# Understanding DDoS Cyber-Attacks using Social Media Analytics

Sumeet Kumar*, Kathleen M. Carley†
*Department of Electrical and Computer Engineering
†School of Computer Science
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, USA
Email: {sumeetku@cmu.edu, kathleen.carley@cs.cmu.edu}

***Abstract*—Cyber-attacks are cheap, easy to conduct and often pose little risk in terms of attribution, but their impact could be lasting. The low attribution is because tracing cyber-attacks is primitive in the current network architecture. Moreover, even when attribution is known, the absence of enforcement provisions in international law makes cyber attacks tough to litigate, and hence attribution is hardly a deterrent. Rather than attributing attacks, we can re-look at cyber-attacks as societal events associated with social, political, economic and cultural (SPEC) motivations. Because it is possible to observe SPEC motives on the internet, social media data could be valuable in understanding cyber attacks.**

**In this research, we use sentiment in Twitter posts to observe country-to-country perceptions, and Arbor Networks data to build ground truth of country-to-country DDoS cyber-attacks. Using this dataset, this research makes three important contributions: a) We evaluate the impact of heightened sentiments towards a country on the trend of cyber-attacks received by the country. We find that, for some countries, the probability of attacks increases by up to 27% while experiencing negative sentiments from other nations. b) Using cyber-attacks trend and sentiments trend, we build a decision tree model to find attacks that could be related to extreme sentiments. c) To verify our model, we describe three examples in which cyber-attacks follow increased tension between nations, as perceived in social media.**

## I. INTRODUCTION

Forensic of cyber-attacks is hard. Traditional tracking methods require logs from infected machines and network routers. Given the fact that attacks could originate from infected bots, IP spoofing is easy to conduct, and activity logs could be deleted, Lipson argued that tracking and tracing cyber-attacks [1] is 'primitive at best' in today's network architecture. To uncover anonymous attackers, we require new technical considerations that could enhance the track and trace capabilities of the internet. However, because primary actors (e.g. countries) have divergent strategic interests, the formation of a stable international consensus is difficult. Shackelford explains that the absence of enforcement provisions in international laws [2] makes cyber attacks an easy to conduct but hard to penalize offense. We argue that even though it may be difficult to identify attackers through forensics; it should still be possible to determine the factors leading to some of the nation-to-nation cyber-attacks, by observing the social tension between nations in social media. We expect that such an approach to understanding cyber-attacks will provide the security researchers a different dimension to better understand their adversary.

Today, cyber-attacks are studied predominantly as a technical issue. However, cyber-attacks are societal phenomena associated with social, economic, cultural and political motivations [3]. Because cyber-attacks are known to be geared towards individuals, services and organizations, we may be able to perceive some of these motivations on social media [4]. In particular, the distributed denial of service (DDoS) attacks resemble proxy wars. These attacks are cheap to conduct and harder to attribute because of their distributed nature, so people, organizations or even governments use DDoS as a form of cyber-warfare to mark their dissent. Apart from defending the attacks, targeted parties often find their options limited. In such a case, for a long-term strategy to control attacks, understanding the motivations behind cyber-attacks is as important as defending against attacks.

In this research, we try to find if we can use social media data to find factors that impact country-to-country cyber-attacks, and if those factors could provide more insights into the motivations behind attacks. Often people use social media to express anger towards an event, a decision or a policy change. The same angst could also lead to cyber-attacks, another way to show disagreement. Given the possibility that social media and cyber-attacks are two separate ways to show anguish, we use social media data to observe anguish expressed by people in one country towards another country and then try to correlate the increased tension with any jump in cyber-attacks in a close time window. This approach could indirectly lead to the motives behind attacks, and thereby, a better understanding of cyber-attacks. For example, Canada's Anti-terrorism Act, 2015 was passed on June 18, 2015. During the same time frame, Canada's government websites saw an extensive series of DDoS attacks. The increased attacks correlated to an increase in negative sentiment towards Canada. Though 'Anonymous' group claimed responsibility for the attacks later, the relation between negative perception towards the Act and increased attacks clearly indicated that the dissent against the bill was a motivating factor. The passing of Canada's Anti-terrorism Act, the negative sentiment towards the bill and a high correlation of negative sentiment with increased cyber-attacks, make a case for using social media to understand attacks. Like the cyber-attack on the Canadian government websites, it should be possible to find other examples where understanding changes in social perception could help to understand reasons behind attacks.

To summarize, in this research, we try to answer the following questions: a) Determine if increased negative sentiment

built towards a country, increases the probability of cyber-attacks? b) Determine if a higher positive sentiment towards a country, decreases the possibility of cyber-attacks in near future? c) Can we find some instances where an increase negative sentiment towards a country followed an increase in cyber attacks on that country? For those instances, can we use tweets (used for determining sentiment) data to understand the motives behind cyber-attacks?

This paper is organized as follows. First we describe the key related work (sec II). Then in section III, we describe our data sources. In section IV, we show that the sentiment trend towards the country influences the probability of attacks received by a country. In the next section (V), we use a decision tree model to relate cyber-attacks with extreme negative sentiments. Finally, we present our conclusion and suggest future directions for this type of research.

## II. RELATED WORK

Social media is commonly used as a tool for sharing information. The data available on social media have been used for decision making and to answer research questions [5]. However, to the best of our knowledge, there is not any research article that uses social media data to understand cyber-attacks. There are publications in related research areas like cyber-attacks and cyber-forensic, motivations behind cyber-attacks and sentiment mining. In this section, we discuss some related papers in each of these research areas.

### A. Cyber Attacks and Forensics

The impact of cyber-attacks, and cyber espionage events is well documented through reviews of key events [6]–[8]. It is estimated that the actual damage by cyber attacks on world economies could run in billion of dollars [8]. However, the exact impact of attacks [9] is difficult to measure. For example, the Estonia cyber attack in 2007 had an almost devastating [10] impact on the country. To investigate such crimes, Cyber Forensic [11] is used for collecting, examining, and preserving evidence of computer crimes. Though forensic is useful, often it is limited in identifying motives behind cyber-attacks.

### B. Motivation behind Cyber-Attacks

Cyber attacks are frequently only examined from a technical perspective. However, cyber-attacks are social events. Mezzour [12] found that the socio-technological sophistication of a country's IT infrastructure and it's economy affected the likelihood that it would be attacked. Gandhi et al. [3] argued that cyber-attacks are associated with social, political, economic and cultural (SPEC) motives. They explained that to effectively prevent cyber-attacks it is necessary to consider the socio-technological sophistication, and the background and motivation of the cyber attackers.

### C. Social Media and Sentiment Mining

With the exponential growth of data on web carrying public opinion, a number of researchers have tried opinion mining and sentiment analysis on social media data. Liu et al. [13] describe a number of such approaches in their survey paper. Some researchers have tried targeted (contextual) sentiment mining as well, e.g. Jiang et al. [14] classify the sentiments of the tweets as positive, negative or neutral according to whether they contain positive, negative or neutral sentiments about a query. There are many applications of such targeted sentiment mining, and one of the interesting application is to understand what people in one country think of other countries. Chambers et al. [15] used Twitter data to model relations between nation states. They verified their model with two public opinion polls and international alliance relationships. We use their nation-to-nation sentiments data to understand the impact of heightened sentiments towards a country on cyber-attacks received by a country.

## III. DATASETS

We build our dataset from two data sources. The first data-source is country-to-country sentiment data [15]. We used the everyday count of positive, negative and neutral tweets to obtain the trends of sentiment. The second dataset used is the ddos-attacks data from Arbor Networks [16]. This dataset is used to build cyber-attacks trend. Both these data-sources are explained in detail next.

### A. Country-to-country Sentiment

The country-to-country sentiment data is sourced from a study conducted at USNA [15], and is publicly available on website http://www.usna.edu/Users/cs/nchamber/nations/index.html. However, the publicly available data only contains weekly ratio of positive tweets to negative tweets. Since for our research, we needed a more fine grained daily count of sentiment tweets, we asked the authors for additional data. They shared additional data which contains the count of positive, negative, neutral tweets between many country pairs for a period of around two years (9/03/2013 to 07/27/2015).

The country-to-country sentiment data [15] was collected using a Twitter API, based on geo-spatial and country names filtering. A number of filters were used to clean the twitter data , and a number of features were used to build directed sentiment towards a country. Some of the features used were from Semeval 2013 and 2014 challenge ( [17]). To know more about their data collection, and the model used for building country to country sentiment, please see the paper [15].

### B. DDoS Cyber-attacks Data

The data on Distributed Denial of Service (DDoS) type cyber-attacks is collected from the website www.digitalattackmap.com. Arbor Networks and Google Ideas together created this website to visualize global DDoS attacks threat. In addition to visualizing recent attacks, the site allows users to explore historical trends of attacks. The website shares the top 2% of global ddos-attacks (reported by Arbor Networks) on the website. The json data used for visualization could be downloaded from the website to build a trend of ddos-attacks.

In this research, we are only interested in the extreme values of attacks. This is because the DDoS attacks are a result of a variety of external events, which in turn show up in Arbor Networks' honeypots. We can make a simplifying assumption that the majority of attacks are some sort of noise, whereas only a minority of bigger attacks are caused by external factors. It is then plausible that these bigger changes contain most of the relevant information for relating cyber-attacks and extreme sentiments. For our analysis, we consider attacks with bandwidth that are two standard deviations above mean attacks bandwidth. Note that this fixed threshold may be over simplistic. Since we consider a period of two years and the time series is not stationary over this long time period, attacks that are considered as extreme in one year may be considered as quite ordinary another year. However, we observe that there is large variation in bandwidth of attacks over time, and a rolling threshold was not a good choice because it even included some very low bandwidth attacks. So for each country pair, we used a fixed threshold of two standard deviation from mean attacks bandwidth over the entire time period of analysis.

## IV. Change in Attacks Probability

In this section, we will try to find if an extreme sentiment (positive or negative) event towards a country changes the probability of cyber-attacks on the country in near future. A simple model to asses this is to use attacks and sentiment trend data, and find the probability of attacks in a time window after extreme sentiment events ($P(ES)$), and find the probability of attacks in a time window when there are no extreme sentiment events ($P(NES)$) i.e. when the sentiment trend is normal. The ratio of $P(Attack|ES)$ and $P(Attack|NES)$ indicates change in probability of attacks when a country is experiencing heightened sentiments when compared to attacks during regular sentiments.

To measure $P(ES)$ and $P(NES)$, we use a simple model based on Bayes theorem. Using Bayes rule, we define:

$$P(Attack|ES) = \frac{P(Attack \cap ES)}{P(ES)} \quad (1)$$

where $ES \Rightarrow ExtremeSentiment$ and $Attack \Rightarrow$ CyberAttacks that are two standard deviations above mean attacks bandwidth.

In equation 1, $P(ES)$ i.e. the probability of extreme sentiment, and can be calculated using sentiment trend data. For calculation, we use sentiment above two standard deviation of rolling mean sentiment as extreme sentiment. The rolling time window was picked to be one week. $P(Attack \cap ES)$ can be calculated by finding the count of attacks event that followed extreme sentiment in a small time window. We took the time window of three weeks after each extreme sentiment event. The three weeks window is based on on the assumption that some time needed for for organization and preparation in conducting a big attack.

Similar to equation 1, using Bayes rule, we define:

$$P(Attack|NES) = \frac{P(Attack \cap NES)}{P(NES)} \quad (2)$$

where $NES \Rightarrow NonExtremeSentiment$ and $Attack \Rightarrow$ CyberAttacks that are two standard deviations above mean attacks bandwidth. We used the same parameters for measuring $P(Attack|NES)$ that we used for measuring $P(Attack|NES)$. The only difference is that in equation 2, $P(Attack \cap NES)$ was calculated using count of attacks event that followed any non-extreme sentiment in a time window of three weeks.

Comparing P(Attack|ES) and P(Attack|NES) gives us an estimate of change in probability of attacks given extreme sentiment. We can measure this for different countries. For each country, we use data of cyber-attacks received by that country and trend of sentiment expressed by people from other countries towards that country. The two plots (1, 2) show the change in probability of attacks for extreme negative sentiment and extreme positive sentiment. For visualization clarity, we have removed the countries for which the probability of attack was less than 1%.

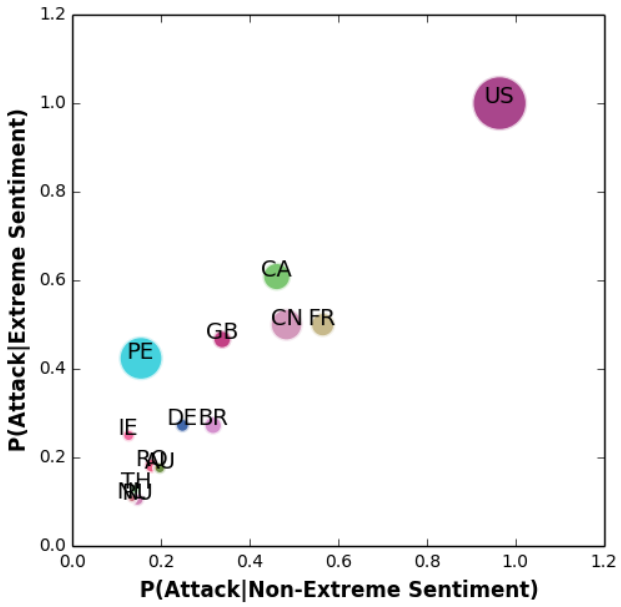


**Fig. 1:** Scatter plot for P(Attack|Extreme Sentiment) vs P(Attack|Non Extreme Sentiment) for Negative sentiment, where extreme negative sentiment implies sentiments that are two standard deviation above the mean sentiment, and 'Non Extreme Sentiment' indicates absence of extreme negative sentiment. The size of country nodes indicate the number of high bandwidth cyber-attacks received by the country. As it can be observed, the probability of attacks increases for most countries experiencing negative sentiment from other countries.

The scatter plot (Fig:1) indicates that, in general, a higher negative sentiment increases the probability of cyber-attacks within three weeks of extreme negative sentiment. For some countries (Peru:26.8%, Canada:14.7%, UK:12.7%, Ireland:12.2%), the increase in probability is as high as 26.8%. Also, we can observe that increase in positive sentiment (Fig:2), in general, decreases the probability of cyber-attacks. The result was obtained using three weeks time window after the date of increased positive sentiment. For some countries

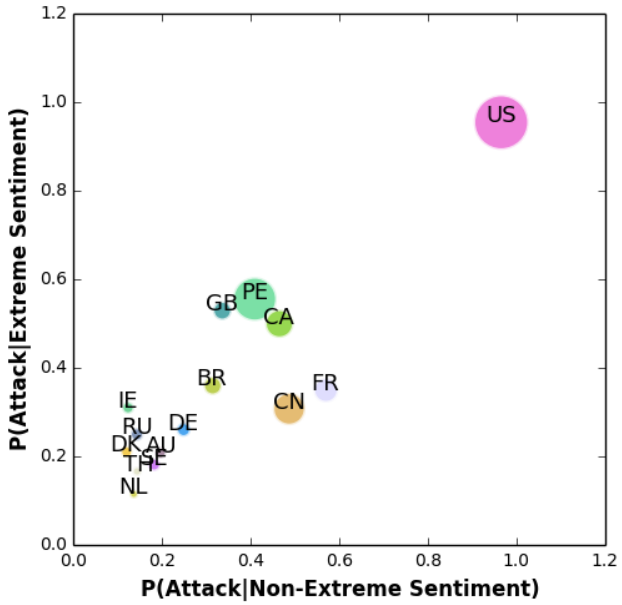


**Fig. 2:** Scatter plot for P(Attack|Extreme Sentiment) vs P(Attack|Non Extreme Sentiment) for Positive sentiment, where extreme positive sentiment implies sentiments that are two standard deviation above mean sentiment, and 'Non Extreme Sentiment' indicates absence of extreme positive sentiment. As it can be observed, the probability of attacks decreases for most countries experiencing positive sentiment from other countries. The size of country nodes indicate the number of high bandwidth cyber-attacks received by the country.

(China:17.9%, Ireland:18.6%, France:22.0%) the decrease in attacks probability is as high as 22%.

Before going forward, we would like to highlight that cyber-attacks are a result of diverse set of reasons, and negative sentiments (because of certain events) is only one of those reasons. Moreover, this research does not aim to find all the reasons behind cyber-attacks, but to find those instances of attacks that are likely to be influenced by negative sentiments. With that goal, a change in attacks probability by 22% because of extreme sentiments, could be considered a good result.

For a country like the USA (US), the impact of sentiment is not clear. This is because the USA receives a huge volume and count of cyber attacks, i.e. almost daily. The continuous attacks increase the probability of attacks for extreme sentiment events as well as non-extreme sentiment events. Therefore, p(Attack|ES) and p(Attack|NES) for both positive and negative sentiment are close to one. Thus, sentiment towards USA is not a great indicator of increased cyber-attacks. However, we can look at cyber-attacks on the USA from a particular country (i.e. A to B attack), and that might give more informative results. This analysis is not covered in this paper, and is a good candidate for future work.

This Bayesian analysis to understand the impact of extreme sentiment only gives an average estimate of the influence of sentiments on cyber-attacks. In the next section, we discuss a way to find attacks that are likely to be a result of extreme sentiments.

## V. Decision Tree for Finding Attacks

Having verified that strong sentiments impact cyber-attacks, we aim to find instances where attacks on a particular country get impacted by extreme sentiments. To find such instances, we built a decision tree model (Fig:4). The model takes time series data for country-to-country attacks and country-to-country sentiment as input, iterates over all extreme sentiment events, and output attacks that are likely to be a result of an extreme sentiments. Given that it requires some planning to conduct DDoS attacks, we expected that the attacks that are a result of extreme sentiments will be conducted in a time window following the extreme negative sentiment. So we used a time window of three weeks as a parameter. To find important cyber-attacks events, we used attacks with bandwidth above two standard deviation from mean attacks bandwidth. For finding extreme sentiments, we used rolling mean and rolling standard deviations of seven days time window. We again used two standard deviation above mean to find extreme sentiment.

We could run this experiment for two combinations: a) A pair of countries b) One country vs rest of the countries. The first case is simple where we consider attacks received by country A from country B, and we consider sentiment towards country A from people of country B. We then use the model (Fig:4) to find attacks than followed heightened negative sentiment. An example of this type of analysis is shown in Fig:3. For the second case, we aggregate the attacks data received by a country A from all other counties, and we combine the sentiments data received by country B from all other countries. This results in a trend of attacks received by country A from rest of the countries, and a trend of sentiments towards country A from people in rest of the world. We again use the model (Fig:4) to find attacks than followed heightened negative sentiment. Two examples of this type of analysis is shown in Fig:5 and Fig:6.

Since there is no ground truth data on cyber-attacks that are related to increased negative sentiment, we are not able to provide an accuracy estimate of this model. Instead, we give some examples where it is evident that cyber-attacks are related to increases negative sentiment. The extreme sentiment could be because of many reasons like increased stress between two countries because of a policy change or a political event. Once we have an extreme sentiment event that relates to cyber-attacks, we can use various methods to further understand the reasons for extreme sentiment. For Twitter data, topic modeling [18] could be used to find the broad range of topics in the set of tweets. Another possibility is to use news search (e.g Google advanced search) to find any important events in a time range. For our results, we used both Twitter search and Google News search for finding events that were likely to result in strong negative sentiments, and in turn to cyber-attacks.

Fig:3 shows the trend of attacks received by Russia from the USA, and the trend of negative sentiment towards Russia from people in the USA. As it could be observed in the trend, there are at least three instances (highlighted in Yellow) where an increased cyber attack followed an increased negative sentiment. One of the increased negative sentiment trend is in

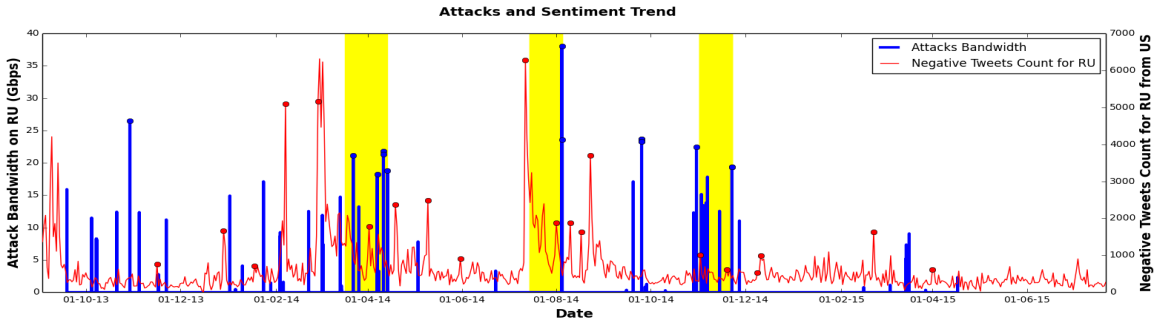


**Fig. 3:** Attacks and Sentiment Trend for Russia. The plot shows the bandwidth of attacks received by Russia from the USA. The red dots show the sentiment that were two standard deviation above mean sentiment using a rolling time window of one week. The blue dots show attacks bandwidth that were higher than two standard deviation above mean bandwidth of attacks. As it could be observed in the trend, there are at least three instances (highlighted in Yellow) where increased cyber attacks followed increased negative sentiment.

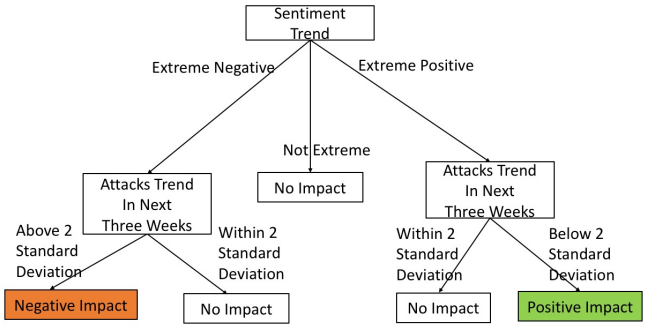


**Fig. 4:** Decision Tree model to find attacks influenced by strong sentiments.

August 2014, and is related to Russia sending combat troops to Ukraine.

Fig:5 shows the trend of attacks received by Canada from the rest of the world, and the trend of negative sentiment towards Canada from people in rest of the countries. As it could be observed in the trend, there are many instances (highlighted in Yellow) where an increased cyber attack followed an increased negative sentiment. One of the increased negative sentiment trend in June, 2015 is related to anti-terrorism act that was passed on June 18, 2015.

Fig:6 shows the trend of attacks received by Peru from the rest of the world, and the trend of negative sentiment towards Peru from people in rest of the countries. As it could be observed in the trend, there are many instances (highlighted in Yellow) where an increased cyber attack followed an increased negative sentiment. One of the increased negative sentiment trend in June, 2014 is related to soccer world-cup in Peru.

## VI. Conclusion and Future Work

In this research, we used country-to-country sentiments data from a USNA study and, compared the sentiments trend with the cyber-attacks trend. We specifically looked at the DDoS type cyber-attacks shared by Arbor Networks. The analysis found, for many countries, a negative opinion towards a country increases the probability of cyber-attacks on the country and, a positive sentiment towards a country decreases the chance of cyber-attacks on the country. For some countries (e.g. Peru), the increase in attacks probability after an increased negative sentiment is as high as 26.8%. In contrast, a positive attitude towards a country (e,g, France) decreases the cyber-attacks probability by up to 22%. We also built a decision tree model to find the instances of attacks that are likely to be a result of extreme negative sentiment. For three countries, we presented the trend of sentiments and cyber-attacks, highlighting those attacks which are likely to be a result of extreme negative sentiments. We also proposed ways to further understand the negative sentiment, by topic modeling of tweets as well as news and Twitter search.

To summarize, this research is unique in that it quantitatively analyzes the impact of sentiment on the reality of cyber attacks. This study finds that as new inter-nation events occur, the expression of opinion towards such events find their way to social media, and sometimes to the hackers community. This leads to an aggregate change of sentiment towards countries, and the change could be measured via social media analytics. Such events, in some cases, also result in increased cyber-attacks experienced by a country. This study argues that the creation of tools that can monitor the increase of the attacks and find the sentiment that triggers such attacks may serve to better understand the attacks and, in turn to, decrease the number of actual attacks. Thus, we suggest that the social media data sources could be used as a sensor for extreme sentiments, and could be integrated with other attacks analysis tools to get a holistic view of the cyber-attacks situation.

In future, we would like to use other publicly available data sets (e.g. news) to mine sentiments. We would also like to build an alliance hostility trend and relate the trend to cyber attacks trend.

## VII. Acknowledgments

This work was supported in part by the NSA under Award No. H9823014C0140, by a MURI N000140811186 on adversarial reasoning, and the Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Security

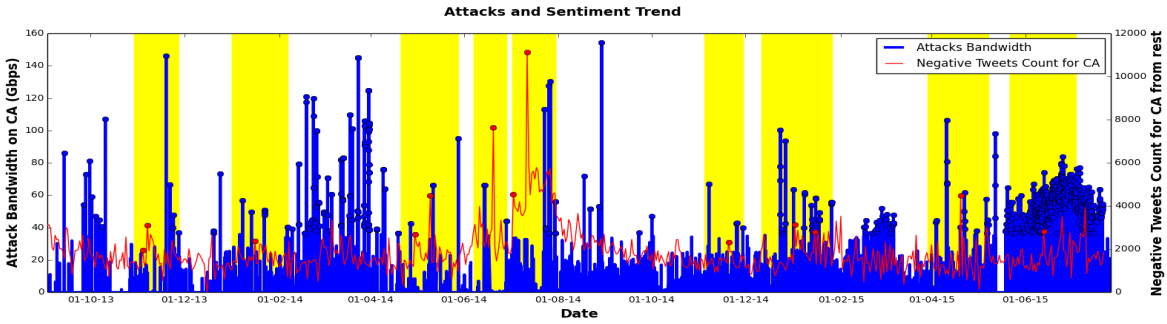


**Fig. 5:** Attacks and Sentiment Trend for Canada from rest of the world. The red dots show the sentiment that were two standard deviation above mean sentiment using a rolling time window of one week. The blue dots show attacks bandwidth that were higher than two standard deviation above mean bandwidth of attacks. As it could be observed in the trend plot, there are many instances (highlighted in Yellow) where increased cyber attacks followed increased negative sentiment.

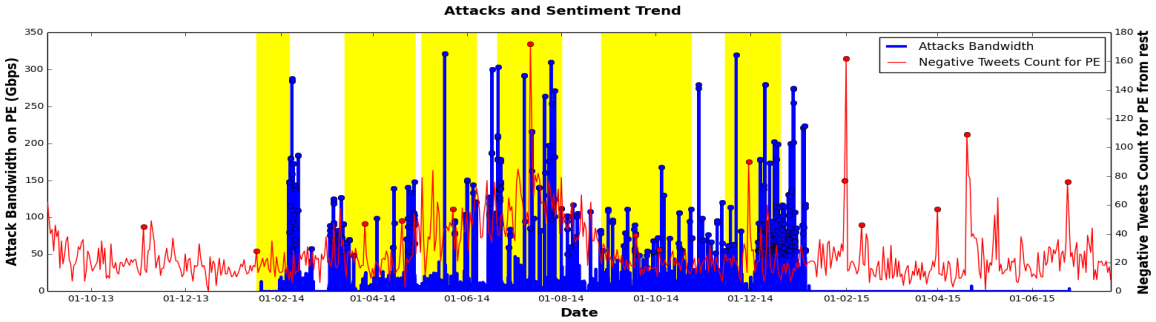


**Fig. 6:** Attacks and Sentiment Trend for Peru from rest of the world. The red dots show the sentiment that were two standard deviation above mean sentiment using a rolling time window of one week. The blue dots show attacks bandwidth that were higher than two standard deviation above mean bandwidth of attacks. As it could be observed in the trend plot, there are many instances (highlighted in Yellow) where increased cyber attacks followed increased negative sentiment.

Agency, the Office of Naval Research, or the U.S. government. We would also like to thank Dr. Nathanael Chambers from the US Naval Academy for sharing his research data on country-to-country sentiments.

## REFERENCES

[1] H. F. Lipson, "Tracking and tracing cyber-attacks: Technical challenges and global policy issues," DTIC Document, Tech. Rep., 2002.

[2] S. Shackelford, "From nuclear war to net war: analogizing cyber attacks in international law," *Berkley Journal of International Law (BJIL)*, vol. 25, no. 3, 2009.

[3] R. Gandhi, A. Sharma, W. Mahoney, W. Sousan, Q. Zhu, and P. Laplante, "Dimensions of cyber-attacks: Cultural, social, economic, and political," *Technology and Society Magazine, IEEE*, vol. 30, no. 1, pp. 28–38, 2011.

[4] S. Kumar and K. M. Carley, "Approaches to Understanding the Motivations Behind Cyber Attacks," in *Intelligence and Security Informatics (ISI), 2016 IEEE International Conference on*, Tucson, Arizona USA, Sep. 2016.

[5] S. Stieglitz, L. Dang-Xuan, A. Bruns, and C. Neuberger, "Social media analytics," *Wirtschaftsinformatik*, vol. 56, no. 2, pp. 101–109, 2014.

[6] G. O'Hara, "Cyber-Espionage: A growing threat to the American economy," *CommLaw Conspectus*, vol. 19, p. 241, 2010.

[7] E. Nakashima, "US Target of Massive Cyber-Espionage Campaign," *Washington Post*, 2013.

[8] J. Lewis and S. Baker, "The economic impact of cybercrime and cyber espionage," *Center for Strategic and International Studies, Washington, DC*, pp. 103–117, 2013.

[9] S. Kumar and K. M. Carley, "DDoS Cyber-Attacks Network: Who's Attacking Whom," in *Intelligence and Security Informatics (ISI), 2016 IEEE International Conference on*, Tucson, Arizona USA, Sep. 2016.

[10] R. Ottis, "Analysis of the 2007 cyber attacks against estonia from the information warfare perspective," in *Proceedings of the 7th European Conference on Information Warfare*, 2008, p. 163.

[11] A. Marcella Jr and R. S. Greenfield, *Cyber forensics: a field manual for collecting, examining, and preserving evidence of computer crimes*. CRC Press, 2002.

[12] G. Mezzour, "Assessing the Global Cyber and Biological Threat," Ph.D. dissertation, Symantec Research Labs, 2015.

[13] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*. Springer, 2012, pp. 415–463.

[14] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011, pp. 151–160.

[15] N. Chambers, V. Bowen, E. Genco, X. Tian, E. Young, G. Harihara, and E. Yang, "Identifying political sentiment between nation states with social media," in *Proc. EMNLP*, 2015, pp. 65–75.

[16] "www.digitalattackmap.com." [Online]. Available: http://www.digitalattackmap.com/

[17] "Semeval Challenge 2014." [Online]. Available: http://alt.qcri.org/semeval2014/task9/

[18] L. Hong and B. D. Davison, "Empirical study of topic modeling in twitter," in *Proceedings of the first workshop on social media analytics*. ACM, 2010, pp. 80–88.