first domain until closing down the webpages took less than 4 months, and the actual spam campaign on Twitter took less than a single month. We refer to this as *'hit-and-run'* pattern, by which we mean that attackers plan and execute a large scale spam attack (e.g. by exploiting a new vulnerability) within days/weeks and then abandon the used accounts and with them remove all traces before updated detection could protect most of their victims.

VI. State-of-the-art classifiers against Bursty Bots

The "Botometer" Detection Framework (previously known as "BotOrNot?" [11]), is freely available through a public API. The "Botometer" Framework employs over 1,000 features of six different categories: user profile, friending, network, temporal, content and language features, which have been reported to provide valuable information to discriminate between bot and human-operated accounts [3], [11], [25]. Thus, it can be argued that it is the state-of-the-art in Twitter botnet detection.

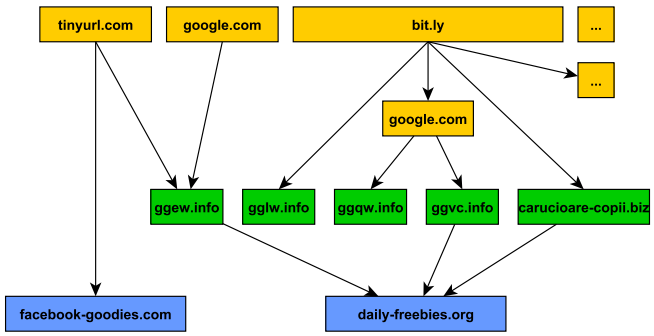

Figure 6: Network of URLs promoted by the Bursty botnet, everything in red owned by a single entity

Timing	Attacker's actions
Before Dec 2011	A large IP pool of 'installs' were acquired and an automatic CAPTCHA solver was developed
December 2011	The domains daily-freebies.org and facebook-goodies.com were registered
Dec - Feb 2012	The landing pages for the spam and phishing campaigns were created. Visitors were told they had won a gift card and had to complete surveys and fill in their personal data in exchange.
21. Feb 2012	A dozen ggXX.info domains were registered. They would be used as a second layer of redirects to easily control traffic flows, make detection harder and act as a 'sacrificial lamb' for black listing services.
Feb - Mar 2012	The Bursty botnet was created on Twitter. Each time new accounts were created they posted shortened URLs, moments after registration, to the ggXX domains and the landing pages of the campaign. Each tweet contains a single mention and a single shortened URL. Many users clicked the links within the first two days after they were posted (see Figure 4). By keeping a low profile the bots have evaded detection up until today.
April/May 2012	The first blacklisting services blocked the promoted URLs, some hosts suspended their web pages. The attacker had long moved on.
After May 2012	Later some of the web pages/domains were sold to new owners promoted with the artificially inflated click statistics.

Table IV: The 'Hit-and-run' attack pattern

	Count	Mean	Std	Min	25%	50%	75%	Max
RU-Eng	899	0.53	0.20	0.05	0.38	0.51	0.67	0.99
RU-Uni	899	0.47	0.23	0.01	0.28	0.44	0.72	0.99
BB-Eng	992	0.66	0.08	0.32	0.61	0.66	0.73	0.85
BB-Uni	992	0.30	0.07	0.19	0.27	0.29	0.31	0.79

Table V: Botometer Scores for Random Users **[RU]** and Bursty Bots **[BB]. [Eng]** Scores include English language features and **[Uni]** Scores are for language independent features

The Botometer API produces a score for each of the features discussed above. In addition, it provides two combined scores, one including language specific features and one without. All scores range from 0 to 1, 0 meaning the user is definitely not a bot. The authors report AUC ROC scores for these classifiers ranging from 0.89 to 0.95.

A. Experiment Setup

To measure the detection accuracy of the "Botometer" Framework, random samples of 1,000 accounts each were drawn from the The Bursty botnet and a random user dataset[7], as collected in [**?**]. These samples were evaluated using Botometer's API.

B. Classification Results

Botometer's API reported scores for over 85% of the accounts. Missing scores are likely deleted, suspended or private accounts.

Table V shows Botometer-'s results for the two samples, with and without English language features. When language features are included, the mean score for Bursty Bots (0.66) is somewhat higher than the mean score for the random users (0.53). However, this changes drastically if language features are ignored ("universal score"). Now, the scores suggest that it is much more likely that random users are bots (0.47) than the Bursty Bots (0.30). With this scores, Botometer would end up classifying more random users as bots, than actual bot accounts, leading to an accuracy below a random baseline. Having noted this shortcoming, in the remainder of this section, only scores including language features are considered. Given the similarity of the Bursty Bots, it is not surprising that the variance of their scores is much lower than for random users.

C. Unsupervised Clustering of Bursty Bots and Random Users

Figure 7 shows histograms of the Botometer scores, illustrating the difficulty to distinguish between random users and Bursty Bots. The two distributions overlap nearly completely, indicating that a detection of all (or most) bots would mean an unacceptable number of false positives. The Bursty bots distribution peaks around 0.6 to 0.7, whereas the random user distribution is broader and peaks around 0.5. [26] states that most normal accounts are in the range of (0.0,0.4) and most bots have scores in between (0.8,1.0), this cannot be reproduced in this experiment.

In order to use "Botometer" to suspend bots, a binary classification is required. Hence, a threshold for the continuous bot scores has to be determined. Varol et. al. [26] suggest thresholds ranging from 0.43 to 0.49 to discriminate between bots and human-operated accounts. If these thresholds were applied, most of the bots would indeed be detected (the recall is ranging from 97 % to 98 %), but also the majority of legitimate users (51 % to 63 %) would be suspended. This extremely high number of false positives is clearly unacceptable for any real world application. Figure 8 plots the Precision-Recall Curve for a binary classifier based on Botometer's scores with the threshold ranging from 0.01 to 1.0. If we optimise the threshold as a hyperparameter of the given dataset, the maximal AUC value is 0.66, well below the values reported by the authors of the system (0.89 to 0.95). If the binary classifier was based on the "universal" scores (excluding English language features), the AUC drops further to 0.30.

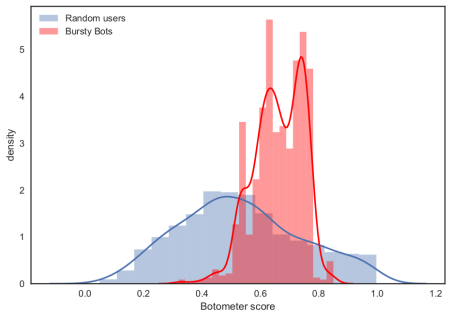

Figure 7: Score distribution for Bursty Bots and Random users

[7]under the assumption that the number of bots should be small

As Fig. 9 shows, score distributions vary widely and deviate from each other in most of the feature categories. The midspreads of the category scores for the Bursty bots are generally narrower and the number of outliers is higher. Apart from the “user" feature category, scores are mostly higher for the Bursty Bots (indicating a higher chance of them being bots). As expected, the scores for the “temporal" feature category are particularly indicative for the Bursty Bots and reflect their abnormal “bursty" tweeting pattern. The scores in the “sentiment" and “friend" feature categories are characterised by extremely narrow midspreads, indicating heavy tailed distributions.

This means that, even though the supervised learning approach failed to accurately detect the Bursty Bots, there are still noticeable differences from random users. It might still be possible to cluster Bursty Bots and Random users based on the distribution of their category scores (unsupervised). Figure 10 plots the two-dimensional t-SNE embedding of the feature category vectors for Bursty Bots and Random users. The t-Distributed Stochastic Neighbor Embedding (t-SNE) [27] is a technique for dimensionality reduction, which is particularly popular for visualisation. Bursty Bots and random users are clustered in a small number of relatively dense clusters. The unsupervised identification of clusters is not sufficient to generate correct cluster labels in order to use this approach to suspend bots.

Another interesting way of clustering the Bursty Bots, that could prove to be promising is by their temporal activity pattern, similar to the approach described in [28]. The distinct ’burst’ pattern of their activity might help to identify lockstep behaviour, and thus uncover the botnet as a whole rather than relying on account-level based classification.

VII. Conclusion

The Bursty botnet is one of the largest Twitter botnets reported in the academic literature. Created in early 2012, it counts over 500,000 accounts. The basic principle of how the Bursty bots work might not appear particularly sophisticated. The botmaster undertook little effort to make the bot accounts look like they were operated by humans (no profile picture or description, barely any friends or follower). But, because of their low-profile actions (only a small number of tweets in the first hour after registration) and, as a result, the scarcity of account data, the Bursty bots were well hidden until now, and many of them are not even suspended.

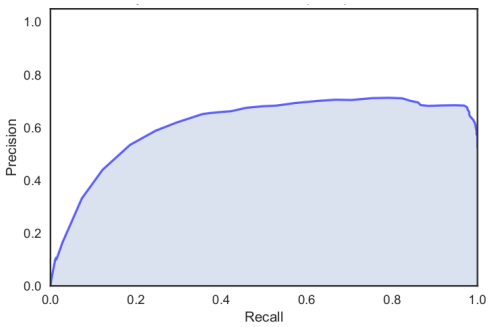

Figure 8: Precision-Recall Curve for Binary Classifier

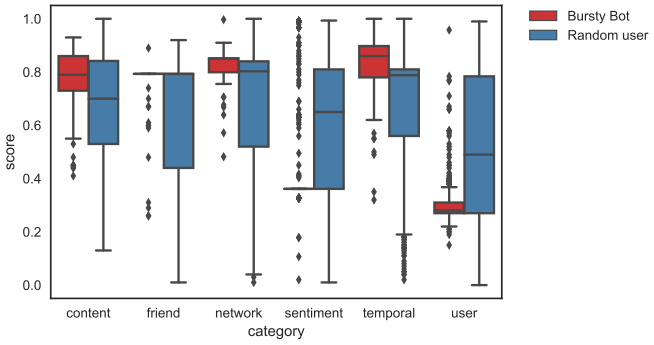

Figure 9: Box plot of category scores for Bursty Bots and Random users

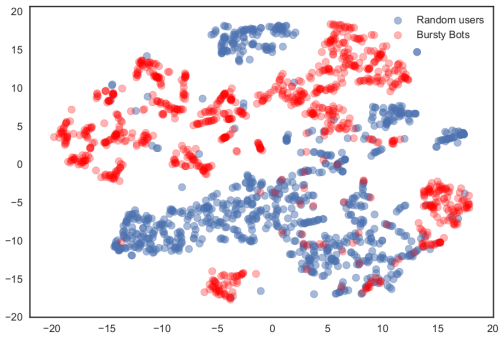

Figure 10: t-SNE embedding of Bursty Bots and Random users

After an in-depth analysis of the botnet’s activity and content, a strategy as simple as it is effective was identified: each bot posted a small number of tweets containing a mention of another user and a shortened URL. The tweets appeared on the timelines of over 1.3m targeted users with obfuscated URLs. Further research showed that the landing pages of the URLs were operating a survey scam claiming visitors had won a ‘gift card’. The analysis of the promoted spam campaigns did not only show that the Bursty bots do indeed belong to the same, large botnet, but also pointed to a threat-actor that, beyond any reasonable doubt, was the botmaster. This research also uncovered that the botmaster is still active in black hat forums and the deep web, and is still selling bot related services.

Most victims clicked on the malicious link within the first two days after being targeted, traditional blacklists are too slow to protect the majority of users from the ‘hit-and-run’ attack pattern employed by the Bursty botnet. The complex network of redirects makes blacklisting even harder. There is a need for designing new counter measures that can effectively mitigate ‘hit-and-run’ spam attacks like that of the Bursty botnet in a reasonable amount of time.

Future work needs to include analysis on the targets of bot links, algorithmically following all redirects to the final

landing page. Furthermore, a shared owner for several landing pages on a spam campaign is likely a good feature to include in further botnet classification strategies.

The need for thinking of new features for bot detection is more important than ever, given that botmasters will only get more sophisticated and creative with their botnet design. There is little doubt that this is an arms race.

It is reported that botnets used for the purpose of political censorship and propaganda adopt the same strategies and rely on the same infrastructure as social spam bots [29]. The Bursty botnet is a clear warning about how easy it is to create a successful and lucrative Twitter botnet. This simple botnet spanning hundreds of thousands of accounts is still unsuspended. It is very probably that it took more effort to detect this botnet, than it took to create it.

In summary, this work provides a comprehensive insight into how a major spam attack on the Twitter social network was carried out, covering the whole life cycle from its planning to execution. This work draws a realistic picture of the success and potential threats of a large Twitter botnet. It does not only provide an important contribution to understanding the inner workings of a large botnet, but also how cybercriminals operate and the infrastructure and underground markets they rely on. The unique traits of the Bursty bots that were uncovered as a result of the analysis can be used to improve relevant systems.

The poor performance of state-of-the-art supervised bot detection systems shows that overly sophisticated techniques are not necessary in order for bots to remain undetected. It suggests that the characteristics of botnets vary so widely and are so different, that at least up to now, generic classifiers cannot be employed in a large-scale or real world scenario. This is especially true for an account-by-account based annotation, as bots of the same botnet have more in common than bots and legitimate users have differences. However, it is possible, as shown in this work, to aid humans to find abnormalities by identifying patterns in informative distributions.

VIII. Acknowledgments

Shi Zhou and Christoph Besel are partially supported by MediaGamma Ltd.

References

[1] J. Zhang, R. Zhang, Y. Zhang, and G. Yan, "The rise of social botnets: Attacks and countermeasures," *IEEE Transactions on Dependable and Secure Computing*, 2016.

[2] F. Morstatter, H. Dani, J. Sampson, and H. Liu, "Can one tamper with the sample api? - toward neutralizing bias from spam and bot content."

[3] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[4] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012, pp. 15–15.

[5] H. Bansal and M. Misra, "Sybil detection in online social networks (osns)," in *Advanced Computing (IACC), 2016 IEEE 6th International Conference on*. IEEE, 2016, pp. 569–576.

[6] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based sybil defenses," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 363–374, 2010.

[7] A. H. Wang, "Detecting spam bots in online social networking sites: A machine learning approach." *DBSec*, vol. 10, pp. 335–342, 2010.

[8] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, "Social turing tests: Crowdsourcing sybil detection," *arXiv preprint arXiv:1205.3856*, 2012.

[9] J. P. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?" in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 2014, pp. 620–627.

[10] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: an analysis of twitter spam," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 243–258.

[11] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 273–274.

[12] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network sybils in the wild," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 8, no. 1, p. 2, 2014.

[13] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 963–972.

[14] M. Washha, A. Qaroush, M. Mezghani, and F. Sedes, "Information quality in social networks: Predicting spammy naming patterns for retrieving twitter spam accounts," 2017.

[15] J. Echeverria and S. Zhou, "Discovery, Retrieval, and Analysis of the 'Star Wars' Botnet in Twitter," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, ser. ASONAM '17.

[16] N. Abokhodair, D. Yoo, and D. W. McDonald, "Dissecting a social botnet: Growth, content and influence in twitter," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015, pp. 839–851.

[17] J. Echeverria, C. Besel, and S. Zhou, "Discovery of the twitter bursty botnet," *Data Science for Cyber-Security*, 2017.

[18] E. Ferrara, "Disinformation and social bot operations in the run up to the 2017 french presidential election," 2017.

[19] S. Yardi, D. Romero, G. Schoenebeck *et al.*, "Detecting spam in a twitter network," *First Monday*, vol. 15, no. 1, 2009.

[20] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@ spam: the underground on 140 characters or less," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 2010, pp. 27–37.

[21] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proceedings of the 26th annual computer security applications conference*. ACM, 2010, pp. 1–9.

[22] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson, "Trafficking fraudulent accounts: The role of the underground market in twitter spam and abuse." in *USENIX Security Symposium*, 2013, pp. 195–210.

[23] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamalytics: An empirical analysis of spam marketing conversion," in *Proceedings of the 15th ACM conference on Computer and communications security*. ACM, 2008, pp. 3–14.

[24] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time url spam filtering service," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 447–462.

[25] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, "The darpa twitter bot challenge," *Computer*, vol. 49, no. 6, pp. 38–46, 2016.

[26] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," *arXiv preprint arXiv:1703.03107*, 2017.

[27] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[28] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copycatch: stopping group attacks by spotting lockstep behavior in social networks," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 119–130.

[29] K. Thomas, C. Grier, and V. Paxson, "Adapting social spam infrastructure for political censorship." in *LEET*, 2012.