Who is Tweeting on Twitter: Human, Bot, or Cyborg?

Zi Chu, Steven Gianvecchio and Haining Wang
Department of Computer Science
The College of William and Mary
Williamsburg, VA 23187, USA
{zichu, srgian, hnw}@cs.wm.edu

Sushil Jajodia
Center for Secure Information Systems
George Mason University
Fairfax, VA 22030, USA
jajodia@gmu.edu

ABSTRACT

Twitter is a new web application playing dual roles of online social networking and micro-blogging. Users communicate with each other by publishing text-based posts. The popularity and open structure of Twitter have attracted a large number of automated programs, known as bots, which appear to be a double-edged sword to Twitter. Legitimate bots generate a large amount of benign tweets delivering news and updating feeds, while malicious bots spread spam or malicious contents. More interestingly, in the middle between human and bot, there has emerged cyborg referred to either bot-assisted human or human-assisted bot. To assist human users in identifying who they are interacting with, this paper focuses on the classification of human, bot and cyborg accounts on Twitter. We first conduct a set of large-scale measurements with a collection of over 500,000 accounts. We observe the difference among human, bot and cyborg in terms of tweeting behavior, tweet content, and account properties. Based on the measurement results, we propose a classification system that includes the following four parts: (1) an entropy-based component, (2) a machine-learning-based component, (3) an account properties component, and (4) a decision maker. It uses the combination of features extracted from an unknown user to determine the likelihood of being a human, bot or cyborg. Our experimental evaluation demonstrates the efficacy of the classification system.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and Protection*

General Terms

Security

Keywords

Automatic Identification, Bot, Cyborg, Twitter

1. INTRODUCTION

Twitter is a popular online social networking and micro-blogging tool, which was released in 2006. Remarkable simplicity is its distinctive feature. Its community interacts via publishing text-based posts, known as *tweets*. The tweet size is limited to 140 characters. Hashtag, namely words or phrases prefixed with a # symbol, can group tweets by topic. For example, #Haiti and #Super Bowl are the two trending hashtags on Twitter in January 2010. Symbol @ followed by a username in a tweet enables the direct delivery of the tweet to that user. Unlike most online social networking sites (i.e., Facebook and MySpace), Twitter's user relationship is directed and consists of two ends, friend and follower. In the case where the user A adds B as a friend, A is a *follower* of B while B is a *friend* of A. In Twitter terms, A follows B. B can also add A as his friend (namely, following back or returning the follow), but is not required. From the standpoint of information flow, tweets flow from the source (author) to subscribers (followers). More specifically, when a user posts tweets, these tweets are displayed on both the author's homepage and those of his followers.

Since 2009, Twitter has gained increasing popularity. As reported in June 2010, Twitter is attracting 190 million visitors per month and generating 65 million Tweets per day [30]. It ranks the 12th on the top 500 site list according to Alexa [5]. In November 2009, Twitter emphasized its value as a news and information network by changing the question above the tweet input dialog box from "What are you doing" to "What's happening". To some extent, Twitter is in the transition from a personal micro-blogging site to an information publish venue. Many traditional industries have used Twitter as a new media channel. We have witnessed successful Twitter applications in business promotion [1], customer service [3], political campaigning [2], and emergency communication [21, 35].

The growing user population and open nature of Twitter have made itself an ideal target of exploitation from automated programs, known as bots. Like existing bots in other web applications (i.e., Internet chat [14], blogs [34] and online games [13]), bots have been common on Twitter. Twitter does not inspect strictly on automation. It only requires the recognition of a CAPTCHA image during registration. After gaining the login information, a bot can perform most human tasks by calling Twitter APIs. More interestingly, in the middle between humans and bots have emerged cyborgs, which refer to either bot-assisted humans or human-assisted bots. Cyborgs have become common on Twitter. After a human registers an account, he may set automated programs (i.e., RSS feed/blog widgets) to post tweets during his absence. From time to time, he participates to tweet and interact with friends. Cyborgs interweave characteristics of both humans and bots.

Automation is a double-edged sword to Twitter. On one hand, legitimate bots generate a large volume of benign tweets, like news and blog updates. This complies with the Twitter's goal of becoming a news and information network. On the other hand, malicious bots have been greatly exploited by spammers to spread spam or malicious contents. These bots randomly add users as their friends, expecting a few users to follow back[1]. In this way, spam tweets posted by bots display on users' homepages. Enticed by the appealing text content, some users may click on links and get redirected to spam or malicious sites[2]. If human users are surrounded by ma-

ACSAC '10 Dec. 6-10, 2010, Austin, Texas USA

[1]Some advanced bots target potential users by keyword search.

[2]Due to the tweet size limit, it is very common to use link shortening service on Twitter, which converts an original link to a short one (i.e., http://bit.ly/dtUm5Q). The link illegibility favors bots to

licious bots and spam tweets, their twittering experience deteriorates, and eventually the whole Twitter community will be hurt. The objective of this paper is to characterize the automation feature of Twitter accounts, and to classify them into three categories, human, bot, and cyborg, accordingly. This will help Twitter manage the community better and help human users recognize who they are tweeting with.

In the paper, we first conduct a series of measurements to characterize the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. By crawling Twitter, we collect over 500,000 users and more than 40 million tweets posted by them. Then we perform a detailed data analysis, and find a set of useful features to classify users into the three classes. Based on the measurement results, we propose an automated classification system that consists of four major components: (1) the entropy component uses tweeting interval as a measure of behavior complexity, and detects the periodic and regular timing that is an indicator of automation; (2) the machine-learning component uses tweet content to check whether text patterns contain spam or not[3]; (3) the account properties component employs useful account properties, such as tweeting device makeup, URL ration, to detect deviations from normal; (4) the decision maker is based on Linear Discriminant Analysis (LDA), and it uses the linear combination of the features generated by the above three components to categorize an unknown user as human, bot or cyborg. We validate the efficacy of the classification system through our test dataset. We further apply the system to classify the entire dataset of over 500,000 users collected, and speculate the current composition of Twitter user population based on our classification results.

The remainder of this paper is organized as follows. Section 2 covers related work on Twitter and online social networks. Section 3 details our measurements on Twitter. Section 4 describes our automatic classification system on Twitter. Section 5 presents our experimental results on classification of humans, bots, and cyborgs on Twitter. Finally, Section 6 concludes the paper.

2. RELATED WORK

Twitter has been widely used since 2006, and there are some related literature in twittering [24, 25, 42]. To better understand micro-blogging usage and communities, Java et al. [24] studied over 70,000 Twitter users and categorized their posts into four main groups—daily chatter (e.g., "going out for dinner"), conversations, sharing information or URLs, and reporting news—and further classified their roles by link structure into three main groups—information source, friends, and information seeker. Their work also studied (1) the growth of Twitter, showing a linear growth rate; (2) its network properties, showing the evidence that the network is scale-free like other social networks [27]; and (3) the geographical distribution of its users, showing that most Twitter users are from the US, Europe, and Japan. Krishnamurthy et al. [25] studied a group of over 100,000 Twitter users and classified their roles by follower-to-following ratios into three groups: (1) broadcasters, which have a large number of followers; (2) acquaintances, which have about the same number on either followers or following; and (3) miscreants and evangelists (e.g., spammers), which follow a large number of other users but have few followers. Their work also examined the growth of Twitter, revealing a greater than linear growth rate. In a more recent work, Yardi et al. [42] investigated spam on Twitter. According to their observations, spammers send more messages than legitimate users, and are more likely to follow other spammers than legitimate users. Thus, a high follower-to-following ratio is a sign of spamming behavior. Kim et al. [10] analyzed Twitter lists as a potential source for discovering latent characters and interests of users. A Twitter list consists of multiple users and their tweets. Their research indicated that words extracted from each list are representative of all the members in the list even if the words are not used by the members. It is useful for targeting users with specific interests.

Compared to previous measurement studies on Twitter, our work covers a much larger group of Twitter users (more than 500,000) and differs in how we link the measurements to automation, i.e., whether posts are from humans, bots, or cyborgs. While some similar metrics are used in our work, such as follower-to-following ratio, we also introduce some metrics, including entropy of tweet intervals, which are not employed in previous research. In addition to network-related studies, several previous works focus on socio-technological aspects of Twitter [21, 23, 32, 35, 44], such as its use in the workplace or during major disaster events.

Twitter is a social networking service, so our work is also related to recent studies on social networks, such as Flickr, LiveJournal, Facebook, MySpace, and YouTube [6, 7, 27]. In [27], with over 11 million users of Flickr, YouTube, LiveJournal, and Orkut, Mislove et al. analyzed link structure and uncovered the evidence of power-law, small-world, and scale-free properties. In [7], Cha et al. examined the propagation of information through the social network of Flickr. Their work shows that most pictures are propagated through the social links (i.e., links received from friends rather than through searches or external links to Flickr content) and the propagation is very slow at each hop. As a result of this slow propagation, a picture's popularity is often localized in one network and grows slowly over a period of months or even years. In [6], Cha et al. analyzed video popularity life-cycles, content aliasing, and the amount of illegal content on YouTube, a popular video sharing service. While YouTube is designed to share large content, i.e., videos, Twitter is designed to share small content, i.e., text messages. Unlike other social networking services, like Facebook or YouTube, Twitter is a micro-content social network, with messages being limited to 140 characters.

As Twitter is a text-based message system, it is natural to compare it with other text-based message systems, such as instant messaging or chat services. Twitter has similar message length (140 characters) to instant messaging and chat services. However, Twitter lacks "presence" (users show up as online/offline for instant messaging services or in specific rooms for chat) but offers (1) more access methods (web, SMS, and various APIs) for reading or posting and (2) more persistent content. Similar to Twitter, instant messaging and chat services also have problems with bots and spam [**?**, 14]. To detect bots in online chat, Gianvecchio et al. [14] analyzed humans and bots in Yahoo! chat and developed a classification system to detect bots using entropy-based and machine-learning-based classifiers, both of which are used in our classification system as well. In addition, as Twitter is text-based, email spam filtering techniques are also relevant [17, 40, 43]. However, Twitter posts are much shorter than emails and spaced out over longer periods of time than for instant messages, e.g., hours rather than minutes or seconds.

Twitter also differs from most other network services in that automation, e.g., message feeds, is a major feature of legitimate Twitter usage, blurring the lines between bot and human. Twitter users can be grouped into four categories: humans, bots, bot-assisted humans, and human-assisted bots. The latter two, bot-assisted humans and human-assisted bots, can be described as cyborgs, a mix between bots and humans [41].

3. MEASUREMENT

In this section, we first describe the data collection of over 500,000 Twitter users. Then, we detail our observation of user behaviors and account properties, which are pivotal to automatic classification.

3.1 Data Collection

Here we present the methodology used to crawl the Twitter network and collect detailed user information. Twitter has released a set of API functions [39] that support user information collection. Thanks to Twitter's courtesy of including our test account to its white list, we can make API calls up to 20,000 per hour. This eases our data collection. To diversify our data sampling, we employ two methods to collect the dataset covering more than 500,000 users. The first method is Depth-First Search (DFS) based crawling. The reason we choose DFS is that it is a fast and uniformed algorithm for traversing a network. Besides, DFS traversal implicitly includes the information about network locality and clustering. Inspired by [15, 18], we randomly select five users as seeds. For

allure users.

[3]Spam is a good indicator of automation. Most spam messages are generated by bots, and very few are manually posted by humans.

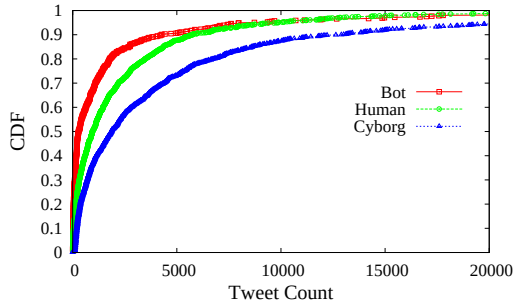


Figure 1: CDF of Tweet Count

each reached user, we record its follower list. Taking the following direction, the crawler continues with the depth constraint set as three. We customize our crawler with a core module of PHP cURL. Ten crawler processes work simultaneously for each seed. After a seed is finished, they move to the next. The crawl duration lasts four weeks from October 20th to November 21st, 2009, and 429,423 users are logged.

Similar to the work in [25] and [42], we also use the public timeline API to collect the information of active users, increasing the diversity of the user pool. Twitter constantly posts the twenty most recent tweets in the global scope. The crawler calls the timeline API to collect the authors of the tweets included in the timeline. Since the Twitter timeline frequently updates, the crawler can repeatedly call the timeline API. During the same time window of the DFS crawl, this method contributes 82,984 users to the dataset. We totally collect 512,407 users on Twitter combining both methods.

3.2 Ground Truth Creation

To develop an automatic classification system, we need a training data set that contains known samples of human, bot, and cyborg. Among collected data, we randomly choose different samples and classify them by manually checking their user logs and homepages. The training set includes one thousand users per class of human, bot and cyborg, and thus in total there are three thousand classified samples. A test set of three thousand samples is created in a similar way. Both sets serve as the ground truth dataset, containing 8,350,095 tweets posted by the sampled users in their account lifetime[4], from which we can extract useful features for classification, such as tweeting behaviors and text patterns.

Our log-based classification follows the principle of the Turing test [36]. The standard Turing tester communicates with an unknown subject for five minutes, and decides whether it is a human or machine. Classifying Twitter users is actually more challenging than it appears to be. For many users, their tweets are less likely to form a relatively consistent context. For example, a series of successive tweets may be hardly relevant. The first tweet is the user status, like “watching a football game with my buds.” The second tweet is an automatic update from his blog. The third tweet is a news report RSS feed in the format of article title followed by a shortened URL.

For every account, the following classification procedure is executed. We thoroughly observe the log, and visit the user’s homepage (http://twitter.com/username) if necessary. We carefully check tweet contents, visit URLs included in tweets (if any), and decide if redirected web pages are related with their original tweets and if they contain spam or malicious contents. We also check other properties, like tweeting devices, user profile, and the numbers of followers and friends. Given a long sequence of tweets (usually we check 60 or more if needed), the user is labeled as a human if we can obtain some evidence of original, intelligent, specific and human-like contents. In particular, a human user usually records what he is doing or how he feels about something on Twitter, as he uses Twitter as a micro-blogging tool to display himself and interact with friends. For example, he may write a post like “I just saw Yankees lost again today. I think they have to replace the starting pitcher for tomorrow’s game." The content carries intelligence and originality. Specificity means that the tweet content is expressed in relatively unambiguous words with the presence of consciousness [36]. For instance, in reply to a tweet like “How you like iPad?", a specific response made by human may be “I like its large touch screen and embedded 3G network”. On the other hand, a generic reply could be “I like it".

The criteria for identifying a bot are listed as follows. The first is the lack of intelligent or original content. For example, completely retweeting tweets of others or posting adages indicates a lack of originality. The second is the excessive automation of tweeting, like automatic updates of blog entries or RSS feeds. The third is the abundant presence of spam or malicious URLs (i.e., phishing or malware) in tweets or the user profile. The fourth is repeatedly posting duplicate tweets. The fifth is posting links with unrelated tweets. For example, the topic of the redirected web page does not match the tweet description. The last is the aggressive following behavior. In order to gain attention from human users, bots do mass following and un-following within a short period of time. Cyborgs are either human-assisted bots or bot-assisted humans. The criterion for classifying a cyborg is the evidence of both human and bot participation. For example, a typical cyborg account may contain very different types of tweets. A large proportion of tweets carry contents of human-like intelligence and originality, while the rest are automatic updates of RSS feeds. It represents a usage model, in which the human uses his account from time to time while the Twitter widget constantly runs on his desktop and posts RSS feeds of his favorite news channel. Lastly, the uncertain category is for non-English users and those without enough tweets to classify. The samples that are difficult and uncertain to classify fall into this category, and are discarded. Some Twitter accounts are set as "private" for privacy protection, and their web pages are only visible to their friends. We do not include such type of users in the classification either, because of their inaccessibility.

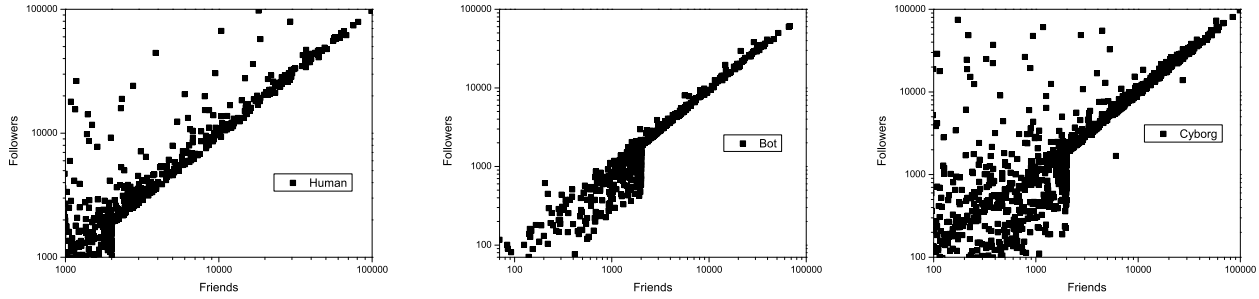
3.3 Data Analysis

As mentioned before, Twitter API functions support detailed user information query, ranging from profile, follower and friend lists to posted tweets. In the above crawl, for each user visited, we call API functions to collect abundant information related with user classification. Most information is returned in the format of XML or JSON. We develop some toolkits to extract useful information from the above well-organized data structures. Our measurement results are presented in the question-answer format.

Q1. Does automation generate more tweets? To answer Question 1, we measure the number of tweets posted in a user’s lifetime. Figure 1 shows the cumulative distribution function (CDF) of the tweet counts, corresponding to the human, bot and cyborg category. It is clear that cyborg posts more tweets than human and bot. A large proportion of cyborg accounts are registered by commercial companies and websites as a new type of media channel and customer service. Most tweets are posted by automated tools (i.e., RSS feed widgets, Web 2.0 integrators), and the volume of such tweets is considerable. Meanwhile, those accounts are usually maintained by some employees who communicate with customers from time to time. Thus, the high tweet count in the cyborg category is attributed to the combination of both automatic and human behaviors in a cyborg. It is surprising that bot generates fewer tweets than human. We check the bot accounts, and find out the following fact. In its active period, bot tweets more frequently than human. However, bots tend to take long-term hibernation. Some are either suspended by Twitter due to extreme or aggressive activities, while the others are in incubation and can be activated to form bot legions.

Q2. Do bots have more friends than followers? A user’s tweets can only be delivered to those who follow him. A common strategy shared by bots is following a large number of users (either targeted with purpose or randomly chosen), and expecting some of them will follow back. Figure 2 shows the scatter plots of the numbers of followers and friends for the three categories. For better illustration, the scale is chopped and a small amount of extraordinary points are not included. In Figure 2, there are three different groups of users: group I where the number of one’s followers is clearly

[4]4,431,923 tweets in the training set, and 3,918,172 tweets in the test set.



(a) Human (b) Bot (c) Cyborg

Figure 2: Numbers of Followers and Friends

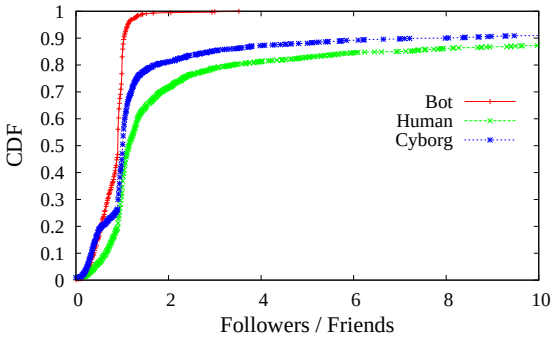


Figure 3: CDF of ratio of Followers over Friends

greater than the number of its friends; group II where the situation is reverse; and group III where the nodes stick around the diagonal.

In the human category, as shown in Figure 2(a), the majority of the nodes belong to group III, implying that the number of their followers is close to that of their friends. This result complies with [27], revealing that human relationships are typically reciprocal in social networks. Meanwhile, there are quite a few nodes belonging to group I with far more followers than friends. They are usually accounts of celebrities and famous organizations. They generate interesting media contents and attract numerous subscribers. For example, the singer Justin Timberlake has 1,645,675 followers and 39 friends (the ratio is 42,197-to-1).

In the bot category, many nodes belong to group II, as shown in Figure 2(b). Bots add many users as friends, but few follow them back. Unsolicited tweets make bots unpopular among the human world. However, for some bots, the number of their followers is close to that of their friends. This is due to the following reasons. First, Twitter imposes a limit on the ratio of followers over friends to suppress bots. Thus, some advanced bots un-follow their friends if they do not follow back within a certain amount of time. Those bots cunningly keep the ratio close to 1. Figure 3 shows the ratio of followers over friends for the three categories. The human ratio is the highest, and the bot ratio is the lowest.

Q3. Are there other temporal properties of Twitter users useful for differentiation among human, bot, and cyborg? Many research works like [11] and [9] have shown the weekly and diurnal access patterns of humans in the Internet. Figure 4(a) and Figure 4(b) present the tweeting percentage of the three user categories on a daily and hourly base, respectively. The weekly behavior of Twitter users shows clear differences among the three categories. While humans are most active during the regular workdays, from Monday to Friday, and less active during the weekend, Saturday and Sunday, bots have roughly the same activity level every day of the week. Interestingly, cyborgs are the most active on Monday and then slowly decrease during the week and become the least active on Saturday and Sunday. The cyborg activity trends are mainly due to their message feeds and high levels of news and blog activity at the start of a week. Similarly, the hourly behavior of human is more active during the daytime that mostly overlaps with office hours. The bot activity is nearly even except a little drop in the deep of night. Some advanced bots have the setting of "only tweet from a time point to another," which helps save API calls [37]. Thus, they can tweet more in the daytime to better draw the attention of humans.

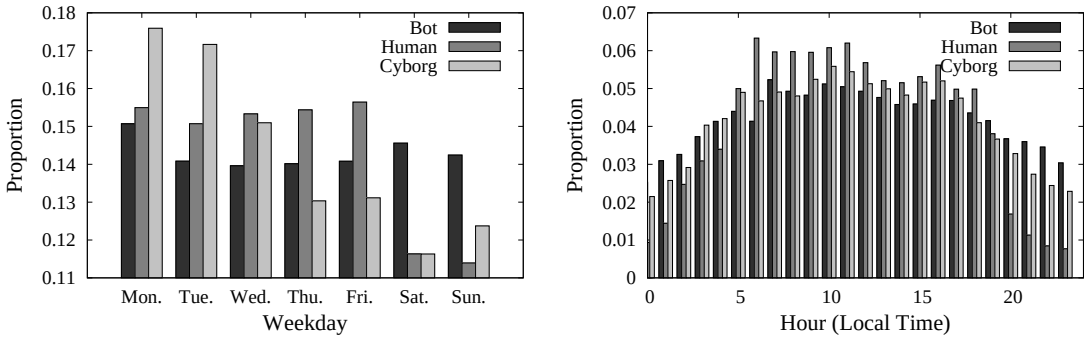
Figure 5 shows account registration dates grouped by quarter due to space limit. We draw two conclusions from the figure. First, the majority of accounts (80.0% of humans, 94.8% of bots, and 71.1% of cyborgs) were registered in 2009. It confirms the skyrocketing development of Twitter in 2009. Second, we do not find any bot or cyborg in our ground truth dataset earlier than March, 2007. However, human registration has continued increasing since Twitter was founded in 2006. Thus, old accounts are less likely to be bots.

Q4. How do users post tweets? Manually or via auto piloted tools? Twitter supports a variety of channels to post tweets. The device name appears below the tweet it posts prefixed by "from." Our whole dataset includes 41,991,545 tweets posted by 3,648 distinct devices. The devices can be roughly divided into the following four categories. (1) Web, the user logs into Twitter and posts tweets via the website. (2) Mobile devices, there are some programs exclusively running on mobile devices to post tweets, like Txt for text messages, Mobile web for web browsers on handheld devices, TwitterBerry for BlackBerry, and twidroid for Android mobile OS. (3) Registered third-party applications, many third-parties have developed their own applications using Twitter APIs to tweet, and registered them with Twitter. From the application standpoint, we can further categorize this group into sub groups including website integrators (twitpic, bit.ly, Facebook), browser extensions (Tweetbar and Twitterfox for Firefox), desktop clients (TweetDeck and Seesmic Desktop), and RSS feeds/blog widgets (twitterfeed and Twitter for Wordpress). (4) APIs, for those third-party applications not registered or certificated by Twitter, they appear as "API" in Twitter.

Figure 6 shows the makeup of the above tweeting device categories. Twitter website is the most widely used that generates nearly half the tweets (46.78%), followed by third-party devices (40.18%). Mobile devices and unregistered API tools contribute 6.81% and 6.23%, respectively. Table 1 lists the top ten devices used by the human, bot, and cyborg categories, and the whole dataset[5].

More than half of the human tweets are manually posted via Twitter website. The rest of top devices are mobile applications (Tweetie, UberTwitter, Mobile web, Txt, TwitterBerry) and desktop clients (TweetDeck, Echofon and Seesmic). In general, tweeting via such devices requires human participation. In contrast, the top tools used by bots are mainly auto piloted, and 42.39% of bot tweets are generated via unregistered API-based tools. Bots can abuse APIs to do almost everything they want on Twitter, like targeting users with keywords, following users, unfollowing those who do not follow back, or posting prepared tweets. Twitterfeed,

[5]The whole dataset contains around 500,000 users, and the human, bot and cyborg categories equally contain 1,000 users in the training dataset.



(a) Tweets by Day of Week (b) Hourly Tweets

Figure 4: Tweets Posted

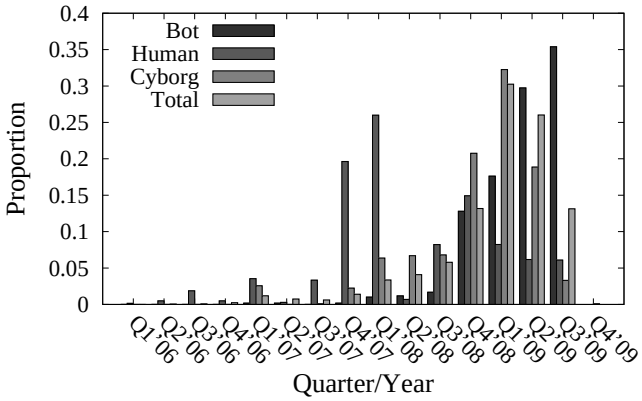


Figure 5: Account Registration Date (Grouped by Quarter)

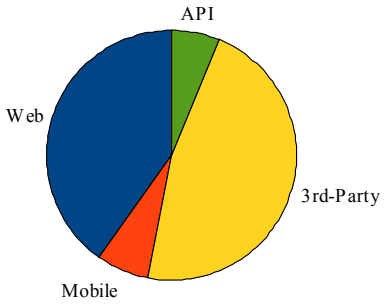


Figure 6: Tweeting Device Makeup

RSS2Twitter, and Proxifeed are RSS feed widgets that automatically pipeline information (usually in the format of the page title followed by the URL) to Twitter via RSS feeds. Twitter Tools and Twitme for WordPress are popular WordPress plug-ins that integrate blog updates to Twitter. Assetize is an advertising syndicator mainly targeting at Twitter, and twitRobot is a bot tool that automatically follows other users and posts tweets. All these tools only require minimum human participation (like importing Twitter account information, or setting RSS feeds and update frequency), and thus indicate great automation.

Overall, humans tend to tweet manually and bots are more likely to use auto piloted tools. Cyborgs employ the typical human and bot tools. The cyborg group includes many human users who access their Twitter accounts from time to time. For most of the time when they are absent, they leave their accounts to auto piloted tools for management.

Q5. Do bots include more external URLs than humans? In our measurement, we find out that, most bots tend to include URLs in tweets to redirect visitors to external web pages. For example, spam bots are created to spread unsolicited commercial information. Their topics are similar to those in email spam, including online marketing and affiliate programs, working at home, selling fake luxury brands or pharmaceutical products[6]. However, the tweet size is up to 140 characters, which is rather limited for spammers to express enough text information to allure users. Basically, a spam tweet contains an appealing title followed by an external URL. Figure 7 shows the external URL ratios (namely, the number of external URLs included in tweets over the number of tweets posted by an account) for the three categories. The URL ratio of

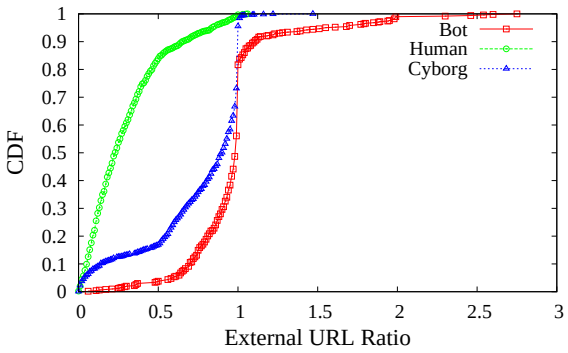


Figure 7: External URL ratio in tweets

bot is highest. Some tweets by bot even have more than one URL[7]. The URL ratio of cyborg is very close to the bot level. A large number of cyborgs integrate RSS feeds and blog updates, which take the style of webpage titles followed by page links. The URL ratio of human is much lower, on average it is only 29%. When a human tweets what is he doing or what is happening around him, he mainly uses text and does not link pages often.

Q6. Are users aware of privacy and identity protection on Twitter? Twitter provides a `protected` option to protect user privacy. If it is set as true, the user's homepage is only visible to his friends. However, the option is set as false by default. In our dataset of over 500,000 users, only 4.9% of them are protected users. Twitter also verifies some accounts to authenticate users' real identities.

[6]A new topic is getting more followers on Twitter. It follows the style of pyramid sales by asking newly joined users to follow existing users in the spam network.

[7]Many such accounts belong to a type of bots that always append a spam link to tweets it re-tweets.

Table 1: Top 10 Tweeting Devices

Rank	Human	Bot	Cyborg	All
#1	Web (50.53%)	API (42.39%)	Twitterfeed (31.29%)	Web (46.78%)
#2	TweetDeck (9.19%)	Twitterfeed (26.11%)	Web (23.00%)	TweetDeck (9.26%)
#3	Tweetie (6.23%)	twitRobot (13.11%)	API (6.94%)	Twitterfeed (7.83%)
#4	UberTwitter (3.64%)	RSS2Twitter (2.66%)	Assetize (5.74%)	API (6.23%)
#5	Mobile web (3.02%)	Twitter Tools (1.24%)	HootSuite (5.22%)	Echofon (2.80%)
#6	Txt (2.56%)	Assetize (1.17%)	WP to Twitter (2.40%)	Tweetie (2.50%)
#7	Echofon (2.22%)	Proxifeed (1.08%)	TweetDeck (1.54%)	Txt (2.13%)
#8	TwitterBerry (2.10%)	TweetDeck (0.99%)	UberTwitter (1.19%)	HootSuite (2.10%)
#9	Twitterrific (1.93%)	bit.ly (0.91%)	RSS2Twitter (1.18%)	UberTwitter (1.71%)
#10	Seesmic(1.64%)	Twitme for WordPress (0.84%)	Twaitter (0.86%)	Mobile web (1.53%)

More and more celebrities and famous organizations have applied for verified accounts. For example, Bill Gates has his verified Twitter account at `http://twitter.com/billgates`. However, in our dataset, only 1.8% of users have verified accounts.

4. CLASSIFICATION

This section describes our automated system for classification of Twitter users. The system classifies Twitter users into three categories: human, bot, and cyborg. The system is made up of several components: the entropy component, the machine learning component, the account properties component, and the decision maker. The high-level design of our Twitter user classification system is shown in Figure 8. The components of the classification system determine what features are present for each user. Based on the combination of these features, the decision maker decides whether the user is a human, bot, or cyborg. The entropy component uses corrected conditional entropy to detect periodic or regular timing, which is a sign of automation. The machine learning component uses a variant of Bayesian classification to detect text patterns of known spam on Twitter. The account properties component uses account-related properties to catch bot deviation from the normal human distribution. Lastly, the decision maker uses LDA to analyze the features identified by the other three components and makes a decision: human, cyborg, or bot.

4.1 Entropy Component

The entropy component detects periodic or regular timing of the messages posted by a Twitter user. On one hand, if the entropy or corrected conditional entropy is low for the inter-tweet delays, it indicates periodic or regular behavior, a sign of automation. More specifically, some of the messages are posted via automation, i.e., the user may be a potential bot or cyborg. On the other hand, a high entropy indicates irregularity, a sign of human participation.

4.1.1 Entropy Measures

The entropy rate is a measure of the complexity of a process [8]. The behavior of bots is often less complex than that of humans [12,22], which can be measured by entropy rate. A low entropy rate indicates a regular process, whereas a high entropy rate indicates a random process. A medium entropy rate indicates a complex process, i.e., a mix of order and disorder [20].

The entropy rate is defined as either the average entropy per random variable for an infinite sequence or as the conditional entropy of an infinite sequence. Thus, as real datasets are finite, the conditional entropy of finite sequences is often used to estimate the entropy rate. To estimate the entropy rate, we use the corrected conditional entropy [28]. The corrected conditional entropy is defined as follows.

A random process $X = \{X_i\}$ is defined as a sequence of random variables. The entropy of such a sequence of random variables is defined as:

$$H(X_1, ..., X_m) = -\sum_{X_1,...,X_m} P(x_1, ..., x_m) \log P(x_1, ..., x_m), \tag{1}$$

where $P(x_1, ..., x_m)$ is the joint probability $P(X_1 = x_1, ..., X_m = x_m)$.

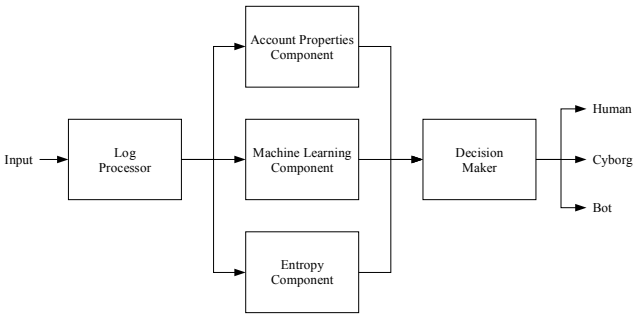


Figure 8: Classification System

The conditional entropy of a random variable given a previous sequence of random variables is:

$$H(X_m \mid X_1, ..., X_{m-1}) = H(X_1, ..., X_m) - H(X_1, ..., X_{m-1}). \tag{2}$$

Then, based on the conditional entropy, the entropy rate of a random process is defined as:

$$\overline{H}(X) = \lim_{m \to \infty} H(X_m \mid X_1, ..., X_{m-1}). \tag{3}$$

The corrected conditional entropy is computed as a modification of Equation 3. First, the joint probabilities, $P(X_1 = x_1, ..., X_m = x_m)$ are replaced with empirically-derived probabilities. The data is binned into Q bins, i.e., values are converted to bin numbers from 1 to Q. The empirically-derived probabilities are then determined by the proportions of bin number sequences in the data. The entropy estimate and conditional entropy estimate, based on empirically-derived probabilities, are denoted as EN and CE respectively. Second, a corrective term, $perc(X_m) \cdot EN(X_1)$, is added to adjust for the limited number of sequences for increasing values of m [28]. The corrected conditional entropy, denoted as CCE, is computed as:

$$\begin{aligned} &CCE(X_m \mid X_1, ..., X_{m-1}) = \\ &\quad CE(X_m \mid X_1, ..., X_{m-1}) + perc(X_m) \cdot EN(X_1), \end{aligned} \tag{4}$$

where $perc(X_m)$ is the percentage of unique sequences of length m and $EN(X_1)$ is the entropy with m fixed at 1 or the first-order entropy.

The estimate of the entropy rate is the minimum of the corrected conditional entropy over different values of m. The minimum of the corrected conditional entropy is considered to be the best estimate of the entropy rate from the limited number of sequences.

4.2 Machine Learning Component

The machine learning component uses the content of tweets to detect spam. We have observed that most spam tweets are generated by bots and only very few of them are manually posted by humans. Thus, the presence of spam patterns usually indicates automation. Since tweets are text, determining if their content is spam

can be reduced to a text classification problem. The text classification problem is formalized as $f : T \times C \rightarrow \{0, 1\}$, where f is the classifier, $T = \{t_1, t_2, ..., t_n\}$ are the texts to be classified, and $C = \{c_1, c_2, ..., c_k\}$ are the classes [31]. A value of 1 for $f(t_i, c_j)$ indicates that text t_i belongs to class c_j, whereas a value of 0 indicates it does not belong to that class. Bayesian classifiers are very effective in text classification, especially for email spam detection, so we decide to employ Bayesian classification for our machine learning text classification component.

In Bayesian classification, deciding if a message belongs to a class, e.g., spam, is accomplished by computing the probability that the document is from that class based on it content, e.g., $P(C = spam|M)$, where M is a message and C is a class. If the probability is over a certain threshold, then the document is from that class.

The probability that a message M is spam, $P(spam|M)$, is computed from Bayes theorem:

$$P(spam|M) = \frac{P(M|spam)P(spam)}{P(M)} = \frac{P(M|spam)P(spam)}{P(M|spam)P(bot) + P(M|not\ spam)P(not\ spam)}. \quad (5)$$

The message M is represented as a feature vector $\langle f_1, f_2, ..., f_n \rangle$, where each feature f is one or more words in the message and each feature is assumed to be conditionally independent.

$$P(spam|M) = \frac{P(spam)\prod_{i=1}^{n} P(f_i|spam)}{P(spam)\prod_{i=1}^{n} P(f_i|spam) + P(not\ spam)\prod_{i=1}^{n} P(f_i|not\ spam)}. \quad (6)$$

The calculation of $P(spam|M)$ varies in different implementations of Bayesian classification. The implementation used for the machine learning component is CRM114 [4]. CRM114 is a powerful text classification system that offers a variety of different classifiers. The default classifier for CRM114 is Orthogonal Sparse Bigram (OSB), a variant of Bayesian classification, which has been shown to perform well for email spam filtering. OSB differs from other Bayesian classifiers in that it treats pairs of words as features.

4.3 Account Properties Component

Besides inter-tweet delay and tweet content, there are some Twitter account-related properties useful for the user classification. As shown in Section 3.3, obvious difference exists between the human and bot categories. The first property is the URL ratio. The ratio indicates how often a user includes external URLs in its posted tweets. External URLs appear very often in tweets posted by a bot. Our measure shows, on average the ratio of bot is 97%, while that of human is much lower at 29%. Thus, a high ratio (e.g., close to one) suggests bot and a low ratio implies human.

The second property is tweeting device makeup. According to Table 1, about 70% tweets of human are posted by web and mobile devices (referred as manual devices), whereas about 87% tweets of bot are posted by API and other auto-piloted programs (referred as auto devices). The third property is the followers to friends ratio. Figure 3 clearly shows the difference between human and bot. The fourth property is link safety, i.e., to decide whether external links in tweets are malicious/phishing URLs or not. We use Google's Safe Browsing (GSB) API project [16], which allows us to check URLs against Google's constantly-updated blacklists of suspected phishing and malware pages. The component converts each URL[8] into hash values based on Google's rules, and performs the local lookup from downloaded Google's blacklists. Appearance in Google's blacklists raises the red flag for security breach. GSB is also applied by Twitter for the link safety inspection [38]. The fifth property is whether a Twitter account is verified. No bot in our ground truth dataset is verified. The account verification suggests human. The last property is the account registration date. According to Figure 5, 94.8% of bots were registered in 2009.

The account properties component extract these properties from the user log, and sends them to the decision maker. It assists the entropy component and the machine-learning component to improve the classification accuracy.

[8]For a shortened URL, the component uses PHP cURL to get the original one from the redirected HTTP response header instead of actually visiting the page.

4.4 Decision Maker

Given an unknown user, the decision maker uses features identified by the above three components to determine whether it is a human, bot, or cyborg. It is built on Linear Discriminant Analysis (LDA) [26]. LDA is a statistical method to determine a linear combination of features that discriminate among multiple classes of samples. More specifically, its underlying idea is to determine whether classes differ in light of the means of a feature (or features), and then to use that feature (or features) to identify class. It is very similar to analysis of variance (ANOVA) [29] and (logistic) regression analysis [19]. However, a big difference is that LDA has a fundamental assumption that independent variables are normally distributed. In other words, it is assumed that variables represent a sample from a multivariate normal distribution. Our classification involves three classes, human, bot and cyborg. Thus, it is a case of multiclass LDA. Multiclass LDA has the following key steps. First, it needs a training set and a test set that contain those samples already classified as one of the *C* classes. Samples in the two sets should not overlap with each other. Second, a discriminant model is created to use effective features to identify classes. Choosing features and assigning weights to features are the two important tasks in the model creation. In the early data collection stage, one usually includes several features to see which one(s) contributes to the discrimination. Some features are of very limited value for discrimination, and should be removed from the model. Our model uses *forward stepwise analysis*. In this way, the model is built step-by-step. At each step, all the features are evaluated, and the one that contributes most to the discrimination is added into the model. The selection process continues to next step. Suppose m features, $< v_1, v_2, ..., v_m >$ are selected. Each class C_i has a classification function. With those functions, we can compute the classification score of an unknown sample for each class, by using the following linear equation:

$$S_i = w_{i0} + \sum_{i=1}^{n} w_{i1}v_1 + w_{i2} * v_2 + ... + w_{im} * v_m \quad (7)$$

where *i* denotes the respective class, S_i denotes the classification score of the sample for class C_i, w_{i0} denotes a constant for class C_i, and w_{ij} denotes the weight of *j*-th feature in class C_i.

The sample is identified to belong to the class with the highest classification score. The model uses the training set to decide feature weights. Every sample in the training set is already known for the actual class it belongs to. The model keeps adjusting weights till it reaches the maximum accuracy for the training set. Third, the test set is used to validate the classification accuracy of the model. Since discriminant functions are derived from the training set, it is inappropriate to reuse it for the validation. It is easier for post hoc identification because it identifies what we already know has happened. The test set contains new data different from the training set, and generates more accurate validation results.

5. EVALUATION

In this section, we first evaluate the accuracy of our classification system based on the ground truth set (both the training and test datasets). Then, we apply the system to classify the entire dataset of over 500,000 users collected, and use the classification results to speculate the current composition of Twitter user population. Finally, we discuss the robustness of the proposed classification system against possible evasions.

5.1 Methodology

Table 2: Multi-class LDA Weights

	Human	Cyborg	Bot
Constant	-25.9879	-15.7787	-17.2416
Entropy	14.2524	9.7128	4.4136
Bayesian text	-0.0018	0.0164	0.1366
URL ratio	-3.4474	3.3059	8.5222
Manual device %	16.4601	13.0164	13.0950
Auto device %	8.5910	7.6849	18.3765
Followers to friends ratio	0.0007	0.0002	0.0003

As shown in Figure 8, the components of the classification system collaborate in the following way. The entropy component calculates the entropy (and corrected conditional entropy) of inter-tweet delays of a Twitter user. The component only processes logs with more than 100 tweets[9]. This limit helps reduce noise in detecting automation. A lower entropy indicates periodic or regular timing of tweeting behavior, a sign of automation, whereas a higher entropy implies irregular behavior, a sign of human participation. The machine learning component determines if the tweet content is either spam or not, based on the text patterns it has learned. The content feature value is set to -1 for spam but 1 for non-spam. The account properties component checks all the properties mentioned in Section 4.3, and generates a real-number-type value for each property. Given a Twitter user, the above three components generate a set of features and input them into the decision maker. For each class, namely human, bot and cyborg, the decision maker calculates a classification score for the user, and identifies him as the class with the highest score. The training of the classification system and its accuracy are detailed in the following sections.

5.2 Classification System Training

The classification system needs to be trained before being used. In particular, the machine learning component and the decision maker require training. The machine learning component is trained on spam and non-spam datasets. The spam dataset consists of spam tweets and spam external URLs, which are detected during the creation of the ground truth set. Some advanced spam bots intentionally inject non-spam tweets (usually in the format of pure text without URLs, such as adages[10]) to confuse human users. Thus, we do not include such vague tweets without external URLs. The non-spam dataset consists of all human tweets and cyborg tweets without external URLs. Most human tweets do not carry spam. Cyborg tweets with links are hard to determine without checking linked web pages. They can be either spam or non-spam. Thus, we do not include this type of tweets in either dataset. Training the component with up-to-date spam text patterns on Twitter helps improve the accuracy.

The decision maker is trained to determine the weights of the different features for classification. We use Statistica, a statistical tool [33], to calculate the feature weights. More specifically, the datasheet of feature values and the actual class of users in the training set are inputted into the classifier. LDA generates a weight table (Table 2) to achieve the maximum accuracy. In other words, it includes as many users as possible whose classified class matches actual class. The weights are then used by the decision maker to classify users.

The larger the (standardized) weight, the larger is the unique contribution of the respective feature to the discrimination. Table 2 shows that, entropy, URL ratio, and manual/auto device percentage are the important features for the classifier. Only those shown to be statistically significant should be used for classification, and non-significant ones should be ignored. Thus, some features collected by the account properties component in Section 4.3, including followers to friends ratio, link safety, account verification and registration date, are excluded from the classifier.

Here we briefly explain why several features, such as followers to friends ratio, link safety, account verification, and registration date, are not as important in the actual discrimination as expected. Bots used to have more friends than followers [25], and the ratio is less than one in this situation. However, there have emerged some more sophisticated bots that un-follow their friends if they do not follow back within a certain amount of time. They cunningly keep the ratio close to one. This strategy makes the ratio feature less useful. Most spam bots spread spam links on Twitter, instead of phishing or malicious links which are the primary target of the link safety inspector. Only 0.2% users in the training set do not pass the link safety inspection. Thus, the link safety feature has little weight under LDA due to its statistical insignificance. Similarly, account verification has a very small weight, because it is also quite rare. Only 1.8% of users are verified. Lastly, account registration dates greatly overlap among bots, humans, and cyborgs, making this feature not useful for discrimination as well.

5.3 Classification System Accuracy

To validate the accuracy of our proposed classification system, we create a test set containing one thousand users of each class. It shares no samples with the training set. The confusion matrix listed in Table 3 shows the classification results on the test set.

The "Actual" rows in Table 3 denote the actual classes of the users, and the "Classified" columns denote the classes of the users as decided by the classification system. For example, 949 in the "Human" row and column means that 949 humans were classified (correctly) as humans, whereas 51 in the "Human" row and "Cyborg" column means that 51 humans were classified (incorrectly) as cyborgs. There is no misclassification between human and bot.

We examine logs of those users being classified by mistake, and analyze each category as follows.

- For the human category, 5.1% of human users are classified as cyborg by mistake. One reason we find out is that, the overall scores of some users are lowered by spam content penalty. The tweet size is up to be 140 characters. Some patterns and phrases are used by both human and bot, such as "I post my online marketing experience at my blog at http://bit.ly/xT6klM. Please ReTweet it." Another reason is that the tweeting interval distribution of some human users is slightly lower than the entropy means, and they are penalized for that.

- For the bot category, 6.3% of bots are wrongly categorized as cyborg. The main reason is that, most of them escape the spam penalty from the machine learning component. Some spam tweets use very obscure text content, like "you should check it out since it's really awesome. http://bit.ly/xT6klM". Without checking the spam link, the component cannot determine if the tweet is spam merely based on the text.

- For the cyborg category, 9.8% is mis-categorized as human, and 7.4% is mis-categorized as bot. The definition of cyborg as human-assisted bot or bot-assisted human, is ambiguous. A strict policy can categorize cyborg as bot, while a loose one may categorize it as human.

Overall, our classification system can accurately differentiate between human and bot. However, it is much more challenging for a classification system to distinguish cyborg from human or bot.

5.4 Twitter Composition

As the last part of this section, we use the classification system to automatically classify our whole dataset of over 500,000 users. We can speculate the current composition of Twitter user population based on the classification results. The system classifies the users as: 48.7% as human, 37.5% as cyborg and 13.8% as bot. We speculate the population proportion of human, cyborg and bot category roughly as 5:4:1 on Twitter.

[9]The inter-tweet span could be wild on Twitter. An account may be inactive for months, suddenly tweets at an intensive frequency for a short-term, and enters hibernation again. It generates noise to the entropy component. Thus, the entropy component does not process logs with less than 100 tweets. Besides, in practice it is nearly impossible to determine automation based on a limited number of tweets.

[10]A typical content pattern is listed as follows. Tweet 1, A friend in need is a friend in deed. Tweet 2, Danger is next neighbour to security. Tweet 3, Work home and make $3k per month. Check out how, http://tinyurl.com/bF234T. Tweet 4, Clothes make the man...

Table 3: Confusion Matrix

		Classified				
		Human	Cyborg	Bot	Total	True Pos %
Actual	Human	949	51	0	1000	94.90%
	Cyborg	98	828	74	1000	82.80%
	Bot	0	63	937	1000	93.70%

5.5 Resistance to Evasion

Now we discuss the resistance of the classification system to evasion attempts made by bots. Bots may deceive certain features, such as the followers to friends ratio as mentioned before. However, our system has two critical features that are very hard for bots to evade. The first feature is tweeting device makeup, which corresponds to the manual/auto device percentage in Table 2. Manual device refers to web and mobile devices, while auto device refers to API and other auto-piloted programs (see Section 4.3). Tweeting via web requires a user to login and manually post via the Twitter website in a browser. Posting via HTTP form is considered by Twitter as API. Furthermore, currently it is impractical or expensive to run a bot on a mobile device to frequently tweet. As long as Twitter can correctly identify different tweeting platforms, device makeup is an effective metric for bot detection. The second feature is URL ratio. Considering the limited tweet length that is up to 140 characters, most bots have to include a URL to redirect users to external sites. Thus, a high URL ratio is another effective metric for bot detection. Other features like timing entropy, bot could mimic human behaviors but at the cost of much reduced tweeting frequency. We will continue to explore new features emerging with the Twitter development for more effective bot detection in the future.

6. CONCLUSION

In this paper, we explore the problem of automation by bots and cyborgs on Twitter. As a popular web application, Twitter has become a unique platform for information sharing with a large user base. However, its popularity and very open nature have made Twitter a very tempting target for exploitation by automated programs, i.e., bots. The problem of bots on Twitter is further complicated by the key role that automation plays in everyday Twitter usage.

To better understand the role of automation on Twitter, we measure and characterize humans, bots, and cyborgs on Twitter. By crawling Twitter, we collected one-month of data with over 500,000 Twitter users with more than 40 million tweets. Based on the data, we identified features that can differentiate humans, bots, and cyborgs on Twitter. Using entropy measures, we determined that humans have complex timing behavior, i.e., high entropy, whereas bots and cyborgs are often given away by their regular or periodic timing, i.e., low entropy. In examining the text of tweets, we observed that a high proportion of bot tweets contain spam content. Lastly, we discovered that certain account properties, like external URL ratio and tweeting device makeup, are also helpful on detecting automation.

Based on our measurements and characterization, we designed an automated classification system that consists of four main parts: entropy component, machine learning component, account properties component and decision maker. The entropy component checks for periodic or regular tweet timing patterns; the machine learning component checks for spam content; and the account properties component checks for abnormal values of Twitter-account-related properties. The decision maker summarizes the identified features and decides whether the user is a human, bot, or cyborg. The efficacy of the classification system is evaluated through the test dataset. Moreover, we apply the system to classify the entire dataset of over 500,000 users collected, and speculate the current composition of Twitter user population based on the classification results.

7. REFERENCES

[1] Amazon comes to twitter. `http://www.readwriteweb.com/archives/amazon_comes_to_twitter.php` [Accessed: Dec. 20, 2009].

[2] Barack obama uses twitter in 2008 presidential campaign. `http://twitter.com/BarackObama/` [Accessed: Dec. 20, 2009].

[3] Best buy goes all twitter crazy with @twelpforce. `http://twitter.com/in_social_media/status/2756927865` [Accessed: Dec. 20, 2009].

[4] The crm114 discriminator. `http://crm114.sourceforge.net/` [Accessed: Sept. 12, 2009].

[5] Alexa. The top 500 sites on the web by alexa. `http://www.alexa.com/topsites` [Accessed: Jan. 15, 2010].

[6] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego, CA, USA, 2007.

[7] Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th International Conference on World Wide Web*, Madrid, Spain, 2009.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 2006.

[9] Marcel Dischinger, Andreas Haeberlen, Krishna P. Gummadi, and Stefan Saroiu. Characterizing residential broadband networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet Measurement*, San Diego, CA, USA, 2007.

[10] Il-Chul Moon Dongwoo Kim, Yohan Jo and Alice Oh. Analysis of twitter lists as a potential source for discovering latent characteristics of users. In *To appear on CHI 2010 Workshop on Microblogging: What and How Can We Learn From It?*, 2010.

[11] Henry J. Fowler and Will E. Leland. Local area network traffic characteristics, with implications for broadband network congestion management. *IEEE Journal of Selected Areas in Communications*, 9(7), 1991.

[12] Steven Gianvecchio and Haining Wang. Detecting covert timing channels: An entropy-based approach. In *Proceedings of the 2007 ACM Conference on Computer and Communications Security*, Alexandria, VA, USA, October-November 2007.

[13] Steven Gianvecchio, Zhenyu Wu, Mengjun Xie, and Haining Wang. Battle of botcraft: fighting bots in online games with human observational proofs. In *Proceedings of the 16th ACM conference on Computer and Communications Security*, Chicago, IL, USA, 2009.

[14] Steven Gianvecchio, Mengjun Xie, Zhenyu Wu, and Haining Wang. Measurement and classification of humans and bots in internet chat. In *Proceedings of the 17th USENIX Security symposium*, San Jose, CA, 2008.

[15] Minas Gjoka, Maciej Kurant, Carter T Butts, and Athina Markopoulou. Walking in facebook: A case study of unbiased sampling of osns. In *Proceedings of the 27th IEEE International Conference on Computer Communications*, San Diego, CA, USA, March 2010.

[16] Google. Google safe browsing API. `http://code.google.com/apis/safebrowsing/` [Accessed: Feb. 5, 2010].

[17] Paul Graham. A plan for spam, 2002. `http://www.paulgraham.com/spam.html` [Accessed: Jan. 25, 2008].

[18] Monika R. Henzinger, Allan Heydon, Michael Mitzenmacher, and Marc Najork. On near-uniform url sampling. In *Proceedings of the 9th International World Wide Web Conference on Computer Networks*, Amsterdam, The Netherlands, May 2000.

[19] Christopher M. Hill and Linda C. Malone. Using simulated data in support of research on regression analysis. In *WSC '04: Proceedings of the 36th conference on Winter simulation*, 2004.

[20] B A Huberman and T Hogg. Complexity and adaptation. *Phys. D*, 2(1-3), 1986.

[21] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. In *Proceedings of the 6th International ISCRAM Conference*, Gothenburg, Sweden, May 2009.

[22] H. Husna, S. Phithakkitnukoon, and R. Dantu. Traffic shaping of spam botnets. In *Proceedings of the 5th IEEE Conference on Consumer Communications and Networking*, Las Vegas, NV, USA, January 2008.

[23] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *American Society for Information Science and Technology*, 60(11), 2009.

[24] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, CA, USA, 2007.

[25] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *Proceedings of the First Workshop on Online Social Networks*, Seattle, WA, USA, 2008.

[26] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience, 2004.

[27] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego, CA, USA, 2007.

[28] A Porta, G Baselli, D Liberati, N Montano, C Cogliati, T Gnecchi-Ruscone, A Malliani, and S Cerutti. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. *Biological Cybernetics*, Vol. 78(No. 1), January 1998.

[29] P. Real. A generalized analysis of variance program utilizing binary logic. In *ACM '59: Preprints of papers presented at the 14th national meeting of the Association for Computing Machinery*, New York, NY, USA, 1959.

[30] Erick Schonfeld. Costolo: Twitter now has 190 million users tweeting 65 million times a day. `http://techcrunch.com/2010/06/08/twitter-190-million-users/` [Accessed: Sept. 26, 2010].

[31] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, Vol. 34(No. 1), 2002.

[32] Kate Starbird, Leysia Palen, Amanda Hughes, and Sarah Vieweg. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proceedings of the ACM 2010 Conference on Computer Supported Cooperative Work*, February 2010.

[33] Statsoft. Statistica, a statistics and analytics software package developed by statsoft. `http://www.statsoft.com/support/download/brochures/` [Accessed: Mar. 12, 2010].

[34] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference on Computer and Communications Security*, Chicago, IL, USA, 2009.

[35] J. Sutton, Leysia Palen, and Irina Shlovski. Back-channels on the front lines: Emerging use of social media in the 2007 southern california wildfires. In *Proceedings of the 2008 ISCRAM Conference*, Washington, DC, USA, May 2008.

[36] Alan M. Turing. Computing machinery and intelligence. *Mind*, Vol. 59:433–460, 1950.

[37] Tweetadder. Automatic twitter software. `http://www.tweetadder.com/` [Accessed: Feb. 5, 2010].

[38] Twitter. How to report spam on twitter. `http://help.twitter.com/entries/64986` [Accessed: May. 30, 2010].

[39] Twitter. Twitter api wiki. `http://apiwiki.twitter.com/` [Accessed: Feb. 5, 2010].

[40] Mengjun Xie, Heng Yin, and Haining Wang. An effective defense against email spam laundering. In *Proceedings of the 13th ACM conference on Computer and Communications Security*, Alexandria, VA, USA, 2006.

[41] Jeff Yan. Bot, cyborg and automated turing test. In *Proceedings of the 14th International Workshop on Security Protocols*, Cambridge, UK, March 2006.

[42] Sarita Yardi, Daniel Romero, Grant Schoenebeck, and Danah Boyd. Detecting spam in a twitter network. *First Monday*, 15(1), January 2010.

[43] Jonathan A. Zdziarski. *Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification*. No Starch Press, 2005.

[44] Dejin Zhao and Mary Beth Rosson. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, Sanibel Island, FL, USA, 2009.