Discovery of the Twitter Bursty Botnet

Juan Echeverria[1*], Christoph Besel, Shi Zhou[2*]

Department of Computer Science
University College London (UCL), London, United Kingdom
[1*] j.guzman.11@ucl.ac.uk
[2*] s.zhou@ucl.ac.uk

Abstract

Many Twitter users are bots. They can be used for spamming, opinion manipulation and online fraud. Recently, we discovered the *Star Wars botnet*, consisting of more than 350,000 bots tweeting random quotations exclusively from Star Wars novels. The bots were exposed because they tweeted uniformly from any location within two rectangle-shaped geographic zones covering Europe and the USA, including sea and desert areas in the zones. In this paper, we report another unusual behaviour of the Star Wars bots, that the bots were created in bursts or batches, and they only tweeted in their first few minutes since creation. Inspired by this observation, we discovered an even larger Twitter botnet, the *Bursty botnet* with more than 500,000 bots. Our preliminary study showed that the Bursty botnet was directly responsible for a large-scale online spamming attack in 2012. Most bot detection algorithms have been based on assumptions of 'common' features that were supposedly shared by all bots. Our discovered botnets, however, do not show many of those features; instead, they were detected by their distinct, unusual tweeting behaviours that were unknown until now.

1 Introduction

Twitter bots are Twitter user accounts created and controlled by hackers, called botmasters, using computer programs. A Twitter botnet is a group of bots that show the same properties and are controlled by the same botmaster. The Twitter company has identified and removed millions of bots. Researchers claim there are much more bots on Twitter [15], and they have introduced a number of methods to detect Twitter bots.

Twitter bots can pose a series of threats to cyberspace security [15]. For example, they can send a large amount of spam tweets to other users; they can create fake treading topics; they can manipulate public opinion; they can launch a so-called astroturfing attack where they orchestrate false 'grass roots' campaigns to create a fake sense of agreement among Twitter users [8, 18, 1]; and they can contaminate the data from Twitter's streaming API [16] that so many research works have been based on; they have even been linked to election disruption [2].

Recently we discovered the *Star Wars botnet* [7] with more than 350,000 bots. These bots were discovered because they tweeted uniformly from any location within two rectangle-shaped geographic zones covering Europe and the USA, including sea and desert areas in the zones. Further inspection showed the bots only tweeted random quotations from Star Wars novels. This botnet was discovered and detected in a way completely different from previous bot detection efforts.

In this paper, we report our discovery of another Twitter botnet, the *Bursty botnet*, which is even larger – with more than 500,000 bots that have not been banned or removed by Twitter the time of writing. The discovery of the Bursty botnet was inspired by another unusual tweeting behaviour of the Star Wars bots. Our preliminary study showed that the Bursty botnet was directly responsible for a large-scale online spamming attack in 2012.

Our work not only provides valuable ground truth data for research on Twitter bots, but also enabled us to reflect on the limitations of existing methods for detecting Twitter bots. Existing methods are mostly based on 'common' features that were supposedly shared by all Twitter bots. Our newly discovered botnets, however, do not show any of those features; instead, they were detected by their distinct, unusual tweeting behaviours that were unknown until now.

2 Background

2.1 Twitter bots detection

Twitter, as a company, has been actively identifying and removing suspicious users, many of which are spammers or bots [21]. Researchers have also proposed various methods to classify or detect Twitter bots, many of which are machine learning methods. Many of these works were based on features either measured from ground truth data or assumed by researchers as common properties of all bots. Such features include user properties, including username length [13], user profile [25], time between tweets [4], Levensthein distance between the

tweets of a user [23], distribution of tweeting frequency [4] and entropy in tweeting frequency [6]; as well as tweet text properties [4], including topics [6] and sentiment analysis [6]. Some researchers claimed that their bot classifiers can achieve high accuracy, as high as 99.99% [22].

The existing efforts on Twitter bots detection, however, suffer a major problem, that despite their ever more sophisticated algorithms, they could only discover small numbers of (mixed types of) bots, mostly hundreds or up to a few thousands – and yet it is well known that there are very large botnets on Twitter [17].

The performance of bot detection methods is ultimately determined by our knowledge and understanding of Twitter bots. However, there is a well acknowledged lack of ground truth data [20]. Twitter's bot datasets are not available to the public. The small number of available datasets are small and contain mixed types of bots [15]. As a result, it is not clear which assumptions on Twitter bots are true, which features are most characteristic, or whether Twitter bots should have common features at all.

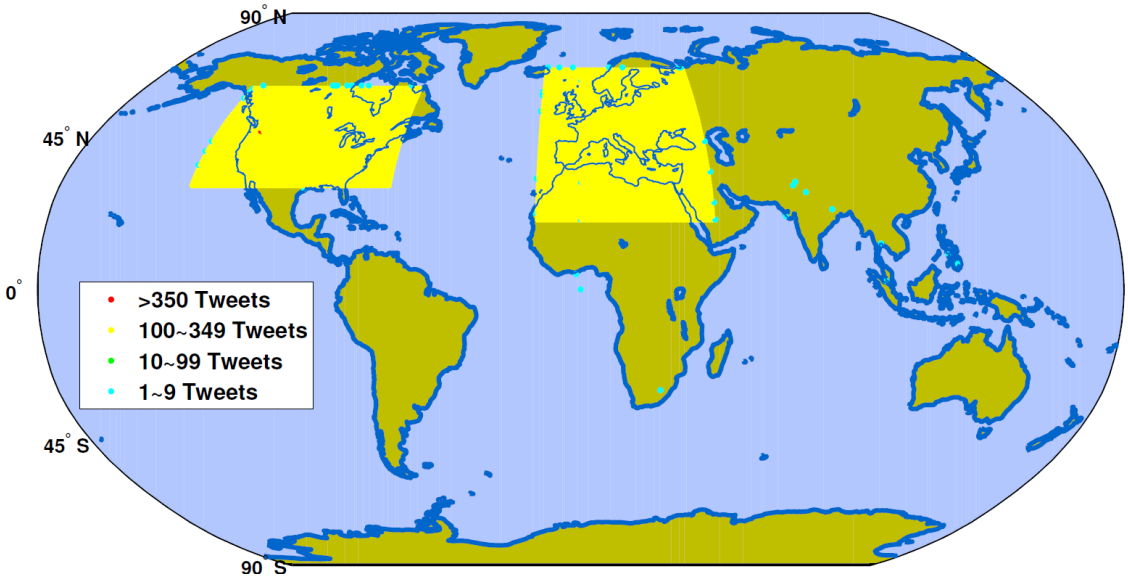
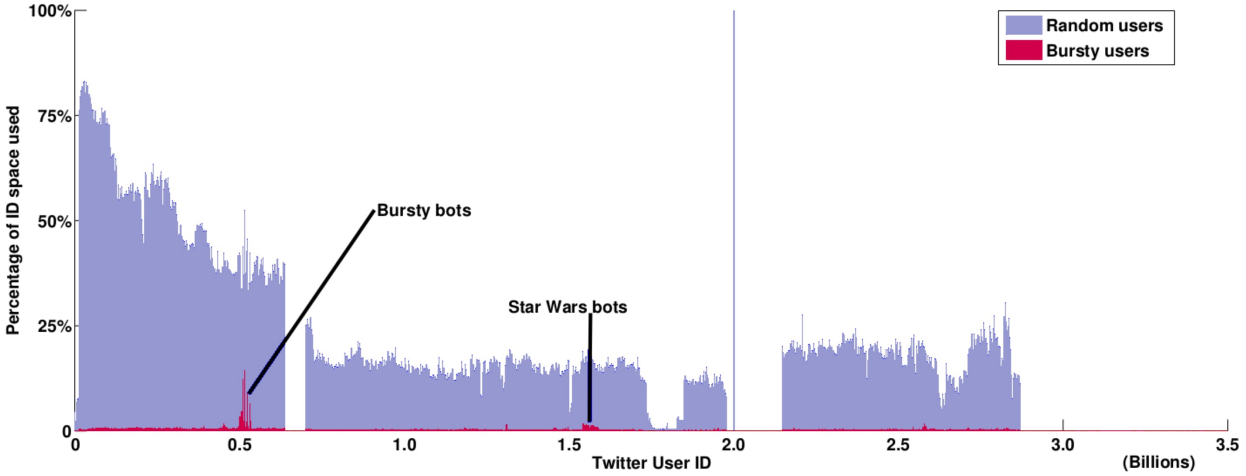


Figure 1: Distribution of tweet locations of the Star Wars bots. The bots exclusively, uniformly tweet from any location within the two rectangle zones over Europe and North America, including sea and desert areas. This unusual tweeting behaviour was the first clue for discovering the Star Wars bots.

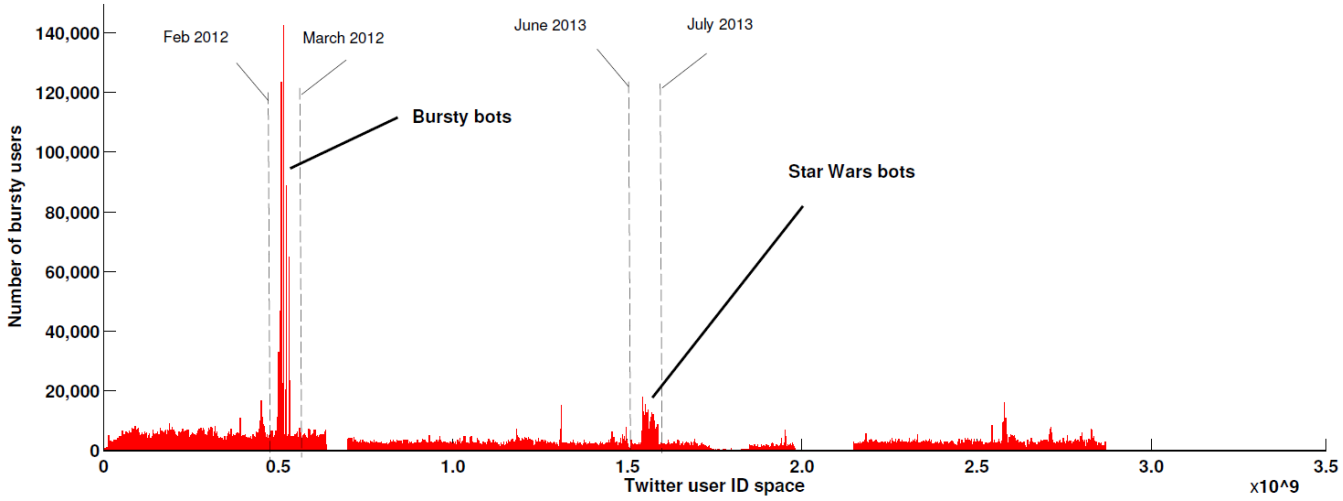
2.2 The Star Wars botnet

Recently we discovered a large botnet, the Star Wars botnet [7]. This discovery stems from the fact that the Star Wars bots create clear rectangular patterns when plotting their location-tagged tweets on a world map. After manual tagging of some of the bots, a Naive Bayes classifier was trained and some filters were applied. Finally, more than 350,000 Star Wars bots were identified showing the following properties.

- The Star Wars bots exclusively and textually tweet quotations from Star Wars novels, with the only exception of inserting the hash character (#) or special hashtags (such as #followme) at random places.
- The bots were registered with Twitter from June to July 2013, and therefore their user ID numbers are within a narrow range between 1.5×10^9 and 1.6×10^9.
- The tweeting source of the bots is exclusively 'Windows Phone', which accounts for only 0.02% of all tweets on Twitter.
- About half of all tweets of the bots are location-tagged. Almost all of the bots have tweeted at least one location-tagged tweet. The tweet location tags are obviously fake because all of them fall in one of the two rectangle zones, uniformly (see Figure 1).
- Because of the fake locations, distance between two consecutive geo-located tweets of a bot is very large, over 2000 km on average.
- The bots have created no more than 11 tweets in their lifetime; and they have no retweets or mentions.
- The Star Wars bots disproportionately follow other Star Wars bots (the botnet follows itself). They have at most 10 followers and at most 31 friends.



(a) Percentage of Twitter user IDs allocated to all users and bursty users in each bin of 1 million IDs.



(b) Number of user IDs allocated to bursty users.

Figure 2: Distribution of the bursty users in the Twitter user ID space in September 2015. A blue line in (a) near 2.0×10^9 are IDs reserved by Twitter.

The Star Wars botnet provides a valuable dataset for studying Twitter bots. It is not only very large, but also contains a single botnet showing the same properties. It provides rich information and clues on how the botnet is designed and created [7]. The Star Wars bots do not show many of the previously assumed 'common' features of Twitter bots; instead they exhibit a number of unusual tweeting behaviours that have not been reported. The discovery of the botnet was accidental, but it illustrates the limitations of existing bot detection methods.

3 Discovery of the Bursty botnet

In this section we will show how we detect another botnet, with over 500,000 users. This was inspired by an unusual tweeting behaviour that we observe on the Star Wars bots: all the bots only tweeted in the first few minutes after their creation. We call it the bursty tweeting behaviour. We define the *bursty* users as those who tweeted at least 3 times in the first hour following user creation, and then never tweeted again. Figure 2 shows the distribution of bursty users in the Twitter user ID space. When a Twitter user is created, Twitter assigns the user a unique ID number between 0 and 2^{32}. In general, the user IDs are allocated in a sequential manner such that a smaller ID means an earlier creation date.

We can see in Figure 2(a) the bursty users are a tiny fraction of Twitter users. As shown in Figure 2(b), the Star Wars bots can be clearly identified as a cluster of spikes with IDs in the range between 1.5×10^9 and

Figure 3: Screenshot of a Bursty Bot live on Twitter.

1.6×10^9, which corresponds to June and July 2013, the period when the Star Wars bots were created. This is as expected, because Star Wars bots are bursty users.

However, in Figure 2(b), there is another much more prominent cluster of spikes. These bursty users were created in February and March 2012 within the ID range between 5.00×10^8 and 5.35×10^8. We manually checked these users and noticed that many of them showed clear characteristics of spamming bots because their tweets contained only a mention and/or a URL. Many of the URLs are shortened and pointed to blocked domains. An example of a bursty bot can be seen in Figure 3. Further study showed that these suspicious users were members of an unknown botnet, we call it the *Bursty* botnet.

3.1 Definition of the Bursty bots

Here we define the Bursty bots as Twitter users with all of the following properties.

- They were registered in February and March 2012 with user IDs between 5×10^8 and 5.35×10^8.
- They show bursty tweeting behaviour. That is, they generated at least three tweets and they only tweeted in the first hour after their creation.
- They only tweeted from the source of 'Mobile Web'.
- They mostly tweet (i) a URL; or/and (ii) a mention of another user.

We retrieved about 21 million users in the ID range of 5×10^8 and 5.35×10^8. We collected all of their user account information and all their tweets up to September 2016. According to the above definition, we identified more than 500,000 Bursty bots.

3.2 Bursty tweeting and bursty creation

The Bursty botnet shows two bursty properties: they tweeted in a bursty way and they were created in bursts. Figure 4 shows most the Bursty bots and the Star Wars bots only tweeted in the first few minutes since their creation. In particular, the Bursty bots show a strong bursty behaviour that almost all tweets were generated less than 2 minutes after account creation and then stayed silent forever. This is a clear sign of automatised behaviour. This feature also gives us a sharp time frame to look for the possible cyber attacks that the bots were created for.

Figure 5a shows that most of the bursty users created in February and March 2012, with user IDs from $500{\times}10^6$ to $535{\times}10^6$, are identified as the Bursty bots, showing all the properties defined above. Only a small, stable number of bursty users are not identified as Bursty bots. These are perhaps normal users who accidentally joined Twitter at that time, tried a few tweets and then never return again. Such users appear with a low and almost constant rate throughout the time.

Although we defined the bursty users as those who tweeted only in the first hour, we now know that most of the Bursty Bots actually tweeted only in the first two minutes. The fact that almost all bursty users can be identified as the Bursty Bots means that bursty tweeting is indeed an unusual, distinct behaviour of this botnet.

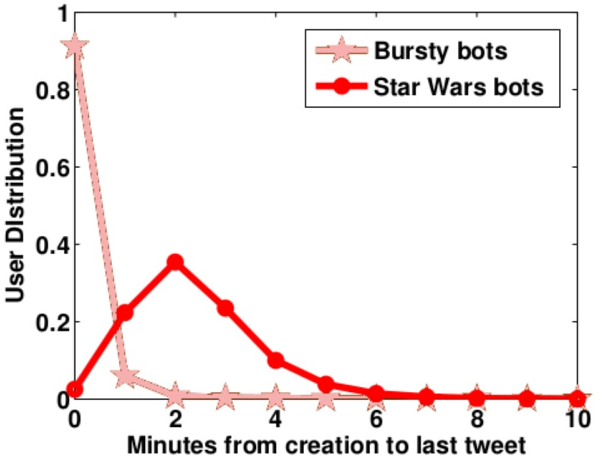


Figure 4: Distribution of the Bursty bots and Star Wars bots as a function of time (in minutes) from their creation to their last tweet. Most of the bots only generated tweets in their first few minutes.

Property	**Value or percentage**
Bots with no friend	99.0%
Bots with no follower	99.0%
Bots with tweets with URLs	98.0%
Bots with tweets with mentions	97.5%
Bots with tweets with hashtags	0.8%
Average number of tweets	4.74
Total number of tweets	2.8 million
Tweets with URLs	97.6%
Tweets with mentions	64.1%
Tweets with hashtags	2.7%

Table 1: Properties of the Bursty bots.

3.3 The 'disappeared' Bursty bots

We collected the bursty users in September 2015 and again in September 2016. As shown in Figure 5b, about 300,000 Bursty bots have disappeared during that period. Notably, there is a whole spike of Bursty bots missing with IDs between 520×10^6 and 525×10^6. It is likely that their accounts were removed by Twitter, we checked many of their accounts, which are indeed suspended. On one hand, this supports our detection of the Bursty bots as computer-controlled, malicious users; on the other hand, it shows that Twitter has not identified the Bursty botnet as a whole, leaving the majority of the botnet still alive online.

3.4 Unusual connectivity

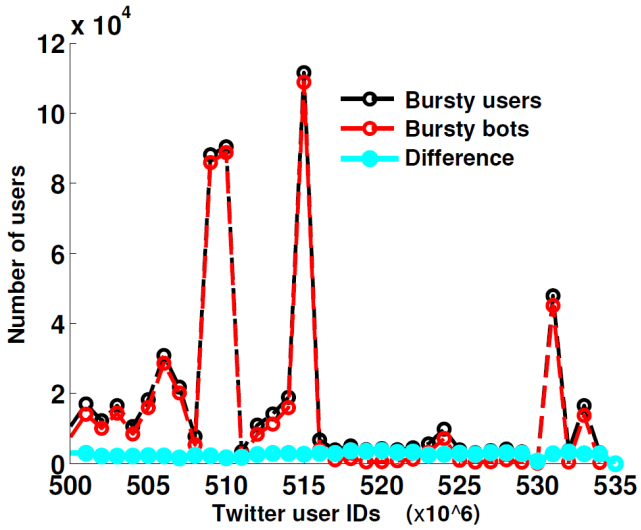
Table 1 shows the Bursty bots have a number of distinct properties. Most of the bots do not have any followers (outgoing links) or friend (incoming links). It is notable that the definition of the Bursty Botnet did not involve user connectivity, yet the detected Bursty Bots exhibited such an unusual feature.

This feature is against a popular assumption in previous studies that Twitter bots should tend to have many connections. This feature, however, is expected from the bursty creation of the Bursty Bots, as they were designed to be used only once immediately after they were created.

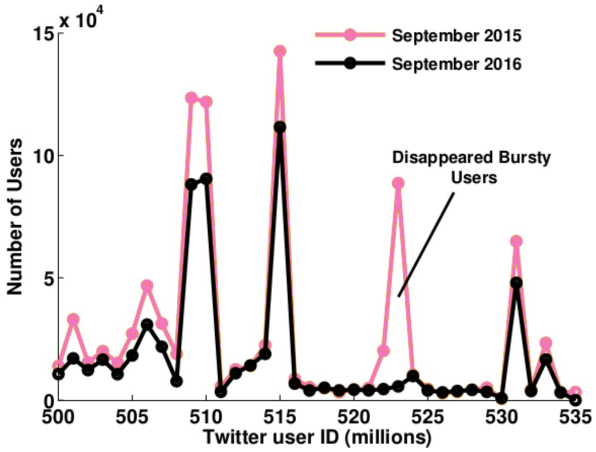
3.5 The Bursty botnet spamming attack

It is notable that almost all of the tweets generated by the bots contain a URL; and about 2/3 have a mention. This means that almost all of the tweets that have a mention also have a URL. This indicates that the bots were likely created for spamming attacks, the mentions were used to maximise the reach of the tweet, both by attracting the user being mentioned and his followers to click on the URL.

To find out more details, we examined all the URLs tweeted by the Bursty bots. Of the 2.8 million tweets that the Bursty botnet has created, almost all (over 99.9%) of the URLs were unique, which means most of the URLs were only tweeted once by a single bot. As shown in Table 2, when divided by domain, the most tweeted domain is the URL shortening and redirect service, `tinyurl.com`, with over 42% (or 1.18million) of the total



(a) Bursty users and the identified Bursty bots as measured in September 2016. Most bursty users are Bursty bots.



(b) Bursty bots in September 2015 and September 2016. Many bursty bots disappeared during that time.

Figure 5: Bursty users and the Identified Bursty bots in the Twitter user ID range of 500×10^6 and 535×10^6.

Domain	Count
tinyurl.com	1,179,369
google.com	562,557
bit.ly	328,016
dietagolder670.ru	54,585
goroskopsiris2346.ru	54,414
dietagoliu4758.ru	52,992
dietaseru858.ru	51,894

Table 2: Domains most tweeted by the Bursty bots

URLs. We investigated the tinyurl links, and found that 99.9% of them pointed to only two destinations: one was a webpage that had been blocked by tinyurl, which means tinyurl had classified it as malicious or spam; the other is a known phishing webpage `www.facebook-goodies.com`[1]. By performing a content analysis, we found that the vast majority of all the URLs could be clustered into only two distinct spam campaigns.

It is almost certain that the Bursty botnet was carefully designed and centrally controlled for the purpose of a spamming attack. A number of tricks have been used to hide the attack.

Firstly, the bots were created in large numbers, and each bot was used to generate a small number of tweets in the first few minutes only and then the bots all became silent. Most existing bot detection methods are not able to identify such inactive bots.

Secondly, the bots used a complex network of URL shorteners and redirects to obfuscate the final landing pages, such that the vast number of URLs were used only once, which could effectively evade most spam filters. Also it was not easy for users to tell on the final destination of the URLs.

Thirdly, the botnet directly targeted over 1.3m distinct Twitter users by mentioning their usernames, which significantly increased the chance of the URLs being clicked. Our analysis revealed just how successful this technique was: on average over 61% of the posted URLs that lead to a phishing campaign were clicked, which could yield a remarkable revenue by selling stolen personal data.

With the above information and further research, we were even able to track down the alleged botmaster of the Bursty Botnet. Our detailed analysis on the spamming attack of the Bursty botnet will be published in another paper.

4 Reflection on Twitter bot detection

4.1 Failure of existing detection methods

In recent years there have been many efforts to detect Twitter bots. Some have produced plausible results. Most of them relied on heuristic assumptions on 'common' features that should supposedly be shared by all bots. It is clear that these assumed features are not shared by all bots. For example the Bursty botnet and the Star Wars botnet show some properties that are diametrically different from those assumptions. As a result the the existing methods have not been successful in detecting large botnets. We have verified that the Bursty bots and the Star Wars bots can fool one of the latest and more advanced bot detection tools [5].

One reason was that previous studies were restricted by the lack of ground truth data. Since the available datasets all contained mixed types of bots, researchers had to search for and focus on general features shared by all bots in the datasets.

Another possible reason was that the previously studied features of bots have been 'fading', because most of these features could be easily avoided in the design of later botnets as botmasters must have closely followed the development on bot detection.

4.2 A long-term battle with no silver bullet

The Bursty botnet and the Star Wars botnet exhibit distinctive properties that have been overlooked so far, namely the tweet location distribution and the bursty tweeting behaviour. But we do not expect that further study on these features will lead to discovery of many other new botnets. There is a strong incentive for botmasters to deliberately create new botnets that do not show any of the features that have already been 'exposed' by researchers. Ideally botmasters would create new botnets that are completely different from existing ones.

[1]This website has now been deleted, but it is available through Wayback Machine, arguably the most comprehensive digital archive of the World Wide Web.

Indeed, we expect the battle on bot detection to be a long-term process, where researchers have to keep proposing new detection methods to catch up with new generations of botnets, which are likely to become ever more deceptive. As such, although it has been highly desired for by the research community, we do not believe it will be possible to develop a 'generalised' method to detect all types of bots.

5 Conclusion

The discovery of the Bursty botnet and the Star Wars botnet provided valuable ground truth data for the Twitter bot research community. Both botnets are unusually large. Each contains hundreds of thousands of bots. They are different from other datasets because each of the botnets contains a single network of bots that exhibit the same properties and were created and controlled by the same botmaster. It seems Twitter has removed the Star Wars botnet since our publication [7]. As of this writing, most of the 500,000 Bursty bots are still alive on Twitter. Researchers can collect them by following instructions in this paper. Researchers, however, should hurry to collect them before Twitter deletes these accounts too.

It is interesting to point out that the Bursty Botnet and the Star Wars Botnet were discovered by their unusual tweeting behaviours, which is in a rather 'unconventional' way that was different from previous bot detection efforts. This also means it does not suffer the biases induced by relying on Twitter's black-box suspension algorithm or URL blacklisting services, which makes it a valuable ground truth for future research.

These new datasets not only enabled us to reflect on the previous assumptions and detection methods, but also provided us a rare and valuable opportunity to investigate how Twitter bots were designed, created, and used for a spamming attack in the cyberspace.

References

[1] 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*, ser. CSCW '15. ACM, pp. 839–851.

[2] A. Bessi and E. Ferrara, "Social bots distort the 2016 u.s. presidential election online discussion" [Online]. Available: http://journals.uic.edu/ojs/index.php/fm/article/view/7090

[3] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy" in *Fourth International AAAI Conference on Weblogs and Social Media*,

[4] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on twitter: human, bot, or cyborg?" in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. ACM, pp. 21–30.

[5] 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*, ser. WWW '16 Companion. International World Wide Web Conferences Steering Committee, pp. 273–274.

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

[7] J. Echeverria and S. Zhou, "Discovery, Retrieval, and Analysis of 'Star Wars' botnet in Twitter." *I*n Proc. of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Available: https://128.84.21.199/abs/1701.02405

[8] E. Ferrara, "Manipulation and abuse on social media" pp. 4:1–4:9. [Online]. Available: http://doi.acm.org/10.1145/2749279.2749283

[9] C. A. Freitas, F. Benevenuto, S. Ghosh, and A. Veloso, "Reverse engineering socialbot infiltration strategies in twitter."

[10] S.-J. L. Gratton, "Follow me! creating a personal brand with twitter" 1st ed. Wiley Pub., Inc.

[11] 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*, ser. CCS '10. ACM, pp. 27–37.

[12] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on twitter" in *Fifth International AAAI Conference on Weblogs and Social Media.*

[13] S. Lee and J. Kim, "Early filtering of ephemeral malicious accounts on twitter" [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2014.08.006

[14] J. Messias, L. Schmidt, R. Oliveira, and F. Benevenuto, "You followed my bot! transforming robots into influential users in twitter"

[15] Morstatter Fred, Carley Kathleen, and Liu Huan. ASONAM 2015 bot detection tutorial. [Online]. Available: http://www.public.asu.edu/~fmorstat/bottutorial/

[16] F. Morstatter, H. Dani, J. Sampson, and H. Liu, "Can one tamper with the sample API?: Toward neutralizing bias from spam and bot content" in *Proceedings of the 25th International Conference Companion on World Wide Web*, ser. WWW '16 Companion. International World Wide Web Conferences Steering Committee, pp. 81–82.

[17] Narang, Satnam. Green coffee and spam: Elaborate spam operation on twitter uses nearly 750,000 accounts. [Online]. Available: http://www.symantec.com/connect/blogs/green-coffee-and-spam-elaborate-spam-operation-twitter-uses-nearly-750000-accounts

[18] J. Ratkiewicz, M. Conover, M. Meiss, B. Gon{\textbackslash}ccalves, S. Patil, A. Flammini, and F. Menczer, "Truthy: mapping the spread of astroturf in microblog streams" in *Proceedings of the 20th international conference companion on World wide web*, ser. WWW '11. ACM, pp. 249–252.

[19] G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and B. Y. Zhao, "Follow the green: Growth and dynamics in twitter follower markets" in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. ACM, pp. 163–176.

[20] V. S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, A. Stevens, A. Dekhtyar, S. Gao, T. Hogg, F. Kooti, Y. Liu, O. Varol, P. Shiralkar, V. Vydiswaran, Q. Mei, and T. Hwang, "The DARPA twitter bot challenge." [Online]. Available: http://arxiv.org/abs/1601.05140

[21] 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*, ser. IMC '11. ACM, pp. 243–258.

[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" pp. 195–210.

[23] A. H. Wang, "Detecting spam bots in online social networking sites: A machine learning approach" in *Data and Applications Security and Privacy XXIV*, ser. Lecture Notes in Computer Science, S. Foresti and S. Jajodia, Eds. Springer Berlin Heidelberg, no. 6166, pp. 335–342

[24] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter" in *Proceedings of the 21st international conference on World Wide Web*, ser. WWW '12. ACM, pp. 71–80.

[25] R. Zafarani and H. Liu, "10 bits of surprise: Detecting malicious users with minimum information" in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, ser. CIKM '15. ACM, pp. 423–431.