Association for Information Systems
AIS Electronic Library (AISeL)

Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics

Special Interest Group on Decision Support and Analytics (SIGDSA)

12-11-2016

Twitter Sentiment Analysis: An Examination of Cybersecurity Attitudes and Behavior

Babita Gupta
California State University - Monterey Bay, bgupta@csumb.edu

Shwadhin Sharma
California State University - Monterey Bay, ssharma@csumb.edu

Anitha Chennamaneni
Texas A&M University Central Texas, anitha.chennamaneni@tamuct.edu

Twitter Sentiment Analysis: An Examination of Cybersecurity Attitudes and Behavior

Research-in-Progress

Babita Gupta
California State University Monterey Bay
Seaside, CA 93955 USA
bgupta@csumb.edu

Shwadhin Sharma
California State University Monterey Bay
Seaside, CA 93955 USA
ssharma@csumb.edu

Anitha Chennamaneni
Texas A & M University Central Texas
Killeen, TX 76549, USA
anitha.chennamaneni@tamuct.edu

Abstract

This exploratory study examines the cybersecurity attitudes and actual behavior over time using the data collected on the social media microblogging platform, Twitter. We plan to use the sentiment analysis and text mining techniques on original tweets related to cybersecurity collected at two different time periods. Upon completion of this research, we would present the analysis of the relationship between the cybersecurity attitudes and behavior and how behaviors may be shaped by the attitudes. This research work aims to contribute to the extant literature in cybersecurity and endeavors to enhance our understanding of cybersecurity attitude and behavior by validating the proposed research model and hypotheses by using real-time, user-generated, social media data.

Introduction

Social media platforms such as Facebook, Instagram, Twitter, Google+, and Pinterest are increasingly becoming an important source of data for understanding consumers' concerns using sentiment analysis. Sentiment analysis refers to the use of text with unstructured data containing individuals' opinions to understand their overall sentiment towards an aspect. Sentiment analysis (or opinion mining) is the study of users' sentiments towards certain entities, such as product review or cybersecurity (Fang and Zhan 2015). Although sentiment analysis has been applied to many different domains, such as tourism (García et al. 2012), politics (Hu et al. 2013), marketing (Mittal and Goel 2012) and epidemiology (Sadilek et al. 2012) to examine individuals' sentiments or opinions about certain events, use of sentiment analysis in digital privacy and security related domain is still nascent (Jurek et al. 2015). At the time of this study, we could not locate a research study that uses social media data to study cybersecurity attitudes and behavior applying sentiment analysis and text mining techniques.

High-profile data breaches at companies, such as Yahoo, Home Depot, Sony Pictures and Target, have brought a lot of attention to cybersecurity issues and the long-term impact of security violations. The hacking of celebrities' accounts in cloud in 2014-2015 further elevated the public discourse of cybersecurity violations and brought greater discussion about cybersecurity in the social media. While it is difficult to accurately estimate the costs associated with cyber security violations (as companies tend not to disclose the resulting financial losses), the estimated costs can easily reach billions of dollars per year (Gordon et al. 2015). The Center of Strategic and International Studies (2014) estimates that in 2014 the annual cost of cybercrime to the global economy exceeded $375 billion. Thus, cybersecurity and individuals' attitude to

cybersecurity and how they might be protecting themselves in digital transactions is an important area of study.

We draw upon the literature of online privacy and security attitudes and behavior as the theoretical foundation for this study. We plan to collect twitter data from individuals at two different time periods. We then analyze these two sets of time-variant data using sentiment analysis and text mining techniques to understand the relationship between cybersecurity attitude and behavior at these two time periods and how the cybersecurity attitudes may be shifting over time. We use a hybrid approach of linguistic and statistical techniques applied at sentence level text (tweet) and feature extraction level. We classify expressed sentiments towards cybersecurity as positive, negative, or as neutral (apathetic). We also plan to analyze these tweets from these two time periods using text-mining techniques to examine the actual behavior. We use regression analysis techniques to study how attitudes (sentiment) towards cybersecurity relate to individuals' actual behavior as evidenced by their engagement in online financial transactions.

This research work makes two contributions to the extant body of knowledge. First, we endeavor to enhance our understanding of cybersecurity behaviors by validating the proposed research model and hypotheses by using real-time, user-generated, social media data. Second, we plan to gather tweets that pertain to cybersecurity behaviors and analyze it to understand the actual cybersecurity behavior.

Literature Review

Cybersecurity

We define cybersecurity using the information and communications technology (ICT) security principles of confidentiality, authentication, integrity of information, non-repudiation, access control and availability (Von Solms and Van Niekerk 2013). Von Solms and van Niekerk (2013) argue that these ICT security principles apply to the cyber space as well where the individuals are concerned about the security of the data they share online in various interactions and also about the ability of the companies to keep their information secure.

Information Systems (IS) research studies have generally focused on behavioral intention rather than the actual behavior using survey instruments (Ifinedo 2012; Johnston and Warkentin 2010; Safa et al. 2016; Sharma and Crossler 2014; Shropshire et al. 2015). IS research measuring IS behavior based on behavioral intention rather than the actual behavior provides a practical way for researchers to study behavior in IS research (Torres et al. 2014; Xu et al. 2012; Xu et al. 2010). Several recent privacy research studies in IS literature have mentioned the limitations of intentions predicting the actual behavior.

Kehr et al. (2015) found that situation-specific assessment of risks and benefits fully mediates the effect of dispositional factors such as attitude on information disclosure and that attitude affects the behavior. Summary of research work in Table 1 highlights the *privacy paradox,* which refers to the discrepancy between the stated privacy risk beliefs, attitudes and/or intentions and the actual behaviors.

Sources	Descriptions
Boss et al. (2015)	Performed an extensive review of protection motivation theory in information security research and examined factors that dissuade non-compliance behaviors
Acquisti et al. (2015)	Examines several reasons of why privacy attitude, intention, and actual behaviors can be different.
Beresford et al. (2012)	Conducted a field experiment, in which subjects bought DVD from the cheaper store (cheaper by 1 Euro) despite the cheaper store asking these individuals sensitive personal information. Survey showed a strong intention to protect information.
Spiekermann et al. (2001)	Found very little evidence of a correlation between intention and behavior when measured the privacy preferences of 171 users and observed their behavior on a mock e-commerce site.
Hughes-Roberts (2013)	Found that user concerns and intention is not a valid indicator of privacy behavior based on comparing questionnaire survey and an examination of participants' Facebook profiles.

Joinson et al. (2012)	Found no association between dispositional privacy attitudes of 759 samples to their actual behavior, suggesting that being concerned about privacy does not influence how a specific privacy-related situation is viewed.
Metzger (2006)	Found no association between people's privacy concerns and their disclosure to an e-commerce site.
Norberg et al. (2007)	There is always a paradox between personal information disclosure intentions and behaviors. Privacy attitudes which are defined broadly, and behaviors which are defined narrowly, should not be expected to be closely related.
Taddicken (2014)	Showed that privacy concerns hardly impact self-disclosure as the study stated that the relationship between the two constructs is moderated by various other variables.
Tufekci (2008)	Found little to no relationship between online privacy concerns and information disclosure in Online Social Network.
Zafeiropoulou et al. (2013)	Examined location data disclosed by individuals and found evidence that supports the existence of privacy paradox for location based applications.

Table 1: IS Research about the Discrepancy between Intention and Behavior

Sentiment Analysis and Text Mining

For examining the cybersecurity attitudes and behavior, we intend to apply sentiment analysis and text mining techniques to gauge people's cybersecurity actions based on what they say in their texts. Text mining is the process of deriving high quality information from natural language text through the application of algorithms and techniques from data mining, statistics, machine, retrieval and knowledge management (Feldman and Sanger 2007). Sentiment analysis and opinion mining are two prominent fields in text mining.

Sentiment analysis is concerned with detecting the sentiment polarity as positive, negative or neutral (apathetic) in a written text that contains an opinion to extract the attitude of the author of the text about a specific topic or theme (Stieglitz and Dang-Xuan, 2013). An opinion in a user generated text is a quintuple (*ei, aij , sijkl, hk, tl*) where *ei* is the name of an entity, *aij* is an aspect of *ei*, *sijkl* is the sentiment on aspect *aij* of entity *ei* , *hk* is the opinion holder, and *tl* is the time when the opinion is expressed by *hk* (Giachanou and Crestani 2016). An opinion quintuple for our research purposes based on Giachanou and Crestani (2016) and Liu (2012) is defined as follows: (*cybersecurity, online shopping, positive, twitter account holder, time*). Sentiment analysis is being used as a way to explore the opinions and emotions and their effects on behaviors (Makarem and Jae 2015; Stieglitz and Dang-Xuan 2013). Opinions play a very important role in decision making process and in influencing our behaviors. Individual's opinions and decision-making are also influenced by opinions of thought leaders and by opinions of significant others.

Sentiment analysis and opinion mining are often used interchangeably to mean the same thing. However, some researchers contend that sentiment analysis and opinion mining address slightly different problems (Tsytsarau and Palpanas 2012). While opinion mining aims to automatically identify and extract people's opinions, attitudes, and emotions towards individuals, entities, and events from the user generated content, sentiment analysis classifies the opinions into subjective categories: positive, negative or neutral (apathetic) and measures sentiment polarity. Pang & Lee (2008) more precisely define sentiment analysis or opinion mining as the "computational study of opinions, feelings and subjectivity in text".

Unlike traditional quantitative or qualitative methods such as surveys or focus groups, which are expensive, time consuming and labor intensive, sentiment analysis methods are cost-effective, non-intrusive and extract opinions or sentiments in real time by applying automatic algorithms to sort through textual data (Chiu et al. 2015; Pang and Lee 2008; Thelwall et al. 2011). Additionally, sentiment analysis methods are not subject to recall biases typically associated with the self-related measurements (Rylander et al. 1995). Wang et al. (2013a) reviewed prior research on sentiment analysis usage, and reported that the findings have a rather high accuracy rate. Sentiment analysis has been successfully applied in various contexts such as predicting financial market performance (Bollen et al. 2011; Xu et al. 2012; Yu and Hatzivassiloglou 2003), providing early warnings (Fu et al. 2012), determining level of happiness (Dodds et al. 2011), predicting movie revenues (Asur and Huberman 2010), understanding online consumer reviews and

opinions (Chiu et al. 2015; Chung 2009; Duan et al. 2013; Yang et al. 2010), tourism (García et al. 2012), politics (Hu et al. 2013), marketing (Mittal and Goel 2012), and epidemiology (Sadilek et al. 2012).

Sentiment analysis can be performed at three levels: document level, sentence level, and aspect or feature level. Document level sentiment analysis treats the entire document as a single unit and classifies the sentiment expressed in the document as positive, negative or neutral. Sentence level sentiment analysis classifies the sentiment expressed at the sentence level. Liu (2012) contends that there is not much difference between document level and sentence level classifications as sentences are merely short documents. Aspect or feature level sentiment analysis classifies the sentiments with respect to the specific entities and their aspects. It is the most fine-grained analysis as the individuals can have diverse opinions for different aspects of the same entity. Sentiment analysis was applied at the document level by Pang et al. (2002), Pang and Lee (2008), and Turney (2002), at the sentence level by Yu and Hatzivassiloglou (2003), and at the aspect level by Singh et al. (2013) and Wang et al. (2013b).

Sentiment analysis classification methods can be grouped into three approaches: machine learning approach, lexicon based approach, and hybrid approach (Maynard and Funk 2011). Machine learning approach uses both supervised and unsupervised learning methods to predict the polarity of the sentiments. The lexicon-based approach uses a list of words that are pre-coded for polarity to identify sentiments (Taboada et al. 2011). The hybrid approach combines both the machine learning and the lexicon based approaches to detect sentiment polarity. Makarem and Jae (2015) studied the consumer boycott behavior by doing a qualitative study of 1,422 tweets using content analysis for analyzing motives, cause and target of boycott, and using sentiment analysis for identifying high intensity tweets about boycott behavior. Stieglitz and Dang-Xuan (2013) reported that the affective dimensions of the sentiments in a tweet were significantly associated with the information sharing behavior in the context of political communication.

Research Methodology

Microblogging with Twitter

Twitter Inc. is an online social networking service that enables Twitter account holders to send and receive 140-character messages called "tweets", follow other users, form communities around a trending topic (#hashtag), and forward tweets to others in their circle of followers. In June 2016, Twitter averaged about 313 million monthly active individuals[1].

We plan to use the microblogging online social network platform Twitter to collect about 15,000 original tweets related to cybersecurity. We chose Twitter because the extant literature posits that web 2.0 platforms, such as the Twitter, allow individuals to generate free flow content in a naturalistic setting and are better suited to discern undiscovered dynamic insights beyond what is discovered in survey-based research questionnaire that focus on numerical or qualitative responses (Ghose and Ipeirotis 2009; Makarem and Jae 2015). Twitter users' posts are generated in real-time with high level of anonymity and therefore more likely to be free of biases identified by Peterson and Wilson (1992). The researchers contend that self-reported measurements associated with survey questionnaires are subject to recall and question framing bias (Peterson and Wilson, 1992).

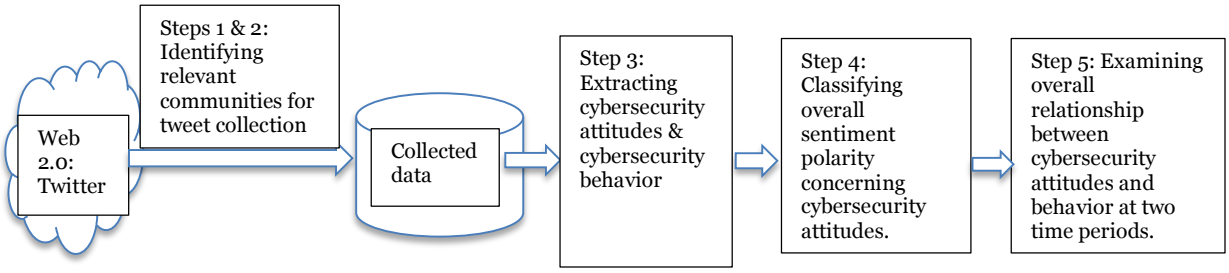

Figure 1: Cybersecurity Research Framework

[1] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

We collect the tweets at two different time periods to examine the cybersecurity attitudes and actual behavior over time. This study uses the mixed-method approach of analysis – sentiment analysis to examine the cybersecurity attitude followed by the qualitative text analysis to identify cybersecurity behavior (Makarem and Jae 2015). We discuss below the overall research framework (see Figure 1) for this study using sentiment analysis and text mining using the data generated on Twitter (Glass and Colbaugh 2011; Liu, 2012).

Steps 1 & 2: Identifying relevant communities for tweet collection

Tweets are small documents of length that are at most 140 characters. Giachanou and Crestani (2016) identify the following tweet characteristics: tweet, user, mentions, replies, followers, retweets, hashtag, privacy, and URLs. Twitter user profile information consists of their user name, user description, gender (if public), and location (if public). Glass and Colbaugh (2011) recommend that for understanding the sentiment of a population related to an issue, it is important to identify the communities in this issue domain that consist of people who are considered influential (represented by users' that have their tweets retweeted most often) and authoritative in the domain.

In order for us to gauge the attitude of the population of twitter users towards cybersecurity, we have identified the communities and themes represented by #Hashtags in twitter related to the topic of cybersecurity using the six defining ICT security principles of confidentiality, authentication, integrity of information, non-repudiation, access control, and availability. Some examples of these #hashtags are: #security, #cybersecurity, #cybercrime, #privacy, #datasecurity, #datatheft, #cybersec, #cyber, #mcgsecure, #IDtheft, #identityprotection, #Passwords, #databreach, #Identitytheft, #surveillance, #infosec, #Breach, #DataBreach, #securitybreach, #Malware, #Hacker, #Cybercrime, #Fraud, #targetdatabreach. These communities share some common attributes, such as the number of original tweets (number of distinct individuals) and the number of retweets (an original tweet shared by other individuals with their followers). These communities carry temporal information such as the time of creation of the hashtag and the number of tweets within this community.

We plan to collect original tweets from two time periods to allow us to examine the relationship between cybersecurity attitudes and cybersecurity behavior. The first time period, T=1, is the six-month period starting from March 1, 2013 to August 31, 2013. The second time period, T= 2, is from March 1 to August 31, 2015. We chose this time frame as several data violations were reported in media including the Target, Sony, healthcare insurance company Anthem, and Home Depot after the first data collection time period ended[2]. Collecting tweets over a six-month period for each time period would minimize bias (Makarem and Jae, 2015). This gap between the two time periods would also allow us to understand if the cybersecurity related violations reported publicly had any effect on the cybersecurity attitudes and behaviors.

After the data collection at the two time periods, we would be filtering out the tweets that may be associated with the companies, bloggers, news channels, and bots. This would enable us to collect tweets representing individuals and thus to examine sentiments and behavior of individuals tweeting about cybersecurity issues.

Step 3: Extracting cybersecurity attitudes & cybersecurity behavior

Step 3a: In order to extract attitude about cybersecurity, we extract the sentiment polarity of tweets collected at T=1 and T=2. Sentiment analysis literature indicates that sentiments generated by individuals in online social networks is a good way to gauge their attitude towards a particular topic, concept or activity (Ceron et al. 2015; Makarem and Jae 2015). We plan to collect tweets in communities (# hashtags) identified with the cybersecurity theme for T=1 and T=2. We plan to have sample size of at least 15,000 original tweets for each of the two time periods at T= 1 and T= 2. In order to measure the sentiment towards cybersecurity expressed by an individual in her/his tweets in communities of hashtags identified earlier, we use following as examples of individual words and phrases:

- For sentiments expressing positive attitude about cybersecurity, words and phrases, such as "smart", "safety", "great", "love", "glad", "I think shopping at X is pretty safe", "awesome experience at bank", "happy to see high security" expressing positive connotations are relevant.

[2] http://www.nytimes.com/interactive/2015/02/05/technology/recent-cyberattacks.html?_r=0

- For sentiment expressing negative attitude about cybersecurity words and phrases, such as "fear", "vulnerable", "fraud", "victim", "dangerous", "online shopping sucks!", "hate giving my info", "disgusted with banking with X", "dislike", "awful" expressing negative connotations are relevant.
- For sentiment expressing neutral attitude about cybersecurity words and phrases, such as "unconcerned", "compliance", "Govt wants to secure data" or "company site mentioned security policy", "forwarding the URL about data safety" expressing no emotion but are simply informational in nature.

Step 3b: We also collect data about the cybersecurity behavior at T=1 and T=2. We do this by applying text mining techniques such as the feature extraction using keywords as a bag of words using lexicon-based dictionary to learn about the online financial transactions expressed in their tweets. Cybersecurity behavior is indicated if the tweets mention online (e-commerce and m-commerce) transactions such as shopping, buying, banking, currency transfers, use of PayPal or bitcoin, paying bills; use phrases, such as installing firewalls, deleting information, not sharing information, encryption, and secure; or tweets disclose personal information, such as location of tweets, home address, health profile, medications, doctors visited, and income.

Step 4: Classifying overall sentiment polarity concerning the cybersecurity attitudes

Sentiment analysis algorithms based on the automated language processing models (Feldman 2013; Feldman and Sanger 2007) are available in the field. We use IBM Watson's Insights for Twitter data analysis. This tool uses the hybrid approach to natural language processing incorporating both the linguistic and statistical analysis techniques that are suitable for texts that are noisy, such as the tweets (Rizzo and Troncy 2011; Saif et al. 2012). It allows us to observe the sentiments at the sentence level (i.e. the tweet), at the community level (i.e. the hashtag), and at the overall level aggregating the polarity of sentiment. IBM Watson's Insight mines opinion words, their semantic orientation, and measures sentiment strength as ambivalent, negative, neutral, positive, and unknown.

Step 5 (Research Model): Examining overall relationship between cybersecurity attitudes and behavior at T=1 and T=2

At T= 1 and T= 2, we collect data using hashtags that relate to cybersecurity attitudes and the actual online behaviors indicated by the tweets mentioning online financial transactions. We use the research model in figure 2 to examine the relationship between cybersecurity attitude gleaned through the sentiment analysis and how it may be related to the actual behavior – are individuals engaging in more, less, or about the same amount of financial transactions online? Our research model in figure 2 draws upon the prior literature on cybersecurity behavior studies (Boss et al. 2015; Foltz et al. 2010).

We expect that positive sentiment about cybersecurity would result in individuals engaging in more financial transactions online and sharing sensitive information, such as the credit card information, with the companies online. We expect individuals with negative cybersecurity attitudes to be more prudent about conducting online financial transactions while individuals with neutral cybersecurity attitude would not exhibit any change in their cybersecurity behavior. Therefore, we derive the following hypotheses:

H1: Positive attitude about cybersecurity at T=1 would be related to higher number of online financial transactions at T =2 compared to the number of online financial transactions at T=1.

H2: Negative attitude about cybersecurity at T=1 would be related to lower number of online financial transactions at T =2 compared to the number of online financial transactions at T=1.

H3: Neutral attitude about cybersecurity at T=1 would be related to same number of online financial transactions at T =2 compared to the number of online financial transactions at T=1.

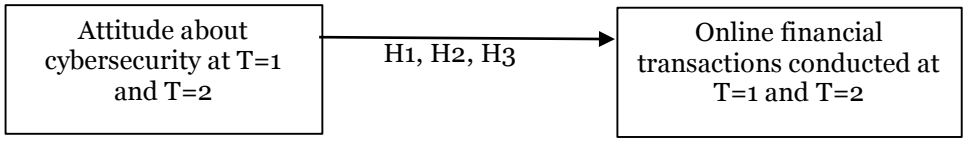

Figure 2: Research model to examine relationship between cybersecurity attitude and behavior over time

In addition, since we examine the cybersecurity attitudes at both the time periods, it would allow us to study how cybersecurity attitudes may have shifted over the period of two years along with changes in how individuals behave with respect to the cybersecurity protective behavior. Any insights gained may help with future research directions.

Data Analysis

We plan to test the hypotheses by applying the regression analysis techniques, similar to the analysis methodology reported by Stieglitz and Dang-Xuan (2013). We aim to use the variable associated with sentiment of a tweet to examine its relationship to how often the online financial transactions were conducted as the dependent variable. We also include control variables, such as number of hashtags in a tweet, individuals' number of followers, and the number of tweets posted by the individual.

Pilot study and results

We conducted a pilot study to test our research methodology. We collected twitter data from the following communities from August 1 to August 3, 2016 (i.e. t=1) and from September 27 to September 29, 2016 (i.e. t=2): #security, #cybersecurity, #cybercrime, #privacy, #datasecurity, #datatheft, #cybersec, #cyber, #mcgsecure, #IDtheft, #identityprotection, #Passwords, #databreach, #Identitytheft, #surveillance, #infosec, #Breach, #DataBreach, #securitybreach, #Malware, #Hacker, #Cybercrime, #Fraud, #targetdatabreach.

Out of the thousands of the tweets collected, we filtered out the tweets that were associated with the companies, bloggers, news channels, and bots to examine the real sentiment of the real twitter users. We finally collected 120 tweets for t=1 and 72 tweets for t=2. Of all the tweets we analyzed, there were total of 36 retweets and with 2 direct messages (tweet that starts with an @sign). The location/s of the tweets were not analyzed for this pilot study (due to the small "*n*" in this pilot study).

For t=1, some of the words that conveyed positive sentiments were: Most important, health, godsend, #smartcities, insights, smart; some of the words that conveyed negative sentiments were: Brawlers, Flood, danger, breach, #hacked, stolen, breached, smear, and infected; and some of the words that conveyed neutral sentiment were: Security, resilience and #efficiency. The aggregate sentiment for t=1 was neutral with overall sentiment score of +0.019. This implies that the tweets in t=1 did not have positive or negative attitude towards cybersecurity in general.

For t=2, some of the words that conveyed positive sentiments were: Safety, famous, great, medical information, too late, security, efficient, succeed, ethical, famous, and great; some of the words that conveyed negative sentiments were: Fear, last thing, lawsuits, lawsuit, vulnerabilities, critical, victims, ransom, deceive, and in danger; and some of the words that conveyed neutral meaning were: #Compliance, #healthcare, major, and huge. The aggregate sentiment for t=2 was a negative sentiment with overall sentiment score of -0.129. This implies that the tweets in t=2 had negative attitude towards cybersecurity in general.

Implication of the pilot study are that moving forward with the full study, we would need to collect data over a longer time frame for each time period and to also include more cybersecurity relevant communities.

Conclusion, Contributions, and Limitations

This exploratory study presents a research framework to examine the cyber security attitude and behavior and their relationship by collecting the Twitter data. While this study is exploratory in nature, we continue to collect data and expect to report the findings of this completed study and the effectiveness of the research framework by summer of 2017.

The completed research aims to contribute to the body of literature in cybersecurity in two ways. First, we endeavor to enhance our understanding of cybersecurity behaviors by validating the proposed research model and hypotheses by using real-time, user-generated, social media data. Much of the extant online privacy and security literature has focused on understanding individual's attitude towards cybersecurity using survey-based methodology. Survey based methods are subject to recall and question framing biases, typically associated with quantitative and qualitative questionnaires and self-reported measurements (Peterson and Wilson 1992; Rylander et al. 1995). Second, in this study, we plan to gather tweets that pertain to cybersecurity behaviors and analyze it to understand the actual behavior about protecting the

cybersecurity. Most studies in online privacy and security literature use intention as a proxy for actual behavior of the individual (see Table 1) and rely on either asking individuals about their intention to engage in a particular set of behavior in future or rely on individuals' past behavior as an indicator of the future behavior. Thus this work may provide an innovative avenue for researchers engaged in research about consumer intention and their behavior.

As web 2.0 & web 3.0 become important avenues for social communication, this study can provide a gauge for measuring the extent to which an average consumer is really concerned about the cybersecurity issues and if that concern impacts their online behavior. These findings may help cybersecurity researchers in planning and executing social media strategies to increase the consumers' awareness of the cybersecurity issues. It is important to recognize that there are some inherent limitations in using social media data such as the tweets. Sentiment analysis is limited to the data that can be downloaded and what the individuals are willing to share. Also, sentiment detection may not work as well for tweets that contain subjective words or do not contain explicitly recognized topic related words. In addition, tweets may contain opinions expressed in complex linguistic ways, such as sarcasm, irony, and implication (Feldman 2013). Sentiment analysis methods have been criticized in the past for being limited in terms of detecting sarcasm (Feldman and Sanger 2007). Additionally, sentiment analysis can be domain and event dependent. Terms and sentences considered as positive in one domain may not be so in another domain.

References

Acquisti, A., Taylor C. R. and Wagman L. *"The economics of privacy." Journal of Economic Literature* (52:2), pp. 1-64.

Asur, S., and Huberman, B. A. 2010. "Predicting the Future with Social Media," *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*: IEEE, pp. 492-499.

Beresford, A. R., Kübler, D., & Preibusch, S. 2012. "Unwillingness to pay for privacy: A field experiment," *Economics Letters*, (*117*:1), pp. 25-27.

Bollen, J., Mao, H., and Zeng, X. 2011. "Twitter Mood Predicts the Stock Market," *Journal of Computational Science* (2:1), pp. 1-8.

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. 2015. "What Do Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Security Behaviors," *MIS Quarterly (MISQ)*,(39:4), pp. 837-864.

Center for Strategic and International Studies (CSIS). 2014. "Net losses: Estimating the global cost of cybercrime," *CSIS and McAfee Intel Security,* pp. 1-23.

Ceron, A., Curini, L., and Iacus, S. M. 2015. "Using Sentiment Analysis to Monitor Electoral Campaigns Method Matters—Evidence from the United States and Italy," *Social Science Computer Review* (33:1), pp. 3-20.

Chiu, C., Chiu, N.-H., Sung, R.-J., and Hsieh, P.-Y. 2015. "Opinion Mining of Hotel Customer-Generated Contents in Chinese Weblogs," *Current Issues in Tourism* (18:5), pp. 477-495.

Chung, W. 2009. "Automatic Summarization of Customer Reviews: An Integrated Approach," In *Proceedings of 15th* Americas Conference on Information Systems, pp. 1-7.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. 2011. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter," *PloS one* (6:12), p. e26752.

Duan, W., Cao, Q., Yu, Y., and Levy, S. 2013. "Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality," *46th Hawaii International Conference on System Sciences (HICSS)*, pp. 3119-3128.

Fang, X., and Zhan, J. 2015. "Sentiment Analysis Using Product Review Data," *Journal of Big Data* (2:5), pp. 1-14.

Feldman, R., and Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

Foltz, C. B., Newkirk, H. E., & Schwager, P. H. 2016. "An Empirical Investigation of Factors that Influence Individual Behavior toward Changing Social Networking Security Settings," *Journal of Theoretical and Applied Electronic Commerce Research*, (*11*:2), pp. 1.

Fu, T., Abbasi, A., Zeng, D., and Chen, H. 2012. "Sentimental Spidering: Leveraging Opinion Information in Focused Crawlers," *ACM Transactions on Information Systems (TOIS)* (30:4), p. 24.

García, A., Gaines, S., and Linaza, M. T. 2012. "A Lexicon Based Sentiment Analysis Retrieval System for Tourism Domain," *Expert Syst Appl Int J* (39:10), pp. 9166-9180.

Ghose, A., and Ipeirotis, P. 2009. "The Economining Project at Nyu: Studying the Economic Value of User-Generated Content on the Internet," *Journal of Revenue & Pricing Management* (8:2), pp. 241-246.

Giachanou, A., and Crestani, F. 2016. "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys (CSUR)* (49:2), p. 1-41.

Glass, K., and Colbaugh, R. 2011. "Web Analytics for Security Informatics," *Intelligence and Security Informatics Conference (EISIC), 2011 European*: IEEE, pp. 214-219.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Zhou, L. (2015). The impact of information sharing on cybersecurity underinvestment: a real options perspective. *Journal of Accounting and Public Policy*, *34*(5), pp. 509-519.

Hu, Y., Wang, F., and Kambhampati, S. 2013. "Listening to the Crowd: Automated Analysis of Events Via Aggregated Twitter Sentiment," In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence,* pp. 2640-2646.

Hughes-Roberts, T. 2013. "Privacy and Social Networks: Is Concern a Valid Indicator of Intention and Behaviour?,". In *2013 International Conference on Social Computing (SocialCom),* pp. 909-912.

Ifinedo, P. 2012. "Factors influencing e-government maturity in transition economies and developing countries: a longitudinal perspective." *ACM SigMIS Database* (42:4), pp. 98-116.

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., ... & Wansink, B. 2012. "Beyond nudges: Tools of a choice architecture," *Marketing Letters*, (23:2), 487-504.

Johnston, A. C., and Warkentin, A. 2010. "Fear appeals and information security behaviors: an empirical study." *MIS Quarterly*, (34:3), pp. 549-566.

Jurek, A., Mulvenna, M. D., and Bi, Y. 2015. "Improved Lexicon-Based Sentiment Analysis for Social Media Analytics," *Security Informatics* (4:1), pp. 1-13.

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. 2015. "Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus," *Information Systems Journal*, (25:6), pp. 607-635.

Liu, B. 2012. "Sentiment Analysis and Opinion Mining," *Synthesis lectures on human language technologies* (5:1), pp. 1-167.

Makarem, S. C., and Jae, H. 2015. "Consumer Boycott Behavior: An Exploratory Analysis of Twitter Feeds," *Journal of Consumer Affairs*, (50:1), pp. 193-223.

Maynard, D., and Funk, A. 2011. "Automatic Detection of Political Opinions in Tweets," *Extended Semantic Web Conference*: Springer, pp. 88-99.

Mittal, A., and Goel, A. 2012. "Stock Prediction Using Twitter Sentiment Analysis," *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*).

Norberg, P. A., Horne, D. R., & Horne, D. A. 2007. "The privacy paradox: Personal information disclosure intentions versus behaviors," *Journal of Consumer Affairs*, (41:1), pp. 100-126.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and trends in information retrieval* (2:1-2), pp. 1-135.

Pang, B., Lee, L., and Vaithyanathan, S. 2002. "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*: Association for Computational Linguistics, pp. 79-86.

Peterson, R. A., and Wilson, W. R. 1992. "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the academy of marketing science* (20:1), pp. 61-71.

Rizzo, G., and Troncy, R. 2011. "Nerd: Evaluating Named Entity Recognition Tools in the Web of Data,").

Rylander, R. G., Propst, D. B., and McMurtry, T. R. 1995. "Nonresponse and Recall Biases in a Survey of Traveler Spending," *Journal of Travel Research* (33:4), pp. 39-45.

Sadilek, A., Kautz, H. A., and Silenzio, V. 2012. "Predicting Disease Transmission from Geo-Tagged Micro-Blog Data," *AAAI*.

Safa, N. S., Von Solms, R., & Furnell, S. (2016). Information security policy compliance model in organizations. *Computers & Security*, 56, pp. 70-82.

Saif, H., He, Y., and Alani, H. 2012. "Semantic Sentiment Analysis of Twitter," *International Semantic Web Conference*: Springer, pp. 508-524.

Sharma, S., & Crossler, R. E. 2014. "Disclosing too much? Situational factors affecting information disclosure in social commerce environment," *Electronic Commerce Research and Applications*, (13:5), pp. 305-319.

Singh, V. K., Piryani, R., Uddin, A., and Waila, P. 2013. "Sentiment Analysis of Movie Reviews: A New Feature-Based Heuristic for Aspect-Level Sentiment Classification," *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on*: IEEE, pp. 712-717.

Shropshire, J., Warkentin, M and Sharma S. 2015. "Personality, attitudes, and intentions: predicting initial adoption of information security behavior." *Computers & Security* 49, pp. 177-191.

Spiekermann, S., Grossklags, J., & Berendt, B. 2001. "E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior," In *Proceedings of the 3rd ACM conference on Electronic Commerce* .pp. 38-47.

Stieglitz, S., and Dang-Xuan, L. 2013. "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior," *Journal of Management Information Systems* (29:4), pp. 217-248.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," *Computational linguistics* (37:2), pp. 267-307.

Taddicken, M. 2014. "The 'Privacy Paradox'in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure," *Journal of Computer-Mediated Communication*, (19:2), pp. 248-273.

Thelwall, M., Buckley, K., and Paltoglou, G. 2011. "Sentiment in Twitter Events," *Journal of the American Society for Information Science and Technology* (62:2), pp. 406-418.

Torres, R. Johnson, V., Rahnamaee, A., and Phillips, B. 2014. "Location Based Service Use: The Role of Trust and Privacy, In Proceedings of *Twentieth Americas Conference on Information Systems*, *Savannah, Georgia*.

Tsytsarau, M., and Palpanas, T. 2012. "Survey on Mining Subjective Data on the Web," *Data Mining and Knowledge Discovery* (24:3), pp. 478-514.

Tufekci, Z. 2008. "Grooming, gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate?," *Information, Communication & Society*, (11:4), pp. 544-564.

Turney, P. D. 2002. "Thumbs up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proceedings of the 40th annual meeting on association for computational linguistics*: Association for Computational Linguistics, pp. 417-424.

Von Solms, R., and Van Niekerk, J. 2013. "From Information Security to Cyber Security," *computers & security* (38), pp. 97-102.

Wang, C., Xiao, Z., Liu, Y., Xu, Y., Zhou, A., and Zhang, K. 2013a. "Sentiview: Sentiment Analysis and Visualization for Internet Popular Topics," *IEEE transactions on human-machine systems* (43:6), pp. 620-630.

Wang, J., Gu, Q., and Wang, G. 2013b. "Potential Power and Problems in Sentiment Mining of Social Media," *International Journal of Strategic Decision Sciences (IJSDS)* (4:2), pp. 16-26.

Xu, H., Gupta, S., Rosson, M. B., and Carroll, J. M. 2012. "Measuring Mobile Users' Concerns for Information Privacy," *Thirty Third International Conference on Information Systems*, Orlando, Florida, pp. 1-16.

Yang, C. C., Tang, X., Wong, Y., and Wei, C.-P. 2010. "Understanding Online Consumer Review Opinions with Sentiment Analysis Using Machine Learning," *Pacific Asia Journal of the Association for Information Systems* (2:3).

Yu, H., and Hatzivassiloglou, V. 2003. "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences," In *Proceedings of the 2003 conference on Empirical methods in natural language processing*: Association for Computational Linguistics, pp. 129-136.

Zafeiropoulou, A. M., Millard, D. E., Webber, C., & O'Hara, K. 2013. Unpicking the privacy paradox: can structuration theory help to explain location-based privacy decisions?. In *Proceedings of the 5th Annual ACM Web Science Conference*, pp. 463-472.