Twitter as a Vector for Disinformation

P. R. Chamberlain

School of Computer & Security Science
Edith Cowan University, Australia
Email: prchamberlain@arkem.org;

Abstract
Twitter is a social network that represents a powerful information channel with the potential to be a useful vector for disinformation. This paper examines the structure of the Twitter social network and how this structure has facilitated the passing of disinformation both accidental and deliberate. Examples of the use of Twitter as an information channel are examined from recent events. The possible effects of Twitter disinformation on the information sphere are explored as well as the defensive responses users are developing to protect against tainted information.

Keywords: Twitter, Disinformation, Social Networks, Information Operations

Introduction
Twitter is an Internet social networking website that allows users to post short messages. Messages are sent from the website or from mobile phones or other devices. Like all networks Twitter is vulnerable to disinformation attacks but Twitter is especially susceptible due to the casual format of the messages and the asymmetrical structure of the relationship between nodes in the network. Twitter and other online communication media provide opportunities for organisations that would otherwise not have the resources to conduct disinformation campaigns with traditional mass media. This wider access to effective disinformation vectors means that there is a greater risk that information networks will be tainted. As a result the constituencies of these networks need to develop a strong sense of information assurance to avoid being compromised by disinformation.

Background
Twitter encourages people to exchange "quick, frequent answers to one simple question: What are you doing?" (Twitter, 2009). Users post updates about their day to day activities and current thoughts, and pass on Twitter messages (Tweets) from others as well as links to other websites. Tweets are by default publically accessible with only limited privacy and security options available. Tweets are limited to 140 characters but Universal Resource Locators (URLs) are encoded to a shorter form to conserve space via services such as Bit.ly or TinyURL (Miller J. L., 2009).

Twitter has powerful tools available to expand the Twitter experience, which also enhance the utility of Twitter for conducting information operations. Twitter has a powerful search system that searches all publically accessible recently posted tweets by keyword, and a trends display that shows what keywords are most popular at the moment (Twitter, 2009). Twitter also has a publically accessible programming interface to allow users to interact with Twitter programmatically. The Twitter API has features that allow for automated posting and datamining (Williams, 2009). The API tracking service at Programmable Web lists over 150 public Twitter API projects. Typical projects involve the aggregation of Twitter data with

other web data or the attempt to group Twitter data by various criteria — for example the geographic source of posts (ProgrammableWeb, 2009).

The Twitter social network

Networks among Twitter users are built with a 'follower/followed' relationship. Users can subscribe to Twitter feeds by ‘following’ them. Following a user on Twitter is not transitive in any way, following a user does not grant any reciprocal effect. This model is different to the more common system of symmetrical relationship where both sides must agree before the nodes become linked (Chen, 2009). This creates three different types of inter-node edges: B follows A; A follows C; B and C follow each other. A Twitter user receives a list of tweets from all followed Twitter streams in chronological order when they access their Twitter account. Information flows from the followed to the follower.

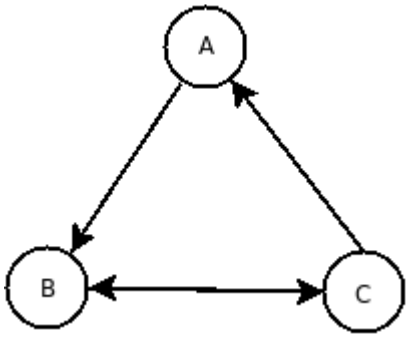

Figure 1: Different types of Twitter relationships, arrows denote the direction of information transfer

Twitter users can be categorised into three types: listeners, talkers, and hubs. Listeners have a low ratio of followers to those who follow, talkers have a high ratio of followers to following users and hubs have a follower to following ratio of approximately 1 (Iskold, 2008). Talkers are information producers using twitter to distribute information rather than collect it — often Talkers are celebrities or syndication sources. Listeners output little data and use Twitter as an information source, more interested in consuming the output of talkers than in producing their own content. Hubs represent the typical use case of Twitter where information is both consumed and produced in roughly equal quantities. Hubs are the most likely group to rebroadcast a tweet that they have received (‘retweet’). Hub users have the most symmetric edges in their local networks representing a close social network. Hub users tend towards dense graph networks with other hub users with only the occasional connection to a talker or listener.

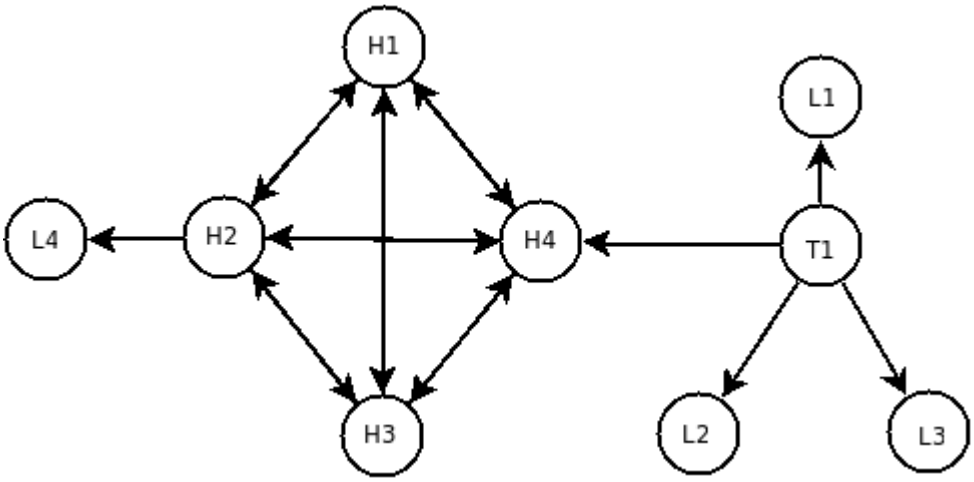

Figure 2: A Twitter network with hubs, listeners and talkers

Information propagates through Twitter networks by being retweeted by recipients. The rate at which a message propagates is the product of the likelihood of each user retweeting the message and the number of users receiving the message. The factors involved in a user's decision to retweet a message include the interest the user has in the content and the credibility of the message. Unlike Internet Relay Chat, Internet forums and email mailing lists there is no Twitter mechanism for sibling nodes to communicate unless those nodes have a pre-existing relationship. If a Twitter user tweets something false and one of the user's followers refutes the information there is no way for the tweeter's other followers to see the rebuttal without it being retweeted. This means that disinformation being spread via Twitter is robust as one user's rebuttal only produces a localised, non catastrophic effect. This allows disinformation to be optimised for the maximum chance to be retweeted rather than optimised for universal believability. In Figure 3 disinformation is shown propagating throughout a Twitter network, the black nodes represent users who have disbelieved or otherwise ignored the information. Figure 3 demonstrates that because the disbelieving nodes have no interaction with their sibling nodes their disbelief does not decrease the chance of the information being further propagated. Figure 3 assumes a 50% rate of disbelief and a 100% rate of retweeting among believers.

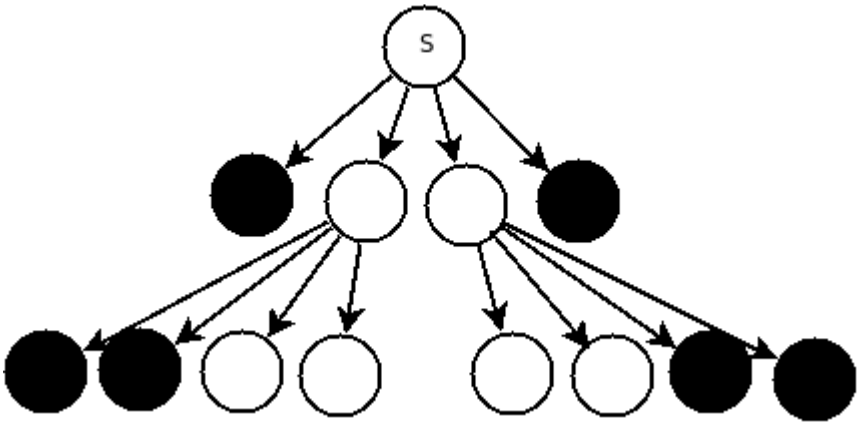

Figure 3: Disinformation spreading through a Twitter network

A Twitter disinformation case study

In March of 2009 Matthew Schneider, a journalist at the online liberal political news site Daily Kos, observed that misinformation about the US Economic Stimulus package was being repeated by "well over a dozen congressmen" and decided to perform an experiment "to test the limits of this phenomenon" (Schneider, 2009). Schneider created a Twitter account called 'InTheStimulus' and started by following all the Twitter users that had followed Republican Twitter feeds. With an initial audience of 1000 users from users reciprocating by following InTheStimulus back Schneider started posting disinformation about the US Economic Stimulus package. The messages all took the format of "InTheStimulus is $x million for _____". The initial messages started with plausible but false statements and progressively became more implausible until eventually the posted statements were blatantly false. Schneider found that users were willing to accept and retweet unsourced information even if that information could be verified or discredited from information in the public domain. Followers of hub users that retweeted the disinformation would often follow the InTheStimulus feed after receiving the retweet, showing an interest in receiving further messages of this type.

While some users requested a source citation from InTheStimulus none of these messages were available to the rest of the audience. As time passed and the disinformation became more blatant some Twitter users started rejecting the disinformation and unfollowing the InTheStimulus feed but the total audience of the feed kept growing from new followers. While Schneider's experiment was not conducted following any rigorous experimental protocol or with any serious objective in mind it still is a reasonable example of a Twitter disinformation campaign. Schneider managed a good retweet rate for the messages and very few negative responses from a position of no initial credibility. Schneider built credibility by starting with subtle disinformation and by choosing a receptive target audience. Schneider's campaign was still proceeding successfully a week after it started despite the extreme fabrications that were distributed later in the campaign and would probably have remained that way if he had not published a report on his experiment online.

Twitter as a favourable environment for disinformation

Twitter messages can seem credible without containing any references to support their claims. The short length of tweets encourages short declarative statements absent of supporting arguments and thus users do not become suspicious of unreferenced assertions. The fact that in some instances Twitter has been the primary source of news about a currently unfolding event also gives it some inherent credibility.

In January 2009 US Airways flight 1549 crash landed into the Hudson River New York and the first news reports and images of the incident were delivered via Twitter from eye witnesses — approximately 15 minutes ahead of any coverage by traditional media sources (Beaumont, 2009). Similarly, in February 2009 Twitter users near the scene of the crash of Turkish Airlines flight 1951 in Amsterdam provided the first public information about the accident (CNN, 2009). Twitter has the capacity to very rapidly disseminate information on unfolding events. Twitter was the primary source of information for the first six hours of the 2008 Mumbai attacks (Mishra, 2009). During this time the volume of Twitter traffic about Mumbai jumped from less than 10 posts an hour to approximately 1000 posts per hour and stayed around that level for the duration of the crisis. As a result of this type of activity tweets are considered credible in the absence of conflicting evidence. Twitter users accept the idea that tweets can represent newly discovered information and this can mean that an absence of confirming sources only reinforces the timeliness of the information rather than undermining the credibility of the information.

Sensational Twitter topics can even create their own feedback loops to sustain themselves. Sensational tweets have a high chance of being retweeted, which widens the audience to the point where the Twitter trends page will start reporting the information. Once a topic appears on the Twitter trends page it becomes visible to Twitter users that are not connected to the social network that originated the information, thus expanding the potential audience to the entirety of the Twitter user population.

This effect was visible during the 2009 H1N1 (Swine Flu) outbreak where from the 20th of April to the 24th of April the percentage of Twitter traffic that referred to H1N1 rose from nothing to 0.2% of all Twitter messages. At this point the information started trending and was visible to larger audiences and on the 25th of April almost 2% of all tweets were H1N1 related (Nielsen Online, 2009). Amongst the common H1N1 memes being tweeted were false rumours about the H1N1 transmission vectors — wrongly attributing the eating of pork as a vector, rumours about the current spread of the outbreak, and speculation as to the source of the outbreak (Morozov, 2009). While there is no evidence to support that any of the

misinformation spread during the H1N1 outbreak was deliberate it would not have been difficult for an external influence to have affected the information flow. Someone seeking to spread misinformation could have taken advantage of the feedback loop to engage in perception shaping operations, or even to encourage the feedback loop to increase the magnitude and prolong the length of the H1N1 Twitter trend.

Identity on Twitter is tied to accounts and therefore if an account is usurped the reputation of that identity and the access to the account's followers can be used to further a disinformation campaign. Several prominent Twitter feeds including those of President Barak Obama, CNN Anchor Rick Sanchez and Fox News have been hacked and tweets have been posted in their names (CNN, 2009). The breached accounts were used to post character damaging information or generally offensive messages and the breaches were quickly identified and fixed, however although the breaches appear to be simple pranks on the part of the hacker future breaches could be used more carefully as a tool for disinformation, borrowing the reputation of the compromised identity to propel subtly shaped information.

Twitter users without any special reputation can be useful to actors spreading disinformation because an average hub user's followers will often have personal relationships with the user and a large degree of trust in the user's messages. While posting disinformation to the small audience of an average hub user is not likely to be effective a coordinated campaign originating from hundreds of compromised accounts could single-handedly propel a piece of disinformation past the point where a feedback loop is created and the disinformation becomes self propelled.

Traditional hacking techniques such as the use of phishing emails or key loggers installed by botnets could be used to compromise the accounts needed for this attack. Additionally, at the AusCERT 2009 Information Security conference Chenette (2009) discussed an avenue through which security flaws in the Twitter API could be used to compromise a large number of Twitter accounts (Chennette, 2009). Chennette also outlined weaknesses in the Twitter API security model that could allow attackers to modify information that is being syndicated to Twitter — for example stock quotes or news information. Modification of syndicated data is as potent as the compromise of the syndication account itself as messages can be subtly altered or completely replaced to further an information operations campaign.

Twitter could also be a conduit for spreading disinformation via the mass media. CNN solicits content from users via its iReports program (CNN, 2009) and via Twitter replies to CNN Anchors' Twitter accounts and these messages are often read out live on the program (Hirsch, 2008). While CNN disclaims responsibility for the correctness of user submitted content creditability is lent to information that is read on air by an anchor on a prominent news network. This channel would probably not be useful for blatant disinformation as tweets are likely filtered before they are read however the opportunity still exists to use Twitter to borrow the reputation of a news network to spread disinformation.

Current situation

Unwary users are at risk from being affected by Twitter borne disinformation. Politicians, political activists and corporations are already using Twitter as a resource to influence opinion (Miller C. C., 2009). During the 2009 German Presidential election the results were leaked by members of two major political parties to Twitter before an official announcement was made (Telegraph, 2009). The threat to people from Twitter disinformation during the 2009 H1N1 outbreak has prompted the creation of guides to help users recognise and

disregard Twitter disinformation and urging users to verify information before retweeting it because “false rumors can cost lives” (Sitaker, 2009). The proliferation of disinformation capabilities represented by Twitter will almost guarantee that users of social networks will be exposed to disinformation and if users do not develop rigorous information sanitation habits they will be manipulated by any organisation that cares to develop an information operations capability.

Conclusion

Twitter is a powerful tool for the dissemination of information and is an equally powerful tool for disinformation operations. Twitter is especially suitable for use in disinformation operations due to the casual nature of the communication and the asymmetrical structure of Twitter networks. Twitter disinformation operations require negligible resources and are an option available to organisations of all sizes. The proliferation of information operation capabilities inherent in the accessibility of online social media will lead to a larger risk of tainted information being assimilated into an organisation's information space. These disinformation campaigns can take advantage of existing trending topics and borrow the reputations of other users through identity theft as a force multiplier, making their campaigns more effective. Constituencies of social networks must be aware of the dangers of disinformation over social media and develop information assurance strategies to avoid being contaminated by disinformation.

References

Beaumont, C. (2009, 01 16). *New York plane crash: Twitter breaks the news, again.* Retrieved 05 24, 2009, from The Daily Telegraph: http://www.telegraph.co.uk/scienceandtechnology/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html

Chen, A. (2009, 03 16). *Friends versus Followers: Twitter's elegant design for grouping contacts*. Retrieved 05 19, 2009, from Futuristic Play: http://andrewchenblog.com/2009/03/16/friends-versus-followers-twitters-elegant-design-for-grouping-contacts/

Chennette, S. (2009, 05 22). *AusCERT 2009 - P0wning the Programmable Web*. Retrieved 05 25, 2009, from Security Labs Blog: http://securitylabs.websense.com/content/Blogs/3402.aspx

CNN. (2009). *iReports*. Retrieved 05 29, 2009, from http://www.ireport.com

CNN. (2009, 01 06). *Twitter accounts of Obama, Britney Spears hacked*. Retrieved 05 25, 2009, from CNN.com: http://www.cnn.com/2009/TECH/01/05/twitter.hacked/index.html

CNN. (2009, 02 26). *Twitter first to publish dramatic crash pictures*. Retrieved 05 24, 2009, from CNN.com: http://www.cnn.com/2009/WORLD/europe/02/25/twitter.amsterdam.plane.crash/

Hirsch, A. (2008, 09 04). *CNN Heavily Promoting Twitter On Air, Making Big Moves in Social Media*. Retrieved 05 25, 2009, from Mashable: The Social Media Guide: http://mashable.com/2008/09/04/cnn-twitter/

Iskold, A. (2008, 03 19). *5 Ways To Have Fun With Twitter When You're Bored*. Retrieved 05 23, 2009, from Read/Write Web: http://www.readwriteweb.com/archives/5_ways_to_have_fun_with_twitter.php

Miller, C. C. (2009, 04 13). *Finding Utility in the Jumble of Tweeted Thoughts*. Retrieved 05 20, 2009, from NYTimes: http://www.nytimes.com/2009/04/14/technology/internet/14twitter.html

Miller, J. L. (2009, 05 11). *Bit.ly Switch Part of Twitter's Realtime Search Strategy*. Retrieved 05 20, 2009, from WebProNews: http://www.webpronews.com/topnews/2009/05/07/bitly-switch-part-of-twitters-realtime-search-strategy

Mishra, G. (2009, 11 28). *Social Media & Citizen Journalism in the 11/26 Mumbai Terror Attacks: A Case Study*. Retrieved 05 20, 2009, from Gauravonomics Blog: http://www.gauravonomics.com/blog/social-media-citizen-journalism-in-the-1126-mumbai-terror-attacks-a-case-study/

Morozov, E. (2009, 04 25). *Swine flu: Twitter's power to misinform*. Retrieved 05 20, 2009, from Foreign Policy: Net Effect: http://neteffect.foreignpolicy.com/posts/2009/04/25/swine_flu_twitters_power_to_misinform

Nielsen Online. (2009, 04 27). *Swine Flu News and Concern Dominates Online Buzz*. Retrieved 05 26, 2009, from Nielsen Wire: http://blog.nielsen.com/nielsenwire/online_mobile/swine-flu-news-and-concern-dominates-online-buzz/

ProgrammableWeb. (2009). *Twitter API Profile*. Retrieved 05 22, 2009, from ProgrammableWeb: http://www.programmableweb.com/api/twitter

Reed, N. (2009, 05). *Counting the number of twitter messages*. Retrieved 05 10, 2009, from GigaTweet: http://popacular.com/gigatweet/analytics.php

Schneider, M. ". (2009, 03 26). *Twitter + Stimulus = Conservative Stupidity*. Retrieved 05 09, 2009, from Daily Kos: http://www.dailykos.com/story/2009/3/26/713407/-Twitter-+-StimulusConservative-Stupidity

Sitaker, K. J. (2009, 04). *How False Rumors Can Cost Lives*. Retrieved 05 30, 2009, from http://canonical.org/~kragen/costs-lives.html

Telegraph. (2009, 05 29). *Germany to investigate leak of election result on Twitter*. Retrieved 05 30, 2009, from The Daily Telegraph: http://www.telegraph.co.uk/scienceandtechnology/technology/twitter/5397150/Germany-to-investigate-leak-of-election-result-on-Twitter.html

Twitter. (2009). *Twitter Search*. Retrieved 05 20, 2009, from http://search.twitter.com

Twitter. (2009). *Twitter: What are you doing?* Retrieved 05 10, 2009, from Twitter: http://www.twitter.com

Williams, D. (2009, 04 21). *Twitter API Wiki*. Retrieved 05 20, 2009, from http://apiwiki.twitter.com/