Approaches to Understanding the Motivations Behind Cyber Attacks

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 appear to have grown in the past few years. Many policy changes have been made to control them, however, a clear path to a safe and secure cyber world is not visible. A mere technical view to confront the cyber-attacks could be futile if motivations behind attacks are not understood.**

To understand the motives behind cyber-attacks, we look at cyber-attacks as societal events associated with social, economic, cultural and political factors. To find factors that encourage unsafe cyber activities, we build a network of aggregate country-to-country cyber-attacks, and compare the network with other country-to-country networks. In particular, we use variable correlation and network correlation (QAP) to examine the relevance of factors like income difference, alliance-hostility, internet penetration, and corruption. We observe that higher corruption and a large internet bandwidth favors attacks origination. We also find that countries with higher Per-capita-GDP and better Information and Communication Technologies (ICT) infrastructure are targeted more often.

I. INTRODUCTION

Cyber attacks are frequently only examined from a technical perspective. However, cyber-attacks are social events. Gandhi et al. [1] explained that cyber-attacks are associated with social, political, economic, and cultural (SPEC) conflicts. The authors argued that to effectively prevent cyber-attacks, it is necessary to consider the socio-technological sophistication and the background and motivation of the cyber attackers. Mezzour [2] found that the socio-technological sophistication of a country's IT infrastructure and it's economy affected the likelihood that it would be attacked. Therefore, establishing an empirical association between cyber-attacks and SPEC factors is important; however, we know of no other literature that provides this linkage quantitatively.

Many researcher have investigated the technologies used in cyber-attacks and ways to defend cyber-attacks, but a good way of determining the SPEC reasons behind cyber-attacks is missing. Understanding SPEC reasons require a broader perspective and detailed analysis, because often the real intent of a cyber attack stays hidden [3]. The situation becomes even more complex when cyber-attacks are across countries [4]. However, even though it may be difficult to explain individual attacks, it should still be desirable to determine from a longitudinal perceptive the SPEC factors leading to attacks. In this research, we take a high-level view of cyber-attacks. Rather that focusing on a few instances of attacks, we consider the network of country-to-country attacks. We try to determine whether such attacks could be related to variables that represent SPEC reasons. Thus, we investigate the feasibility of finding SPEC factors that impact country-to-country cyber-attacks. The broad goals of this research are:

1) Discover central members in country-to-country cyber-attacks e.g. Which countries are attacked more often? Which country attacks the most?
2) Are cyber-attacks increasing or decreasing?
3) Determine whether there is a correlation between cyber attacks and characteristics of the country (like GDP, Internet bandwidth) ?
4) Determine whether the country-to-country cyber-attacks network could be compared to other country based networks (e.g. alliance-hostility network)?

In this paper, we present the initial results on quantitatively measuring SPEC factors behind cyber-attacks using variable correlation and network QAP. In sec:III, we first describe our data sources. We find the trend of attacks in sec:IV Then, We use two types of correlation analysis a) variable correlation (Section V) b) network analysis (Section VI). Finally, we present our conclusion and suggest some possible future directions.

II. RELATED WORK

Several researchers [5], [6] have highlighted the impact of cyber-attacks. While it is estimated that the actual damage by cyber attacks on world economies run in billion of dollars [7], the exact impact of attacks is difficult to measure. Some argue that theft constitutes the "greatest transfer of wealth in history" [8]. But most cyber-attacks [9] are innocuous to the general public, and only a few create major impacts. For example, Estonia cyber attack in 2007 had an almost a devastating [10] impact on the country. To understand the factors behind cyber-attacks, Gandhi et al. [1] explained that cyber-attacks are associated with social, political, economic and cultural conflicts (SPEC), and argued that an effective prevention of cyber-attacks need to consider the socio-technological sophistication, background, and motivation of cyber attackers. Though there are many research articles on instances of cyber-attacks, and techniques used in attacks, not many researchers discuss cyber-attacks trend and motivations behind cyber-attacks. That is the topic of this research.

III. DATASETS

We used various data sources to build our dataset. We use Arbor Networks DDoS-attacks data from the website www.digitalattackmap.com to create country-to-country attacks network. The dataset starts from May 2013 and is updated daily, and shares top 2% of global ddos-attacks registered by Arbor Networks. We used data till March, 2106 for this analysis. The data includes a time-series of cyber-attacks with source country (if available), target country and bandwidth-of-attack (bps) information. Using this attacks data, we can build a country-to-country average (or total) attack-bandwidth network, based on source and target country information. Apart from the attacks data, we used data from the 'World Bank' website to represent country level parameters like Information and Communication technologies (ICT) infrastructure, Per-capita-GDP, Country Policy and Institutional Assessment (CPIA) for corruption and 'Internet Users per 100 population'. We used www.econstats.com to get international internet bandwidth information and 'Correlates of War' (www.coorelatesofwar.org) website to build alliance-and-hostility network. We used data from USNA for country-to-country sentiments trend [11].

IV. Time Series Analysis

Since we have a time series data of cyber-attacks with source, target and bandwidth-of-attack information, we can process the data to build the attacks trend. For example, for a particular year, we can find which countries are the top attacking countries and, which countries are top receivers. Moreover, if we compare different years, we can find if attacks on a particular country are increasing or decreasing. Furthermore, for a given country, we can observe which country is attacking it the most. Similarly, we can discover which countries are getting attacked by a country.

Using the above mentioned approach, we did a time-series trend analysis of attacks. Fig:1 shows the trend of cyber attacks for a few top target countries. The plot uses the bandwidth of attacks as the measure of attacks, and is based on Arbor Networks data (top 2% of global attacks). The USA remains the top target of DDoS attacks followed by China, Peru, France, and Canada.

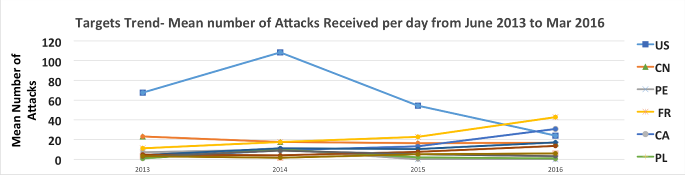


Fig. 1: Yearly Trend of Attacks Received

V. Country Level Correlation Analysis

We use country level correlation measure to understand the relationship between attacks received and sent, and variables associated with a country. We use Pearson's correlation, and p-value for significance testing. We selected *Attacks Sent* and *Attacks Received* as dependent variables and *Network Bandwidth*, *GDP* and *Internet Users per 100 population*, *ICT*, *CPIA* and *country-to-country average sentiment Score* as independent variables.

The results (Table:1) indicate a moderate correlation between network bandwidth and attacks sent. We also observe a moderate correlation between network bandwidth and attacks received. The results are significant with a p-value of 0.000. There is a weak correlation between GDP and attacks sent (0.15) with a p value of 0.05. The internet users per 100 population measure and attacks sent/attacks received, also correlate weakly.

Table 1: Cyber-attacks Correlation with Country level measures

	Network Bandwidth	GDP Per Capita	Internet Users per 100 population	ICT	ICT Import as Percent of Trade	CPIA[1]	Sentiment Score (USNA data)
Attacks Sent	**0.53** (P =0.000)	0.15 (P = 0.0544)	0.17 (P = 0.0283)	0.18 (P = .0194)	0.19 (P = .0116)	-0.11 (P = .1591)	-0.028 (P= .7197)
Attacks Received	**0.49** (P = 0.000)	0.12 (P = 0.122)	0.15 (P =0.0549)	.16 (P = .0418)	.25 (P = .0009)	-.098 (P = .2031)	-0.063 (P = 0.4178)

VI. Network Analysis

In this section, we use network visualization to find the major source and target countries of cyber-attacks. After building the attacks network, we first use ORA network visualizer, to find important nodes and high weight links. We then use Quadratic assignment procedure (QAP) [12] to find any correlation between cyber-attacks network, and networks created using other variables.

The cyber-attacks network was built using the steps mentioned in sec:III. Figure 2 visualizes the ddos-attacks received network. In the figure, the node size reflects the attacks received, and the edge color indicates the mean attack bandwidth of the attacks. We can observe that China and the US share the two center positions. For clarity, edges with a value less than 200 Gbps (total bandwidth of attack) were hidden in Figure 2, so we observe some one-way directed links, indicating either a higher level of attacks received.

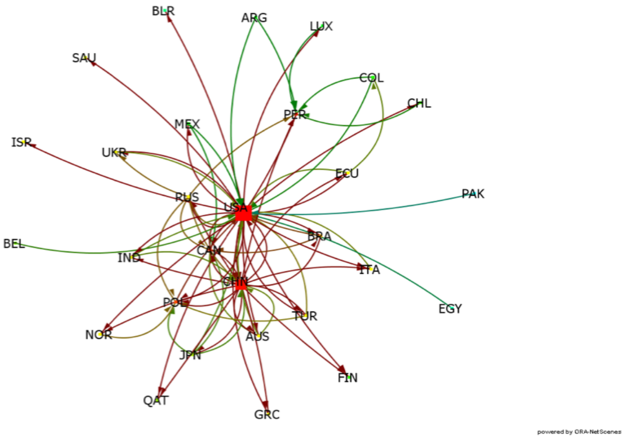
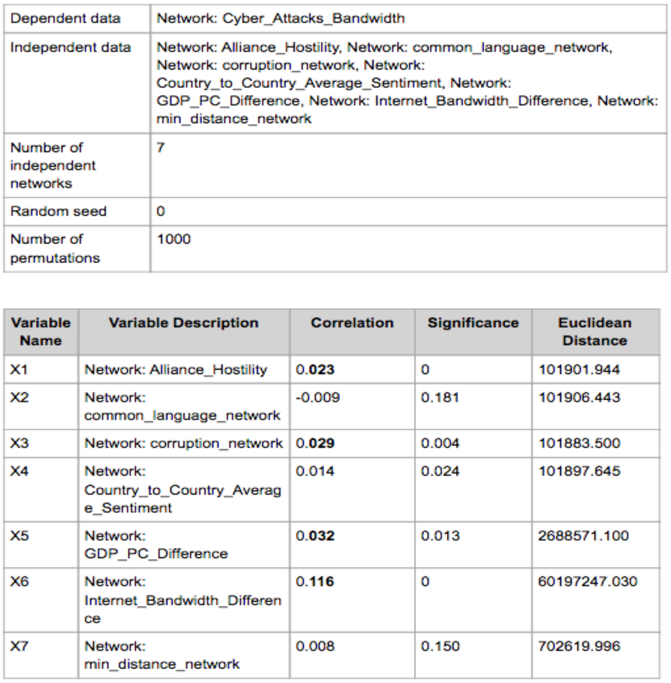


Table 2: Attacks Received Network of country-to-country cyber-attacks. The network diagram was generated using ORA software.

Since we have built a cyber-attacks network, we can do the Quadratic assignment procedure (QAP) analysis to compare attacks network with other country-to-country networks (like GDP difference network). QAP [13] tests an arbitrary graph level statistics against a QAP null hypothesis using Monte Carlo simulation of likelihood quantities while preserving row-column dependencies.

In the network level QAP analysis (Table: 3), we use attacks sent and received by each country to create a directed network, which is treated as the dependent network, and we use alliance hostility network, corruption index difference network, GDP (per-capita) difference network, internet bandwidth and minimum distance between two countries as independent networks. Then we did a QAP using ORA software with a random seed value 0 and 1000 number of permutations.

Table 3: QAP Correlation



Dependent data	Network: Cyber_Attacks_Bandwidth
Independent data	Network: Alliance_Hostility, Network: common_language_network, Network: corruption_network, Network: Country_to_Country_Average_Sentiment, Network: GDP_PC_Difference, Network: Internet_Bandwidth_Difference, Network: min_distance_network
Number of independent networks	7
Random seed	0
Number of permutations	1000

Variable Name	Variable Description	Correlation	Significance	Euclidean Distance
X1	Network: Alliance_Hostility	0.**023**	0	101901.944
X2	Network: common_language_network	-0.009	0.181	101906.443
X3	Network: corruption_network	0.**029**	0.004	101883.500
X4	Network: Country_to_Country_Average_Sentiment	0.014	0.024	101897.645
X5	Network: GDP_PC_Difference	0.**032**	0.013	2688571.100
X6	Network: Internet_Bandwidth_Difference	0.**116**	0	60197247.030
X7	Network: min_distance_network	0.008	0.150	702619.996

Table:3 shows the correlation and related statistics between the dependent network variable (cyber attacks = Y) and each independent network variable (X). As we can see in the correlation matrix, the networks have fairly low correlation coefficient. Given small correlation coefficient for corruption index and network bandwidth and 0 significance (p value), we can argue that cyber attacks network has small correlation with corruption index difference, network bandwidth, GDP-PC difference and alliance-hostility network .

VII. CONCLUSION AND FUTURE WORK

In this research, we looked at the cyber-attacks as a social phenomenon. We used network analysis to understand the motivations behind the attacks. Using network analysis of attacks, we find China (#1) and US (#2) attack the most, and that the US is attacked the most. Using variable correlation and network correlation, we find some important correlating factors. The analysis highlights that there is a medium correlation between bandwidth of cyber attacks and network bandwidth of a country. This correlation indicates that high bandwidth countries are good source of bots for making ddos attacks, possibly because high bandwidth countries could help to host computers that can execute high bandwidth ddos attacks. A weak correlation (Corr=0.032, P = 0.013) between cyber attacks network and GDP-per-capita network indicates wealth (economic) difference to be an important factor, if not the most important factor. A weak correlation of attacks-received-network with corruption-index-difference-network (Corr: 0.029. P = 0.004), indicates honest countries receive more attacks from corrupt countries.

This research is a beginning in understanding the complex network of country-to-country cyber-attacks. We would like to explore a few more options in future: a) We are currently limited by the absence of openly accessible data on cyber attacks. Using discussions on Twitter and News, we would like to build a data set of cyber-security events. b) We want to re-look at the alliance-hostility between countries as a time series network dictated by inter-nation events. We plan to use GDelt news data to build such a network.

VIII. ACKNOWLEDGMENT

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 Agency, the Office of Naval Research, or the U.S. government.

REFERENCES

[1] 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.

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

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

[4] 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.

[5] G. O’Hara, “Cyber-Espionage: A growing threat to the American economy,” *CommLaw Conspectus*, vol. 19, p. 241, 2010.

[6] E. Nakashima, “US Target of Massive Cyber-Espionage Campaign,” *Washington Post*, 2013.

[7] 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.

[8] J. ROGIN, “NSA Chief: Cybercrime constitutes the âĂIJgreatest transfer of wealth in historyâĂİ,” *http://foreignpolicy.com*. [Online]. Available: http://foreignpolicy.com/2012/07/09/nsa-chief-cybercrime-constitutes-the-greatest-transfer-of-wealth-in-history/

[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] S. Kumar and K. Carley, “Understanding DDoS Cyber-Attacks using Social Media Analytics,” in *Intelligence and Security Informatics (ISI), 2016 IEEE International Conference on*, Tucson, Arizona USA, Sep. 2016.

[12] D. Krackhardt, “Predicting with networks: Nonparametric multiple regression analysis of dyadic data,” *Social networks*, vol. 10, no. 4, pp. 359–381, 1988.

[13] J. L. Martin, “A general permutation-based QAP analysis approach for dyadic data from multiple groups,” *Connections*, vol. 22, no. 2, pp. 50–60, 1999.