RT @IWantPrivacy: Widespread Violation of Privacy Settings in the Twitter Social Network

Brendan Meeder, Jennifer Tam, Patrick Gage Kelley, and Lorrie Faith Cranor
School of Computer Science, Carnegie Mellon University
Pittsburgh, PA USA
{*bmeeder, jdtam, pkelley, lorrie*}*@cs.cmu.edu*

***Abstract*—**
Twitter is a social network that focuses on creating and sharing short 140 character messages know as *tweets*. Twitter's sole privacy policy is a binary option that either allows every message a user creates to be publicly available, or allows only a user's *followers* to see posted messages. As the Twitter community grew, conventions organically formed that allow for rich expressiveness with only 140 characters. Repeating what someone else says is called *retweeting*; this behavior facilitates the spread of information and commentary in real-time.

We have performed a large-scale collection of data from the Twitter social network by means of the publicly available application programming interface (API) they provide. Our data set contains over 2.7 billion messages, 80 million user profiles, and a 2.6 billion edge social network. We analyze these data and uncover the growing trend where users defeat Twitter's simple privacy mechanism of "protecting one's tweets" by simply retweeting a protected tweet.

We have shown through quantitative and qualitative analysis that these privacy-violating retweets are a growing problem. More than 4.42 million tweets exist in our corpus that expose protected information. As Twitter gains popularity over time, we see an increasing trend in the number of privacy-violating retweets. Although there are many users who are unaware that their private tweets are being broadcast to the public, there are some who are aware of this problem.

I. Introduction

We have identified a serious and growing privacy problem within Twitter, an extremely popular web service which allows its users to easily broadcast anything they want and follow what others are saying. The only privacy setting Twitter offers its users is the option to "protect one's tweets." The idea behind protecting one's tweets is to allow only the protected user's followers to view his tweets. In theory, this would allow a Twitter user to control who can see his tweets, but in reality, his tweets and the information contained within them can be broadcast to the public.

In this paper we analyze exactly what it means to "protect one's tweets." We examine the leaked tweets of 5.27 million users who have protected accounts. The community-driven phenomenon of retweeting enables protected tweets to leak to the public sphere by users copying and pasting the text of protected tweets into their own Twitter feed. Unfortunately, retweeting allows protected tweets to be made public without the approval of the original tweet's author. Our results illustrate that not only are one's tweets never guaranteed to be "protected," but for those who think their tweets are protected, they may have more information leaking into the public realm than they were aware of.

Although this problem may not be surprising to frequent users of Twitter, we define the extent and rapid growth of the problem. We have a unique opportunity to study the behavior of over 62.6 million Twitter users by examining over 118.74 million retweets. Twitter users will soon, if not already, face the same problems that many Facebook users encountered when the service became open to the general public. For example, Twitter is already indexed by Google. Users will need to understand that their tweets can be cached and searched, future and current employers may be reading their tweets, and they have no way of monitoring exactly who can read their tweets.

In this paper we will describe some Twitter fundamentals and related work, the data we have collected with accompanying quantitative and qualitative analyses of possible information leaking, and conclude with a few comments on the future of this privacy issue.

II. Related Work

Early research on Twitter focused mainly on understanding the size, relationships, and tweeting behavior as the network grew [1], [2]. While these census-style reports have continued based on ever growing Twitter API data dumps, other more sociological approaches to studying Twitter have also arisen.

The Pew Internet & American Life Project's "Twitter and Status Updating" report, updated for Fall 2009, describes the data from a sample of over 2000 adults [3]. This approach allows us to better understand the people who are actually using Twitter. Additionally, this report estimated about "19% of Internet users now use Twitter or another service to share updates about themselves, or to see updates about others."

Research has examined retweeting as a way to classify influential tweeters [4], and conversation between users as a method of communication [5]. However, boyd's "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," is the definitive paper on this emergent behavior [6]. boyd's work explores retweeting behavior in depth, detailing: the varying methods for retweeting, why users retweet, the types of content that are retweeted, and finally details a number of specific ways retweets propagate through Twitter. These papers do not explicitly examine privacy concerns.

Privacy has been studied extensively in terms of ubiquitous computing environments and other online social networks (OSNs) [7], [8]. While OSN privacy studies have most commonly been framed around policy and configuration [9], [10] or specific types of information leaking (e.g. location [11], photos [12], SSNs [13]), Acquisti and Gross's initial work broadly examined user awareness of privacy across the network. Their results show that many users rely on their own ability to control information dissemination even as they may have remained unaware of how their information actually travelled through the network and the policies that governed the system.

Our work goes beyond previous work in two ways. First, Twitter is a largely untapped domain in terms of privacy studies. Generally, Twitter is seen as an entirely public system, however their "protected" account feature breaks that model and creates an avenue for examining privacy leakage on Twitter. Second, status updating is becoming more prevalent across the web, from Twitter, to Facebook statuses, to the latecomer Google Buzz. Understanding the semantics of privacy with Twitter updates will help inform future studies of status updates on other networks. For example, earlier this year, Facebook rolled out status updates with privacy protection.

III. Twitter Fundamentals

Twitter is a social networking website that allows users to publish short messages known as *tweets*. These messages may be up to 140 characters long, but other than length there are no restrictions on the message. Moreover, a user can *follow* other users they find interesting in order to get real-time updates of content those users publish. A plethora of desktop- and mobile-based applications make Twitter more accessible.

We first describe the registration process and focus on how the privacy policy and options are presented to the user. We then provide an overview of the interfaces for several Twitter applications. We also discuss two popular community conventions, @-mentioning and retweeting, that are critical to the analysis we perform.

A. Registering For An Account

Although having a Twitter account is not required to see content shared on the network, more than 80 million people have created accounts allowing them to post content. Signing up for an account is straightforward and requires a user to enter their full name, a *username* that will be used to identify the user, a password, and an email address. During the entire registration process the only mention of privacy on Twitter is in the Terms of Service, which tells users to read the privacy policy for the relevant information. Their privacy policy then links to their privacy settings; however this is a third, optional step users must follow while registering to configure their optimal privacy settings. At no point does the option to protect their account appear in the registration process, without taking these branching steps. A user must specifically seek out this functionality *after* their account has been created.

B. Tweet Conventions

Because the messages are so short, certain conventions have been devised by the Twitter community to increase the expressiveness of a tweet. The first convention is making an *@-mention* (at-mention) of another user. @-mentioning allows users to reference a specific user in the body of a tweet. Users simply place an @ symbol in front of the username they wish to reference. When used at the beginning of a tweet, the @-mention indicates that the message has an intended recipient. We call this convention *@-messaging* and note that @-messages only appear in the timeline of the recipient (and those following both the sender and recipient) in order to avoid crowding other users' timelines. Although these messages don't necessarily appear in other users' timelines, they are still visible by going to the authoring user's page.

Often, a Twitter user will see another user's tweet and wish to rebroadcast that message. This is traditionally accomplished with a *retweet*. A user will preface their tweet with the token "RT", @-mention the user who originally posted it, and include the original message. For example, if Bob tweeted "It is sunny in California." and Alice wanted to retweet that message, she would tweet "RT @Bob: It is sunny in California". Retweeting also allows users to provide commentary by adding additional content. An example of such commentary would be "That's no surprise! RT @Bob: It is sunny in California."

Retweeting was accomplished by the copy-and-paste method illustrated above. In November 2009, Twitter introduced functionality making retweeting a feature of the service; these 'official' retweets are treated differently than ones created by the copy-and-paste method. In particular, retweets appear differently in a user's timeline, and user commentary cannot be added. The author of the user being retweeted is displayed in the timeline, and a small footnote link indicates which person retweeted the message. The official retweet functionality respects user privacy settings in that a tweet of a protected user cannot be retweeted. However, many users continue to use the copy-and-paste retweet convention as it allows them to add comments of their own.

C. Methods of Posting to Twitter

A detailed analysis of the source of tweets, which describes the method or application by which the user sent their tweet to Twitter, indicates that on any given day over half of the content on the service was created in the web interface. Any single application rarely has more than 8% of daily tweet volume and a specific application's usage varies greatly as new applications are developed.

The method used to author each tweet is provided through the API methods and is also visible on the website and in a number of clients. We aggregate the number of tweets generated by each particular method across our entire corpus. Additionally, we calculate the fraction of tweets on a given day that are generated by the various applications and interfaces. A summary of overall application usage is found in Table I. The 'web' method corresponds to tweets authored through the Twitter.com website interface, 'txt' represents tweets created through the text message service available on cell phones, and the 'API' method is a blanket category; any application that uses the API but chooses to not specify an application name falls into this category. The rest of the list describes common, third-party applications used for tweeting.

Application	Percent
web (Twitter.com)	49.32%
TweetDeck	5.56%
txt (SMS)	4.42%
Echofon	4.11%
twitterfeed	4.07%
API	3.33%
UberTwitter	2.41%
Mobile web	2.40%
Tweetie	2.03%
Twitterrific	1.76%
All Others	20.59%

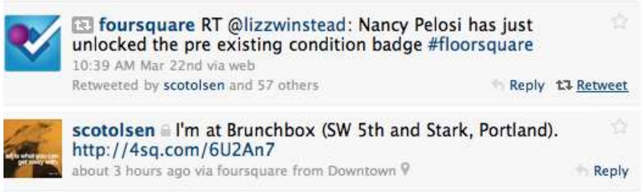
Table I
USAGE STATISTICS FOR THE MOST POPULAR TWEET SOURCES

Most of these clients are either dedicated interfaces to Twitter that run on mobile platforms or desktop software that integrates multiple social platforms such as Facebook, Twitter, MySpace, and Flickr.

D. Retweeting Through Different Clients

As discussed above, we will examine the phenomenon of retweeting in detail. Retweeting, as it is performed through the interface, varies depending on the client used. In order to examine a sample of interfaces we will highlight the interfaces for retweeting on the web and using both Tweetdeck and Echofon, two popular clients using both their desktop and mobile applications.

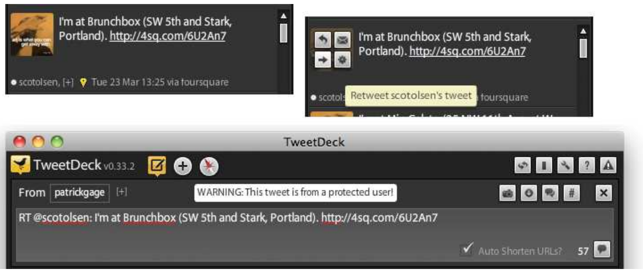
Using Twitter On The Web: After logging in the user is taking to their main view of service, which is similar on most clients. This is a view that shows a reverse-chronological list of tweets, displaying only tweets that have been written or retweeted by Twitter users that the logged-in user is following. Individual tweets contain not only the message text, but also the user's selected icon, a lock icon if the tweet is private, the time it was tweeted, the client through which it was tweeted, a hollow star by which a user can "favorite" the tweet, and up two action buttons. The first is Reply, which automatically fills the field where a user can create a new tweet with the @ symbol followed by that user's handle. The second is Retweet, which only appears if the user is not protected, allowing the user to Retweet to their followers what their follower had said. For a comparison between a non-protected tweet, a protected tweet, and an official retweet see Figure 1.



foursquare RT @lizzwinstead: Nancy Pelosi has just unlocked the pre existing condition badge #floorsquare
10:39 AM Mar 22nd via web
Retweeted by scotolsen and 57 others
Reply Retweet

scotolsen I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
about 3 hours ago via foursquare from Downtown
Reply

Figure 1. Twitter's website, showing a retweeted tweet, and a user's ability to further retweet it, compared to an un-retweetable protected tweet below.



I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
scotolsen, [+] Tue 23 Mar 13:25 via foursquare

I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
Retweet scotolsen's tweet

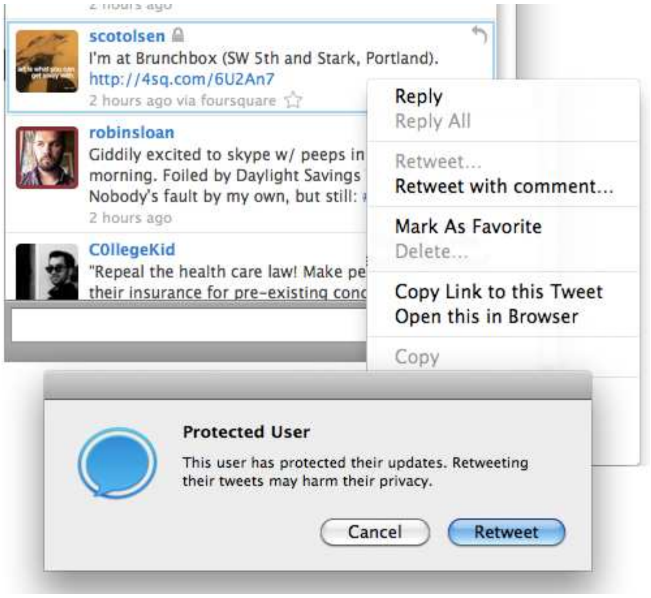
TweetDeck
TweetDeck v0.33.2
From patrickgage [+]
WARNING: This tweet is from a protected user!
RT @scotolsen: I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
✓ Auto Shorten URLs? 57

Figure 2. TweetDeck's desktop client, showing a user about to retweet protected content, and the passive warning they receive during the process.

Using TweetDeck and Echofon on the Desktop: Using a desktop can provide a very different user interface than using Twitter on the web. Here we examine two different desktop Twitter clients, TweetDeck and Echofon.

Tweetdeck has a similar view of a single tweet, where users can retweet by hovering over the user's icon and clicking the forward icon. Note there is no indication that this user has protected tweets. In the example shown in Figure 2 the user is retweeting a protected tweet. While the client still fills the field with the protected tweet, in the copy-and-paste style there is a brief warning message that states "WARNING: this tweet is from a protected user!" yet still allows and facilitates the retweeting.

Echofon has a similar display interface, where retweeting is performed by right-clicking on the tweet. In Figure 3 we see there are two options for retweeting: the official methods and the copy-and-paste methods, denoted by "Retweet with comment..." Since the official retweeting API blocks retweeting protected tweets this option is grayed out, further reinforcing the lock icon shown with the tweet. However, unofficial retweeting can still be selected, leading to the pop-up shown below, which further attempts to block the user



I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
2 hours ago via foursquare

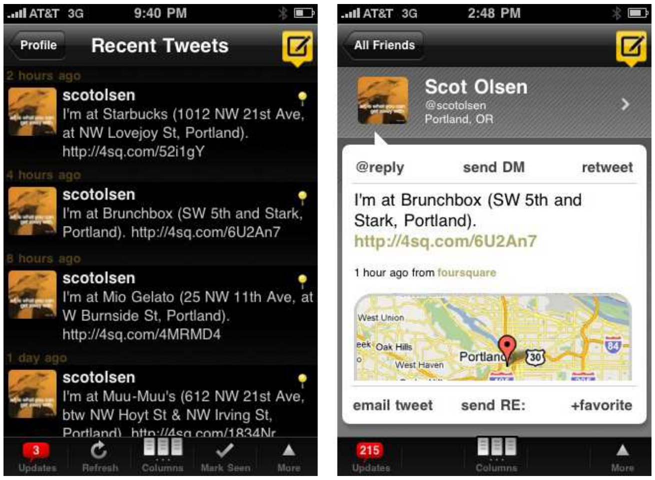
Giddily excited to skype w/ peeps in morning. Foiled by Daylight Savings Nobody's fault by my own, but still:
2 hours ago

"Repeal the health care law! Make pe their insurance for pre-existing conc

Protected User

This user has protected their updates. Retweeting their tweets may harm their privacy.

Figure 3. Echofon's desktop client, showing a user viewing a protected tweet, the contextual menu that would allow retweeting, and the notification a user would receive if they attempted to retweet it.



I'm at Starbucks (1012 NW 21st Ave, at NW Lovejoy St, Portland). http://4sq.com/52i1gY

I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7

I'm at Mio Gelato (25 NW 11th Ave, at W Burnside St, Portland). http://4sq.com/4MRMD4

I'm at Muu-Muu's (612 NW 21st Ave, btw NW Hoyt St & NW Irving St, Portland). http://4sq.com/1834Nr

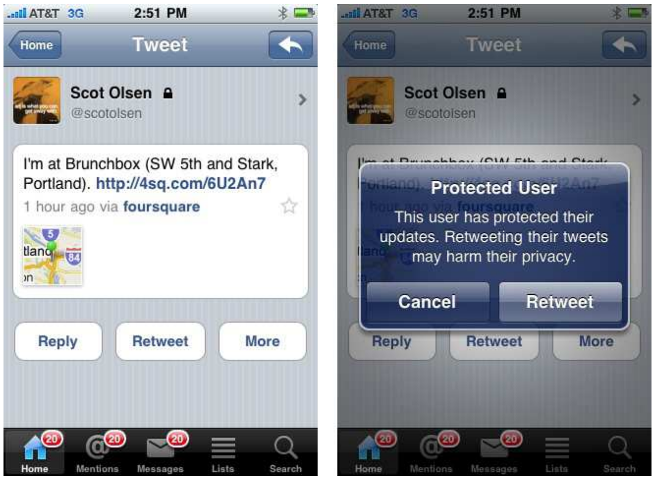
I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
1 hour ago from foursquare

Figure 4. TweetDeck's mobile client, showing a user about to retweet protected content.

from retweeting protected content.

Using TweetDeck and Echofon on Mobile Phones: Switching to the mobile interfaces where screen real-estate is at a premium we will walk through the method for retweeting from TweetDeck and Echofon's mobile iPhone clients.

TweetDeck shows tweets in a scrolling list, which can then be clicked for more information and to retweet, as shown in Figure 4. There is no lock icon or any information regarding the fact that this tweet was made from a protected user, although there is information regarding where the tweet was made, including a map, information to interact with it, the client it was made with, and basic interface controls. Here selecting re-tweet moves the user straight to the "Retweet Classic" interface with no interstitial notification, pop-up, or any other notice that you might be trying to retweet protected content.



I'm at Brunchbox (SW 5th and Stark, Portland). http://4sq.com/6U2An7
1 hour ago via foursquare

This user has protected their updates. Retweeting their tweets may harm their privacy.

Figure 5. Echofon's mobile client, showing a user viewing a protected tweet, and the notification a user would receive if they attempted to retweet it.

Echofon has an interface similar to their desktop client. Again from the single tweet view shown in Figure 5 we see the lock icon indicated that user has his tweets protected. While the option to retweet exists, clicking it yields the pop-up on the right informing the user that this action may "harm their privacy." Even in the mobile client Echofon has managed to encourage users not to retweet protected tweets.

IV. Methodology

This paper focuses on understanding the extent to which the sole privacy mechanism in Twitter is compromised. Our analysis is comprised of a quantitative and qualitative examination of tweets that violate a user's privacy settings. We have collected a large corpus of Twitter data as part of a broader research agenda and use this data set in our analysis. Our data set contains over 2.7 billion messages, profile information for more than 80 million users, and a snapshot of over 2.6 billion relationships between these users. A detailed summary of the quantity of data we have collected is found in Table III. After examining the literature ([4], [14], [15]), we have concluded that this data set is one of the largest used for such analysis. The remainder of this section describes the methodology used to collect the corpus, as well as the processes used to extract the subset of data used in this analysis.

A. Data Collection

Like many web services, Twitter provides a publicly accessible Application Programming Interface (API) which allows programmers to develop applications that utilize and extend the service. In our case we use this API solely to collect user data including: messages, social graph data, and user profiles. Twitter implements their API through HTTP and an inquisitive reader can simply use a web browser and the Twitter API documentation [16] to explore the functionality we use.

To avoid excessive use and abuse of this service, the number of requests per client (IP address or [user, IP address] pair) is *rate-limited*. The baseline rate of requests is 150 per hour per client; however, user accounts and IP addresses can make up to 20,000 requests per hour after getting *white-listed*. Even at this rate it would take more than 660 days for a white-listed machine to crawl the amount of data we have. These limitations necessitate the creation of a distributed crawling infrastructure to acquire data at the scale we have achieved. Our crawler runs on over 80 machines simultaneously and is capable of downloading tens of gigabytes of data per hour.

We crawl the network by sequentially probing the user ID space. Twitter assigns unique user identifiers in order, so collecting information for each user is a straightforward task. For each user account that exists we retrieve the profile, up to the last 3,200 tweets published, and the social connections. It is interesting to note that the API restricts retrieving more than 3,200 tweets, but allows retrieving arbitrarily old tweets so long as the authoring user hasn't created more than 3,200 messages. Almost all users we have crawled have created fewer than 3,200 tweets, and for these users we possess a complete history of messages.

B. Measured Features

We use only four API methods to get each user's profile, tweets, and social connections. A description of what data is provided by each of these API calls is given in Table II. All profile information is available regardless of the user's privacy settings; privacy-conscious users should provide only a minimal amount of profile information. A protected user's tweets are clearly unavailable to us and we make no efforts to circumvent these settings by attempting to follow protected users. The methods for getting social connections are restricted in that through the API it is not possible to get a protected user's friends or followers list; however, this information is available on the website. This inconsistency is even more confusing because the fact that an unprotected user U is following a protected user P is discoverable by requesting U's friends. Although protecting an account should have no effect on social link visibility, it is clear that Twitter treats protected and unprotected accounts differently and may make such concealment part of protecting an account in the future.

C. Data Processing and Extraction

After downloading massive quantities of raw data from Twitter, we extract and format the responses into a more usable format. Most of the analysis we perform for this research and other projects involves using a combination of AWK and Python scripts, grep, and the Hadoop MapReduce infrastructure. We use a set of flat, tab-delimited text files to organize and manipulate the corpus.

API Method	Retrieved Information
users/show (always available)	user ID, screen name, account creation time, number of friends / followers / tweets / favorites
users/show (optionally provided)	'real' name, location, description website, time zone, geo-coordinates
followers/show **friends/show**	IDs of users following specified user IDs of users followed by specified user
statuses/show	unique message ID, message content, posting user ID, in response to user ID, in response to message ID, post time, update method, geo-coordinates

Table II
SUMMARY OF API METHODS USED

Starting in July 2009 we developed the crawling infrastructure and performed some preliminary crawls. By the end of August 2009 we had completed an initial crawl of approximately 62.6 million user accounts. At this time changes in computing infrastructure prevented us from performing subsequent recrawls until November 2009 and January 2010. During both of these recrawls we focused on gathering information from new users and were unable to fully revisit all users with accounts older than September 1, 2009. Due to possible gaps in our overall data set, we restrict ourselves to analyzing accounts and Tweets authored before September 1, 2009.

In this paper we focus on extracting and analyzing a subset of our corpus. We have extracted all retweets containing the string "RT @UserName" where UserName is a protected user. Pulling out retweets is somewhat error prone because of the different variants people use. For example, one such variant is to give attribution using the string 'via @username.' However, because the conventional retweet pattern is so widely used, it is the only one we use to extract our data set. We do not attempt to resolve typos of user names and only perform basic parsing so that 'RT @Username[punctuation]...' is correctly handled.

V. QUANTITATIVE DATA ANALYSIS

We first analyze the extracted tweets in a quantitative manner without regard to the semantic content of the messages. In the following discussion, a tweet is said to be *privacy-violating* if it exposes content that should be inaccessible because of privacy settings. Beyond simply counting the raw number of privacy-violating tweets (4.42 million) we also analyze several distributions that provide greater insight into the nature of this problem. The occurrences of privacy-violating tweets is relatively rare. Although only approximately 0.1% of all tweets fall into this category, approximately 4% of all retweets are privacy-violating. The remainder of this section summarizes the extracted data set, analyzes how often varying degrees of egregious retweeting behavior occurs, and examines the growth of this phenomenon over time.

A. Data Set Summary

Over the course of five months we performed three bulk crawls of data and collected over 700 GB of data through the Twitter API. A summary of the overall data set we collected is found in Table III.

Data type	Number collected
User Profiles	+80,000,000
Social Links	2,662,969,000
Tweets	2,782,666,000

Table III
SUMMARY OF ALL DATA COLLECTED

As previously mentioned, the first crawl performed in August 2009 was complete in that we had visited all users that existed at the time. As stated earlier, we restrict our analysis to accounts and tweets created before September 1, 2009. In Table IV we present a breakdown of user accounts created before September 1, 2009. Approximately 8.4% of users at that time were marked as protected. Also in Table IV we provide a count of how many retweets are in the data set, as well as the number of retweets that involve a protected user. We find that nearly 5% of protected users have had at least one tweet retweeted by a public user. Furthermore, nearly 1% of all Twitter accounts have retweeted private information.

Account type	Number of accounts	Percentage
All accounts	62,644,601	
Protected accounts	5,277,654	8.42% of all
Protected accounts with at least one tweet retweeted	247,161	4.68% of protected
Accounts having retweeted at least one protected tweet	563,760	0.90% of all

Tweet type	Number of Tweets	Percentage
Retweets	118,736,531	
Retweets containing content from a protected user	4,418,272	3.7%

Table IV
SUMMARY OF ACCOUNTS AND RETWEETS CREATED BEFORE SEPTEMBER 1, 2009.

B. Distribution of Offending Behavior

In absolute numbers, it is clear that retweeting of protected content was already a widespread problem through September 2009. With the continued growth of the network, this issue is surely worse today.

In Figure 6 we plot in a log-log graph the number of protected users that have had their content retweeted K times vs. K. We see that of the 247,000 protected users that have had content retweeted, over 86,000 (35%) of them have been retweeted only once. Overall, 186,000 protected users have been retweeted less than six times. These users account for 75% of all protected users who have been retweeted. This means that most protected user's privacy has either never been violated or has been violated only a handful of times. On the other hand, it is also more likely that these users are unaware that they have been retweeted because it is not a recurring problem. This distribution also provides insight into the other end of the spectrum, where over 18,400 protected users have been retweeted at least fifty times, and nearly 9,600 at least one-hundred times (including tweets retweeted multiple times).

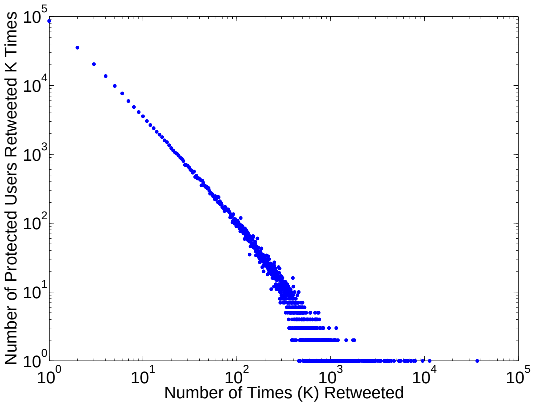


Figure 6. Number of protected users vs. number of times protected user is retweeted (log-log plot)

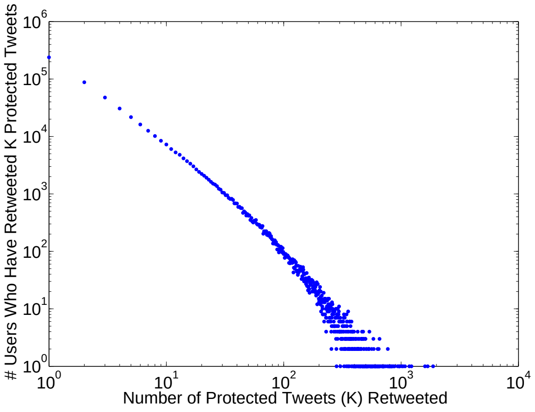


Figure 7. Number of users vs. number of protected tweets retweeted (log-log plot)

We also examine the distribution of how many users have made K egregious, privacy-violating retweets. Similar to the previous analysis we find that those users who have retweeted protected content have done so only a handful of times. 42% of users who have retweeted protected content have done so exactly once, and about 76% of users who have retweeted protected content have done so less than six times. Unfortunately, some individuals have blatant disregard for other's privacy settings. More than 16,400 users have retweeted protected content at least fifty times. Almost 5,800 users have done so at least one-hundred times.

Finally, we explore how many people were exposed to protected content every time a privacy-violating retweet was made. For each retweet containing protected content, we look at how many followers the author of the retweet has. The number of followers is based the number of followers the user had when we crawled the profile page. Therefore, this number is only an approximation since we do not have the exact number of followers the user had at the time the retweet was made.

This analysis results in a set of 4.42 million *audience sizes*, one for each of the offending retweets. The audience size of a retweet is our estimate of how many users would have that retweet appear in their timeline, as described above.

The cumulative distribution plot for the set of audience sizes is given in Figure 8. The average audience size is 538 users and the median audience size is 135 users. 90% of privacy-violating retweets reach no more than 730 users, while only 20% of privacy-violating tweets reach fewer than 20 users. Approximately 16,300 (0.4%) of privacy-violating tweets reach an audience of more than 10,000 users. If any reasonable percentage of Twitter users read their timeline, broadcasting a compromising statement to 10,000 user timelines would be devastating.

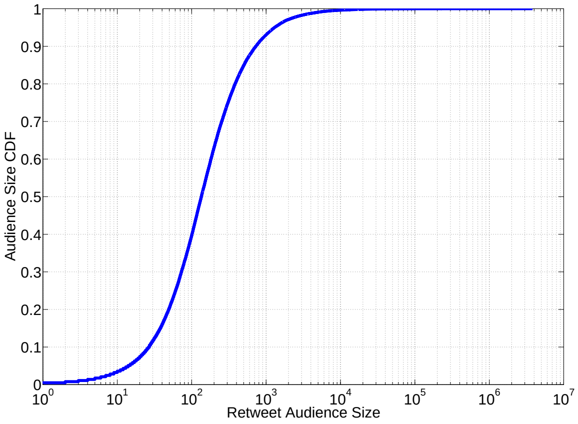


Figure 8. Cumulative density plot of audience size of privacy-violating retweets.

C. *Increase in Violations Over Time*

The impressive growth of Twitter has come with an increased risk for having private information leaked and repeated across the network. We have aggregated the number of privacy-violating tweets over time by week in Figure 9. It becomes apparent that this problem has grown significantly over the lifetime of Twitter; in the last week of the analyzed data set more than a quarter-million retweets contained protected content. Comparing this growth to the overall growth of Twitter as illustrated in Figure 10, it is clear that Twitter's increasing popularity has lead to an explosive growth in these kinds of retweets.

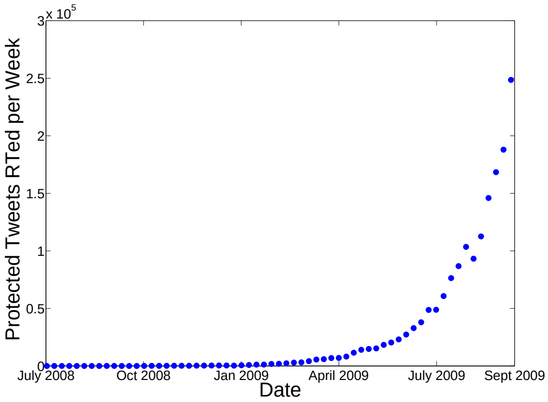


Figure 9. Number of retweets of protected tweets (aggregated weekly) over time.

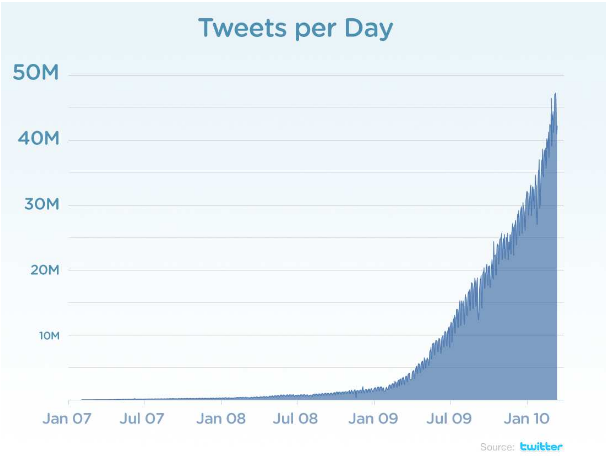


Figure 10. Volume of tweets per day over the lifetime of Twitter. [17].

Finally, we examine how the number of users who have their accounts protected, based on when they registered their account. In Figure 11, we show for each day what percentage of users who signed up on that day have protected accounts.

Several aspects of the graph stick out: the sudden drop in early 2007, the general decrease in variance over time, and the decrease of account protection throughout 2009. The anomaly in 2007 occurs right after Twitter wins the South-By-Southwest Web Awards [18]. Before that point, Twitter was only known to a small group of hardcore technology users. The publicity associated with winning this award brought Twitter to the attention of mainstream technology users, a large fraction of whom appear not to be concerned with online privacy. The decreasing of this percentage over time could be explained by the increase in registration volume. Finally, the decreasing number of protected accounts

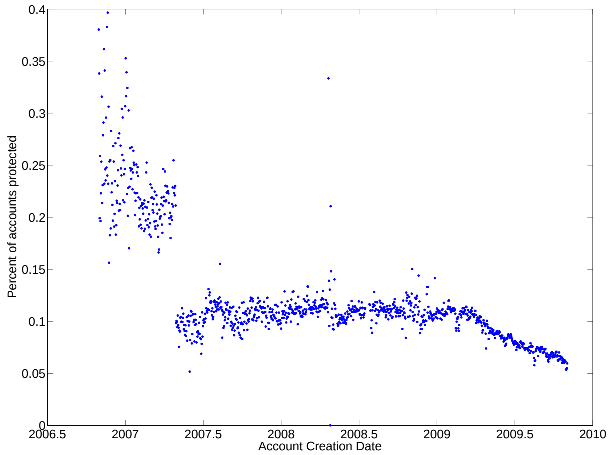


Figure 11. Fraction of accounts protected vs. account registration time.

throughout 2009 is likely driven by Twitter truly going mainstream. News outlets started referencing Twitter throughout 2009 when events such as the Iran election protests and Michael Jackson's death generated an enormous surge of traffic on Twitter [19].

D. *Continued Use of Copy-and-Paste Retweeting*

Because our comprehensive crawl of the Twitter network stops before the introduction of the official retweet mechanism, it is unclear that the privacy violations examined in this paper still occur with great frequency today. We attempt to address this matter by showing that the original style copy-and-paste retweets are still prevalent today. To do so, we examine data from the Gardenhose stream provided by Twitter. This stream provides access to a real-time random subsample of approximately 5% of all tweets. In Figure 12 we show the number of Tweets per day in the Gardenhose stream that are retweets (of any kind), copy-and-paste style retweets, and new official retweets. The significant increase in retweet volume over time immediately stands out. Between January 1, 2010 and April 1, 2010 there is nearly a doubling in total retweet volume as well as a doubling in copy-and-paste style retweet volume. What is most shocking is that on April 1, 2010 twice as many copy-and-paste style retweets were found the subsampled stream (no more than 5% of tweets) as were found in the whole last week of August 2009.

In order to clearly show how the Twitter retweet feature has changed the retweeting behavior of users over time, in Figure 13 we plot the fraction of retweets in the Gardenhose stream that are crafted by the copy-and-paste method. The sharp drop-off corresponds to the introduction of the official retweet feature in November 2009. We notice that within several months the fraction of retweets that are manually authored has stabilized at around 60%.

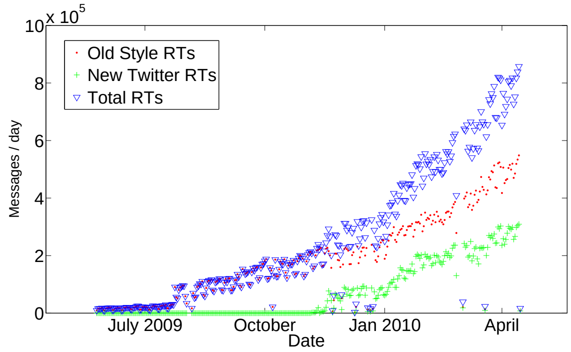


Figure 12. Number of retweets, broken down by manual- and official-style, in the Gardenhose stream over time.

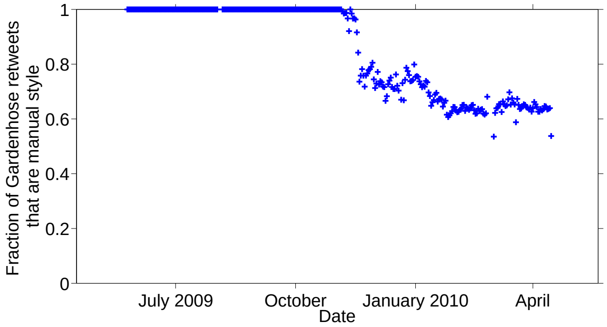


Figure 13. Fraction of reweets in daily stream that are copy-and-paste retweets over time.

VI. Analysis of Privacy-Violating Tweets

Twitter users often treat their accounts like journals or a place to vent their frustrations. Through qualitative analysis we have found many instances of users complaining about their family, friends, coworkers, personal lives, etc (Table VII). This might be considered acceptable if Twitter users could be sure that those whom they are complaining about will never find out. A protected user could feel this way as only his followers can see his tweets, but as we see below, as soon as a protected tweet is retweeted anyone can see it.

Although ”protecting one's tweets” may have been a well intentioned idea, the general lack of understanding by Twitter users has rendered this feature nearly useless. From the data, we know that of the 62.6 million Twitter users, 5.27 million have protected accounts (8.42%). We can also see that $247,000$ (4.7%) of the protected users are being retweeted.

We decided to analyze the retweets of protected users to see exactly what types of information were leaking from protected users. In the following sections we give estimations for the frequencies at which various topics are mentioned in tweets as well as examples of actual tweets which we have anonymized.

Non-English Language	Advertisement	Family
Opinion	News	Relationship
Description of Self	Joke	Work
Friends	Location or Contact Information	other

Table V
TAGS USED TO CODE RETWEETS OF PROTECTED USERS

Tag	Count	Frequency
Non-English Language	373	35.32%
Opinion	246	23.30%
Description of Self	116	10.98%
Friends	87	8.24%
Advertisement	74	7.01%
News	65	6.16%
Joke	39	3.69%
Location or Contact Information	18	1.70%
Family	11	1.04%
other	11	1.04%
Relationship	8	0.76%
Work	8	0.76%

Table VI
FREQUENCY OF TAGS CODED FOR 1000 RETWEETS OF PROTECTED USERS

Frequencies of Topics Found in Protected Tweets

To get a sense of the amount of private data being leaked, we decided to due an analysis of how often various topics came up in our retweet analysis. After reading through many tweets, we came up with a list of 12 tags (see Table V) which we feel accurately describe most tweets. To calculate the estimated frequencies of these tags, we randomly selected and manually tagged $1,000$ tweets from the 4.42 million total retweets. Tagging involved having two people independently code each of the $1,000$ tweets with at most two tags. If a tag match occurred between the two coders then the matching tag was assigned to the tweet, otherwise, a third person was brought in to break ties. The frequencies or topics are show in Table VI. We chose to hand code the tweets to achieve as much accuracy in tagging as possible. In total, we had $1,056$ tags for $1,000$ tweets. Although the frequencies for tags which may be considered privacy sensitive (like Location and Contact Information) are rather low, we illustrate in the following subsections how the tag "Description of Self'" and others might be privacy sensitive as well.

References to Family & Friends

In this section we examine retweets which mention family. We find examples of protected tweets which should probably remain private. We also find examples of photographs which were originally shared with a select group of people making their way into the public realm. The links to photographs often lead to the protected user's photo account and therefore more photos.

- "RT @ProtectedUserName: At my in-laws. Pure hell. $<-$ happy holidays!"
  - This protected user is complaining about being at his in-laws and the public user has retweeted this and added her own comment, "happy holidays!".
- "*DONE* RT @ProtectedUserName: My sister is shaped like a linebacker. I'm glad that I didn't get those genes."
  - This protected user insults her sister, and the public user has retweeted the insult so that the sister could possibly see it now.
- "Wooohoooo! RT @ProtectedUserName: Here's a pic of my hot *ss wife on our way out to Tao http://twitpic.com/????"
  - This protected user shares a picture of his wife with his followers who then broadcasts it out to the public by retweeting. By clicking on the link, people are directed to the protected user's twitpic account which displays other photos posted by the protected user.

References to Contact Information

We now examine some common examples of people sharing their contact information. There are many Twitter users who post their phone numbers or their locations. This could result in uninvited guests like thieves. In fact the site pleaserobme.com, was started to prove just how easy it could be to compile this type of public information.

- "#PHONEBLAST B*TCH! RT @ProtectedUserName GUYS - my phone number is +########### and will prb stay that until I get to the UK"
  - This public user intentionally and maliciously broadcasts the protected user's phone number.
- "RT @ProtectedUserName: If you need to reach me tonight, I'll be at (###) ###-#### Where is that? "
  - Perhaps unitentionally, this public user retweets the protected user's phone number so that he can ask a question regarding the number.
- "RT @ProtectedUserName Come to my birthday tonight at DBA Gallery & Wine Bar.. The address is 256 S Main St Pomona, CA 91766.. No cover or dresscode.."
  - The public user is sharing the location of a protected user's party and at the same time broadcasting that the protected user will not be at home in the evening.

References to Employment

Many Twitter users are frustrated with their current employment situations, however, we highly doubt that these users would want their co-workers and especially their bosses to be reading their tweets. Below we have examples of Twitter users admitting to misusing company time and insulting/threatening their coworkers and bosses.

- "Haha! Don't hurt 'em! RT @ProtectedUserName: I'm about to use company time to look for a new job."

  - The protected user is misusing her job time and the public user is now broadcasting this information.
- "I Hear YAAA RT @ProtectedUserName: ugh stressin about work, f*ckn bitches i work w/ have NO work ethic what so ever! grrr so frustrating!!"
  - This protected user is badmouthing her coworkers and if any of them are friends with the public user who retweeted this post then they will find out immediately.
- "RT @ProtectedUserName: restraining the urge to assassinate my boss"
- "Lol I agree RT @ProtectedUserName: I wish my boss would grow some f*cking testicles and quit being a c*nt"
  - Both of these protected users are referring to their bosses negatively.

Poor representation of self

The public should be aware that many employers scour the internet to check up on their employers, or potential employees. Similarly to how many people realized that they should edit their public Facebook information and photos, Twitter users should consider editing their tweets. Users who protect their tweets are under the false impression that they can tweet whatever they want. Below we have some examples of tweets that protected users would probably want to hide from potential employers.

- "And proud. RT @ProtectedUserName: i am racist and i admit it."
  - This protected user shares something negative about himself which is then broadcast to the public.
- "RT @ProtectedUserName: i love to smoke weed"
  - This protected user shares that he enjoys smoking an illegal substance.
- "Dam u look good bak then 2 RT @ProtectedUserName – me when i was 15; lmao. http://www.twitpic.com/?????"
  - This protected user shares a racy photo of herself as a minor which is then broadcast to the public.
- RT @ProtectedUserName I googled myself & my mess of a twitter page came up... Not a good look 4 my employer & future employers to witness. Protected.
  - This protected user understands that his online persona reflects poorly on himself, but he doesn't understand that his formerly public information can still be searched.

Public users disregarding and misunderstanding the protected status

Although public users may disregard or not care about how they are perceived online, users who protect their tweets may be more concerned about their online reputation. Because of this, there are some users who request that their followers not retweet their protected tweets. Unfortunately the users who have protected tweets cannot force their followers to heed their wishes to keep things private, and as we will see, public users often disregard these requests. There are also cases, where public users do not understand what a protected tweet is, as is illustrated by one user who asks why he cannot use the automatic retweet feature on a protected tweet.

- "RT @ProtectedUserName do me a favor twitter that follow me please dont rt my sh*t i get introuble every time ya do((so don't tweet lmao))"
  - This protected user has experienced unwanted people reading his tweets and requests that his follower stop retweeting his tweets, possibly in a joking manner.
- "LOL Yes, sure are! But my tweets aren't protected my dear! LOL RT @ProtectedUserName @PublicUserName its okies, we're all friends"
  - This public user understands that retweeting her friend's protected tweets makes them public to all and is now informing her protected friend of such.
- "RT @ProtectedUserName: Twitter is d only safe place for her to not follow me here. Feeling so protected here.
  >>; Haha, I hear u. Driving now?"
  - This protected user believes that his tweets are unreadable by the public and the public user who is broadcasting them to the public believes that as well.
- "RT @ProtectedUserName Why can't I RT protected tweets?"
  - This protected user does not understand that retweeting allows other users who are not the original author's followers to see the protected tweet. This would imply that this protected user does not understand that when his tweets are retweeted, they can be seen by people who are not his followers. This public user does not answer the question but instead shows that protected tweets can indeed be retweeted using the copy-and-paste method.
- "Doesnt surprise me! RT @ProtectedUserName: I just read on LATIMES.COM that protected tweets can be googled!"
  - This protected user falsely believes that her protected tweets can be searched, but this is only true if public users RT her, like her public follower is currently doing.

VII. Discussion

The protected tweet feature in theory allows one to keep tweets private; however, in reality many protected tweets leak into the public realm. Many protected users are probably unaware of the extent that their protected tweets are

Retweets of A Sensitive Nature		
Category	Sample keywords	# retweets
Past & present boyfriends & girlfriends	boyfriend bf ex-boyfriend ...	74829
Immediate family & in-laws	mother-in-law dad sister in-laws ...	444113
Job & coworkers	boss employees work ...	94704
Phone numbers	(###)###-#### ###.###.#### ...	2015
Pictures & videos	twitpic flickr twitvid tweetube ...	109668
Taboo topics	penis n*gger c*nt slut ...	247223

Table VII
NUMBER OF PRIVACY-VIOLATING RETWEETS WHICH REFERENCE SENSITIVE TOPICS

readable by the public, and although there are some users who are aware of this problem, there is little they can do to prevent it.

It is also evident that many Twitter users do not censor their tweets, broadcasting information and language which could be detrimental to them in the future. Since only tweets can be protected and not the rest of their profile, which often includes their actual first and last name and location, leaked tweets could be tied back to these users. It is important for Twitter users to realize that this leaked information, as well as their social connections on Twitter, is available to anyone who has their username. Users who post pictures are inviting the public to view their pictures and potentially their entire web album. Twitter users should also be aware that everything that is publicly tweeted can be cached and retrieved for later viewing.

An interesting phenomenon to look at more closely in the future is the trend of decreasing protected accounts. Of the 5.27 million protected users we looked at back in September 2009, only 4.83 million are still protected (March 1, 2009). We know that 95,600 of the protected users have deleted their accounts and 353,000 users have changed their accounts to public. However, we do not know of the users who have joined since September 2009 how many currently have protected accounts.

Twitter's introduction of the official retweet feature could help mitigate protected tweets becoming public. This requires widespread adoption of the official method from both users and client developers. Based on the Gardenhose data we have analyzed, copy-and-paste retweeting is still widely used by users, with more than half of all retweets still using the older more organic method. Follow-up crawls will let us determine whether privacy-violating retweets still occur with non-negligible frequency. Any client-side impediments to retweeting protected content will help slow this trend. However it is currently unclear how many developers are considering this issue and future research should examine how much these pop-ups, notifications, and lock icons actually effect user behavior.

Our analysis of privacy-violating tweets only covered retweets which contained protected content as we could automatically extract these tweets from our data set. Another interesting type of privacy-violating tweet would be @-messages directed at protected users. This type of tweet would be similar to listening to half of a conversation where we know who is involved in the conversation. From the 819 million @-messages that we have collected, 106.33 million of those are directed at protected users (13%). An interesting direction for future work would include analyzing these privacy-violating @-messages.

VIII. CONCLUSIONS

We have analyzed, in detail, the extent to which information that is protected by a user's privacy settings is made public on Twitter. The popular community convention of retweeting enables this information leakage. While only a small fraction of tweets are retweets of protected users, it seems improbable that individuals are likely to systematically exploit these privacy violations. Yet, in the case of a single person targeting a specific 'victim' or a particular named place or type of content, it is possible to use the search API to uncover protected leaks. To some extent this would only be possible through some sort of automated processing on a large corpus of data, or through persistent and repeated search queries. Despite these difficulties, it is clear that the private information being made public through this analysis could be of significant concern for many users of the service, who may believe they have protected their content from the public. Methods for users to audit the spread of their private information and interfaces that encourage users to more thoughtfully consider the retweeting of protected content should be developed and applauded by the community.

IX. ACKNOWLEDGMENTS

Brendan Meeder and Jennifer Tam are supported by National Science Foundation Graduate Research Fellowships. We acknowledge Brendan O'Connor for providing us with his copy of the Gardenhose stream, a resource we have not yet collected. Finally, we give great thanks to Greg Kesden for allowing us to use computational resources without which this research would not be possible.

References

[1] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. New York, NY, USA: ACM, 2007, pp. 56–65.

[2] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about twitter," in *WOSP '08: Proceedings of the first workshop on Online social networks*. New York, NY, USA: ACM, 2008, pp. 19–24.

[3] S. Fox, K. Zickuhr, and A. Smith, "Twitter and status updating, fall 2009," 2009. [Online]. Available: http://www.pewinternet.org/Experts/~/link.aspx?_id=6C747837133C4A54A4D0351E2683478B&_z=z

[4] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *WWW 10: Proceedings of the 19th international conference on World Wide Web*, 2010.

[5] C. Honeycutt and S. C. Herring, "Beyond microblogging: Conversation and collaboration via twitter," *Hawaii International Conference on System Sciences*, vol. 0, pp. 1–10, 2009.

[6] danah boyd, S. golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." in *HICSS-43*, 2010.

[7] V. Bellotti and A. Sellen, "Design for privacy in ubiquitous computing environments," in *ECSCW'93: Proceedings of the third conference on European Conference on Computer-Supported Cooperative Work*. Norwell, MA, USA: Kluwer Academic Publishers, 1993, pp. 77–92.

[8] A. Acquisti and R. Gross, *Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook*, 2006. [Online]. Available: http://dx.doi.org/10.1007/11957454_3

[9] S. Patil and J. Lai, "Who gets to know what when: configuring privacy permissions in an awareness application," in *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2005, pp. 101–110.

[10] J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh, "Who's viewed you?: the impact of feedback in a mobile location-sharing application," in *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*. New York, NY, USA: ACM, 2009, pp. 2003–2012.

[11] J. Tsai, P. Kelley, L. Cranor, and N. Sadeh, "Location-sharing technologies: Privacy risks and controls," in *37th Research Conference on Communication, Information and Internet Policy (TPRC '09)*, 2009.

[12] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair, "Over-exposed?: privacy patterns and considerations in online and mobile photo sharing," in *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2007, pp. 357–366.

[13] R. Gross, A. Acquisti, and H. J. Heinz, III, "Information revelation and privacy in online social networks," in *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. New York, NY, USA: ACM, 2005, pp. 71–80.

[14] J. Leskovec, "Dynamics of large networks," Ph.D. dissertation, Machine Learning Department, School of Computer Science, Carnegie Mellon University, 2008.

[15] ——, "Stanford large network dataset collection." [Online]. Available: http://snap.stanford.edu/data/index.html

[16] "Twitter API Documentation." [Online]. Available: http://apiwiki.twitter.com

[17] "Twitter Blog: Tweets per day." [Online]. Available: http://blog.twitter.com/2010/02/measuring-tweets.html

[18] "Twitter Blog: We Won!" [Online]. Available: http://blog.twitter.com/2007/03/we-won.html

[19] "How michael jackson's death shut down Twitter, brought chaos to Google... and 'killed off' Jeff Goldblum." [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-1195651/How-Michael-Jacksons-death-shut-Twitter-overwhelmed-Google--killed-Jeff-Goldblum.html