

Global Perspectives on Cyber security Using Latent Dirichlet Allocation Algorithm

Lowell A. Quisumbing

Leyte Normal University, Tacloban City, Philippines.

Orcid: 0000-0002-1914-2274

Abstract

Cybersecurity is a global concern which affects every government, business, and individual. It is perceived to be the best solution towards online fraud, theft, and exploitation. However, the new corpus of studies reveals different viewpoints towards this phenomenon. This study investigated the plethora of opinions from the top eight countries having a high incidence of cybersecurity-based on Google trends. The data gathered from academic articles were processed utilizing the Web Mining procedure. Unsupervised Machine Learning with the utilization of Latent Dirichlet Allocation (LDA) algorithm applied for content examination was employed. Five underlying themes extracted by the researcher in light of the outcomes created by the R-Programming application upheld the writing, literature, and analysis of data.

Keywords: Cybersecurity, Latent Dirichlet Allocation Algorithm, Social Science, Unsupervised Machine Learning

INTRODUCTION

The United Nations Department of Economic and Social Affairs (UN DESA) described the online crime as a scheme which amasses a trillion dollars a year in online fraud, identity theft, and lost intellectual property that affects people, businesses, and governments of every nation. It is a complex transnational issue that requires global cooperation to ensure a safe Internet (Cyber security: A global issue demanding a global approach | UN DESA Department of Economic and Social Affairs. (n.d.). Retrieved August 21, 2017, from <http://www.un.org/en/development/desa/news/ecosoc/cybersecurity-demands-global-approach.html>).

A 2011 Norton study stated threats to cyberspace had increased dramatically afflicting 431 million adult victims globally or 14 adult victims every second, one million cybercrime victims every day (United Nations Department of Economic and Social Affairs, 2011). With the integration of the Internet of Things (IoT) and cloud computing, data resources are sure to be made available vastly online. These factors will open greater risks and threats for online attacks. Coincidentally, researchers agree that the best solution to cut off the risk and threats of

cybercrime is to develop an efficient and responsive cybersecurity.

Cybersecurity as defined, are the technologies and processes created to protect computers, equipment, software, networks and data from unauthorized access and vulnerabilities supplied through the Internet by cyber criminals, terrorist groups, and hackers. Cyber security is related to protecting the internet and network-based digital equipment and information from unauthorized access and alteration. Moreover, it includes the collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used to protect the cyber environment of organization and user assets (Maurer, T., 2011).

As the issue of cyber security becomes a national concern for most countries today, it is necessary to explore global opinions and views relative to this phenomenon. Using real-time data sets and research articles taken from the top countries in Google trends, this study examines the common and uncommon patterns derived from online articles to explain the issues of cyber security utilizing Unsupervised Machine Learning and the Latent Dirichlet Allocation algorithm for content analysis of online documents. It is on the premise that examining what netizens search for or writes about online may provide us with a unique perspective on what they are currently interested in and curious about, hence providing us with more timely and relevant data analyzed for policy and decision-making processes concerning cyber security in the future.

THEORETICAL FRAMEWORK

This study anchored on the Routine Activity Theory (RAT) which states that for crime to take place three particular criteria must be present. These criteria are as follows: there must be a motivated offender, the presence of a suitable target, and the absence of a capable guardian. This theory shows that crime rates involving those that are related to online theft, fraud, and terrorism are not thoroughly affected by macro changes such as economic recessions and unemployment rates, but rather by the role played by the general lifestyle of an individual. The more a person exposed to criminal behavior in his or her everyday lifestyle, the higher the likelihood that individual will commit

detrimental activities.

Routine Activity Theory also suggests that gender affects both minor and severe victimization mediated by a measured lifestyle (Cullen et al., 2010). Gender does play a role in the opportunity to commit a crime since men have a more unsafe way of life than ladies. As indicated by the examination, young women tend to have a more organized way of life. Hence, there is more supervision of exercises, and in this way, young ladies are more protected more from illegal activities. This idea can extrapolate to age. As indicated by Cullen the group with the most affinity to wrongdoing is between the ages of 15 and 20. This behavior attributed to the way that young people have a greater amount of an unstructured way of life compared to grown-ups who hold occupations and have more obligations.

However, on the contrary, Elizabeth Groff offers a different view on Rational Activity Theory; she claimed as time spent away from home increases, so does the chance that the person will commit a crime (Groff, 2008). Furthermore, those who spent 70% of their time away from home were more involved with the police and more likely to commit crimes. Jasinki & Navarro, (2012) accept that there are plenty of motivated offenders, so people are more vulnerable when online, and they become "suitable targets." The lack of guardianship is also a concern for online users. Parents tend to assume that parental controls installed on a computer can act as a replacement guardian; however, this is not the case. Parental controls cannot filter what young people can post, explore or access on websites that they are allowed to access, such as email or social-networking sites such as Facebook.

Jeffery contends that RAT presents three components for crime, and fails to address social aspects of perpetrating a crime, such as personal education and socio-economic status. It states that theory is simply a description of the crime, not an explanation (Jeffery, 1993). While these issues have been brought up, it is certain that the Routine Activity Theory works when used to clarify why crime is more predominant for fixed groups, as well as why certain types of crime occur more often.

Therefore, RAT can be utilized as an approach to clarify why crimes happen, relative to Cybersecurity and to foresee who has the most astounding probability of carrying out crimes based on three factors: willing offender, a target, as well as inadequate parental supervision. In the assumption that, when confronted with the situation where they have the ability and occasion to commit a crime, people will pursue this route. There are downfalls to this theory like it does not take into account any other aspects of the individual committing the offense, nonetheless, with the use of RAT, we can efficiently create programs that make it more difficult for people to commit crimes and victimize or cause harm to our society.

RESEARCH QUESTIONS

The study examined the online documents on cyber security to further understand the different topics, related conversations, and documents found on the internet, as an input to the artificial intelligence collection of knowledge through unsupervised machine learning. The study seeks the answer to the following questions:

1. What are the topics and its trends related to cyber security for the last five years?
2. What are the countries that are active on cyber security on the web on these issues?
3. What are the underlying themes generated from the online documents?
4. What are the hidden issues or topics prevalent in the different articles as produced by the Latent Dirichlet Allocation Algorithm?

METHODS

Research Design

This study utilized *sequential exploratory design* using content analysis of online published articles. This type of design has two phases (qualitative and quantitative) but allows the theoretical perspective of the researcher to guide the study and determine the order of data collection. The heart of the qualitative data analysis is the task of discovering themes. By themes, it is meant the abstract, often ambiguous constructs which investigators identify before, during and after data collection. The discovery of these themes comes from reviewing the literature; richer writing produces more themes. Ideas originate from the characteristics of the phenomena and already agreed-upon professional definitions, from local common-sense constructs, and from researchers' values, theoretical orientation, and personal experience with the subject matter (Bulmer 1979; Strauss 1987; Maxwell 1996). The results from both methods are integrated together at the end of the study during the interpretation phase (Creswell, 2013). Furthermore, the published articles are derived from the following countries presented in the table below.

Top Twelve Countries with Documents Published Related to Cyber Security from April 2012 to April 2017

Ranking	Country	# of Documents
		Published
1	Singapore	100
2	United States	78
3	Malaysia	56
4	Canada	23

5	Netherlands	22
6	United Kingdom	19
7	Australia	18
8	India	18
9	Italy	9
10	Germany	9
11	France	8
12	Japan	5

Source: Google Trends

Research Method

The method used in this study was Gibb Sampling and Latent Dirichlet Allocation Algorithm.

Gibbs Sampling:

Gibbs sampling is one of the Monte Carlo Markov Chain (MCMC) technique suitable for the task. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values (Yildirim, 2012). The underlying logic of MCMC sampling is that we can estimate any desired expectation by ergodic averages. That is, we can compute any statistic of a posterior distribution as long as we have N simulated samples from that distribution:

$$E[f(s)]p \approx \frac{1}{N} \sum_{i=1}^N f(s^{(i)})$$

Where P is the posterior distribution of interest, $f(s)$ is the desired expectation, and $f(s^{(i)})$ is the i^{th} simulated sample of P . For example, we can estimate the mean by $E[x]p = \frac{1}{N} \sum_{i=1}^N x^{(i)}$. How do we obtain samples from the posterior distribution? Gibbs sampling is one MCMC technique suitable for the task. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values. For instance, consider the random variables $X_1, X_2,$ and X_3 . We start by setting these variables to their initial values $x_1^{(0)}, x_2^{(0)},$ and $x_3^{(0)}$ (often values sampled from a prior distribution q). At iteration I , we sample $x_i^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)})$, sample $x_2 \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)})$, and sample $x_3 \sim p(X_3 = x_3 | X_1 = x_1^{(i)}, X_2 = x_2^{(i)})$. This process continues until “convergence” (the sample values have the same distribution as if they were ample from the true posterior joint distribution (Yildirim, 2012).

Gibbs Sampler General Algorithm:

```

Initialize  $x^{(0)} \sim q(x)$ 
for iteration  $i = 1, 2, \dots$  do
     $x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)}, \dots, X_D = x_D^{(i-1)})$ 
     $x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)}, \dots, X_D = x_D^{(i-1)})$ 
    .
    .
    .
     $x_D^{(i)} \sim p(X_D = x_D | X_1 = x_1^{(i)}, X_2 = x_2^{(i)}, \dots, X_D = x_D^{(i-1)})$ 
end for
    
```

Latent Dirichlet Allocation Algorithm

Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus. The basic idea is that documents represent random mixtures over latent topics, where each topic is characterized by a distribution over words (Abramowitz & Stegun, 1966; as cited by Blei, Ng, & Jordan, 2003).

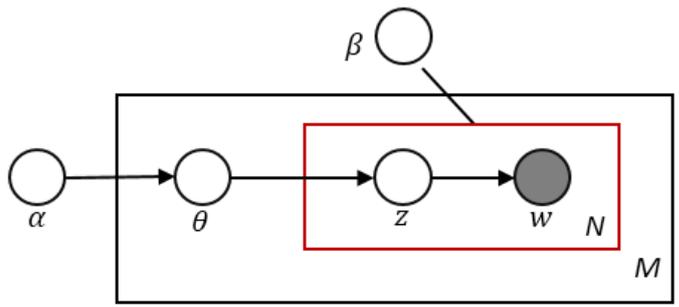


Figure 3: Plate Notation Representing LDA

With plane notation, the dependencies among the many variables are captured concisely. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. M denotes the number of documents, N the number of words in a document. Thus:

α is the parameter of the Dirichlet prior on the per-document topic distributions,

β is the paramant of the Dirichlet prior on the per-topic word distribution,

ϕ_m is the topic distribution for document m .

ϕ_k is the word distributed for the k ,

$simon$ is the topic for the n th word in document m , and

w_{in} is the particular word.

We w_{ij} are the only observable variables, and the other variables are latent variables. Mostly, the basic LDA model will be extended to a smoothed version to gain better results. The plate notation is shown on the right, where K denotes the number of topics considered in the model and φ is a $K * V$ (V is the dimension of the vocabulary) Markov matrix (transition matrix), and each row of which denotes the word distribution.

The generative process (algorithm):

1. Choose $\theta_i \sim Dir(\alpha)$, where $i \in \{1, \dots, M\}$ and $Dir(\alpha)$ is the Dirichlet distribution for parameter α
2. Choose $\varphi_k \sim Dir(\beta)$, where $k \in \{1, \dots, K\}$
3. For each of the word position i, j , where $j \in \{1, \dots, N_i\}$, and $I \in \{1, \dots, M\}$
 - a. Choose a topic $z_{ij} \sim Multinomial(\theta_i)$
 - b. Choose a word $w_{ij} \sim Multinomial(\varphi_{z_{ij}})$

(Note that the Multinomial distribution here refers to the Multinomial with only one trial. It is formally equivalent to the

categorical distribution.)

The lengths N_i are treated as independent of all the other data generating variables (w and z). The subscript is the often dropped, as in the plate diagram shown here.

Finally, this paper uses software to generate the output for Latent Dirichlet Allocation Algorithm such as: Google Trend, for identifying the number of recent documents related to cyber security for the past five years. Rstudio and R-programming, for LDA algorithm application using Python programs like gibbs sampler and lda.

Ethical Considerations

The information used in this research was raw data derived from the top twelve countries with various published articles related to cyber security. Additional data obtained via Google Scholar. To protect the authors from future predicaments, their identities will remain confidential. This study presents the philosophical views of the researcher, and its results need further validation and evaluation.

RESULTS AND DISCUSSION



Figure 4: Cyber Security Trends Worldwide

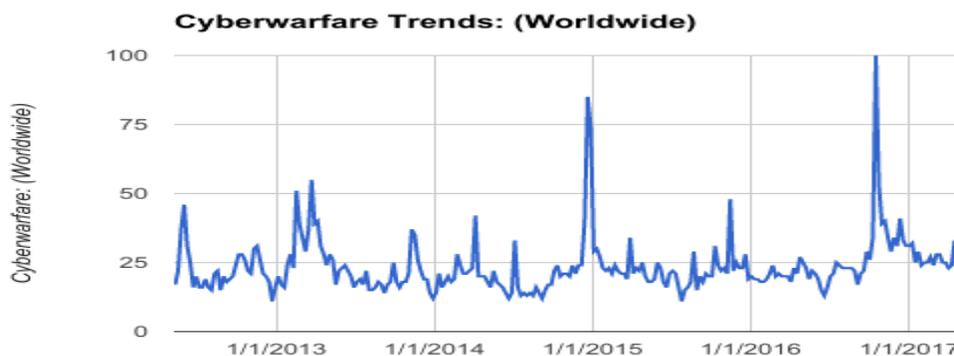


Figure 5: Cyber warfare Trends Worldwide



Figure 6: Computer Security Trends Worldwide

Cyber warfare and Computer security are terms that are synonymously popular with Cyber security. The above figures (fig. 4, 5, and 6) were typical indexed topics trending for the last five (5) years as illustrated in Google Trends. These suggests that there is a relevant association between these three phenomena. When searching for information or topics about cyber security on the Internet, there is a significant interdependence between cyber warfare and computer security. Cyber security is closely related to the domain of computer security as the latter is a subset of it, while cyber warfare is an underlying threat or attack that motivates the need for cyber security (USC, S. 3502. CNSSI-4009; Kissel, R., Ed., 2011; Walls, A.; E. Perkins; J. Weiss; Gartner, 2013; Hathaway, M. E., & Klimburg, A., 2012).

The figures represent parallel search trends on Cyber security and Cyber warfare. Both topics had an erratic series starting from 2012 to January 2017, the lowest point for cyber security was in December 2012 on a scale of 14, while a range of 8 for cyber warfare on December 28, 2013. The lowest level of cyber security/ Cyber warfare trend in the timeline points to the fact that there was a decrease in Cyber security activities for 2012 and 2013, the 0.03% decrease in email malware and the 58.7% decrease of malicious websites globally may have affected user activity (Symantec, 2012). Furthermore, 2012 witnessed the shutdown of the Cyber Security Act of 2012 in the United States (Couts, A., 2012; Rosenzweig, P., & Lieberman, S. J. (2012) thus, the number of articles and document search related to both topics declined.

However, as the timeline for both topics Cyber security and Cyber warfare progressed in the succeeding years, it followed

a constant and steady upward trend having the highest scale of 100 in February 2017 on both counts. It is not surprising to know that Cyber security and cyber warfare will achieve a rising trend. Considering that Cyber-offense and cyber-defense capacities, Ransom ware and extortion, Industrial IoT (Internet of Things) hacks, Internal threats and Business security spending all of which are related topics have the potential to grow in the coming years (Patterson, D., 2016).

Moreover, 2012 to 2017 has seen the significant rise of cyber warfare and alleged state-sponsored Cyber Attacks or Ransom ware on government institutions and industrial systems. Symantec reports confirm that in the last eight years, International bank heists, disrupted elections, and state-sponsored attacks will define the security landscape (Security in 2017 and Beyond) thereby increasing the popularity of search interest relative to Cyber security and Cyber warfare.

Surprisingly, Computer security has continued to plunge in its recent timeline. From a scale of 84 in 2012, its popularity to netizens fell downward to a level of 44 in 2017. It implies that fewer netizens were interested in the topic; as recent trends in ICT on that timeline were more focused on Cyber security and how to deal with cybercrimes (Cyber security current and emerging trends for 2017. (n.d.). The focus of Computer security is on securing the Confidentiality, Integrity, and Availability (CIA) of data. It is less concerned with cybercrime, online fraud and other forms of attack using the Internet of things (IoT). These contribute to the decline of its popularity as Netizens focus more on issues like Big Data, Social Media, Cyber terrorism; topics which are Cyber security-related, frequent and timely.

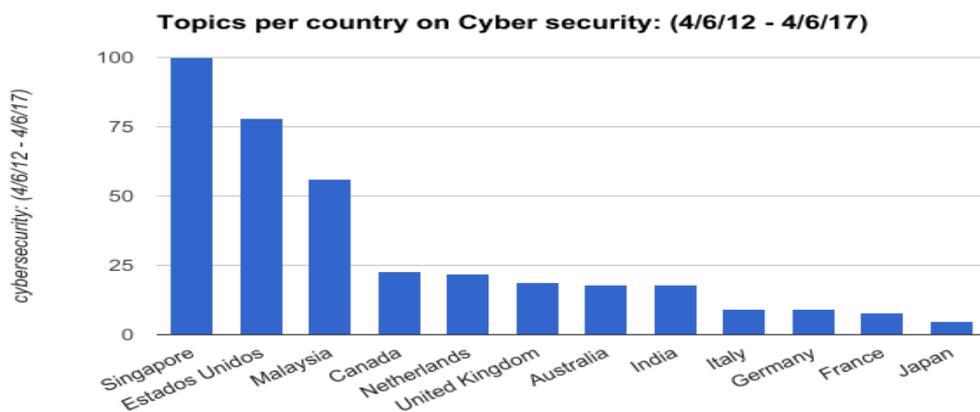


Figure 7: Topics per country on Cyber Security

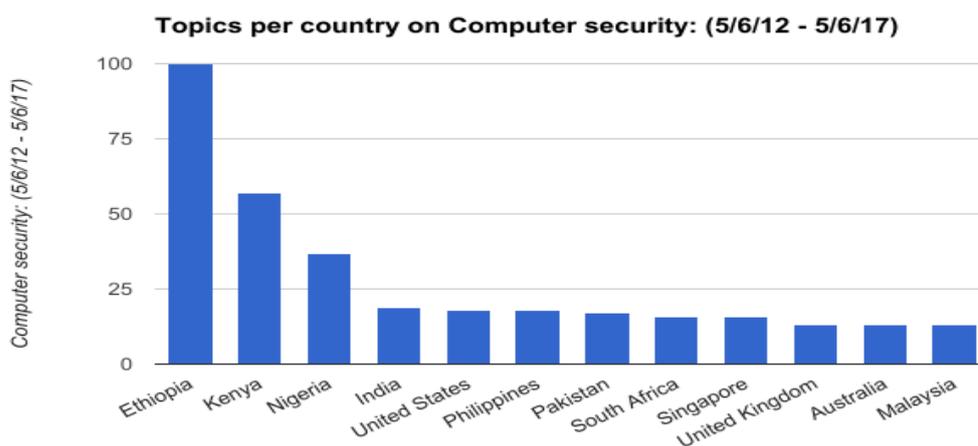


Figure 8: Topics per country on Computer Security

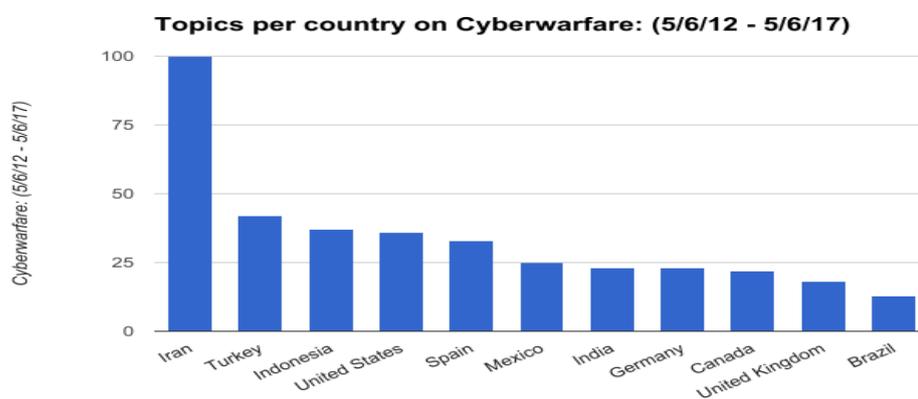


Figure 9: Topics per country on Cyber warfare

Countries with the most documents and conversations uploaded to the web relative to Cyber security. For the last five (5) years. Singapore (100), United States (78), Malaysia (56), Canada (23), Netherlands (22), United Kingdom (19), Australia and India (18), Italy and Germany (9), France (8) and Japan (5). The majority of these countries belonged to the NATO, or the North Atlantic Treaty Organization and the US allies like Japan,

Malaysia, and India. All of the mentioned countries are active in the development of their own National Cyber security strategy by NATO Cyber Defense cooperation agreement signed in February 2016. NATO's states and its allies reaffirmed their defensive mandate and recognized cyberspace as a domain of operations in which NATO must defend itself as efficiently as it does in the air, on land, and at sea. The

organization likewise emphasized the importance of protecting their national networks and enhancing information-sharing and mutual assistance in preventing, mitigating and recovering from cyber-attack (Cyber defense, 2017). It is also a fact that the United States, one of the leaders of NATO has an impending conflict with North Korea on their Nuclear Missile Program. This issue also has an impact on cyber security as there have been speculations that the recently foiled missile launch of the DPRK was due to cyber-attacks coming from the US (Waddell, K., 2017).

Countries with the most number of documents and conversations uploaded to the web about Computer Security. (figure 8) Ethiopia (100), Kenya (57), Nigeria (37), India (19), United States (18), and Philippines (18). Pakistan (17), South Africa (16), Singapore (16), United Kingdom (13), Australia (13), and Malaysia (13). These countries belong to East Africa, Asia Pacific, and the West. A few years ago Africa was referred to as a new safe harbor for cybercriminals (Kharouni, 2013). Over the last few years, computer security has gone from a concern to an issue of pressing importance in East Africa. Kenya, one of Africa's largest economies and considered as East Africa's central tech hub had estimated losses of more than 2 billion Kenyan shillings (US\$22.56 million) in 2013 due to cybercrime (Otieno 2014). On the contrary, Ethiopia, another east African nation has emerged as a paradox in East Africa about ICTs and cyber security (ITU 2015). The Ethiopian government had developed increasingly advanced legal and technical means to ensure greater control over communication networks and defend the country from cyber-attacks (Dlamini, I. Z., Taute, B., & Radebe, J., 2011). Similarly, the Philippines, Singapore, and Malaysia are all members of the ASEAN organization. As a developing region, it has the most advanced implementation of harmonized e-commerce laws. Nine of the 10 ASEAN countries have laws related to electronic transactions, while eight countries have laws concerned with

cybercrime. Surprisingly, ASEAN leaders have not yet established a unified cyber security framework (Elina, N., 2014). Thus, ASEAN countries share a common predicament. All are vulnerable to cyber-attacks because ASEAN's current cyber security initiatives are largely ineffective due to the absence of compliance mechanisms and divergences in security priorities and capabilities (Lee, S. 2016, April 4).

Countries with the most documents and conversations uploaded to the web relative to Cyber warfare. Iran(100), Turkey(42), Indonesia(37), United States(36), Spain(33), Mexico(25), India(23), Germany(23), Canada(22), United Kingdom(18), and Brazil(13). These were countries that belonged to the Top 20 most powerful military in the World. These countries are at the forefront of cyber warfare. The majority of these nations have experienced cyber-attacks and suspected of launching coordinated attacks against other rival countries. Iran became the most popular victim of a coordinated attack on its nuclear development program using a malware named STUXNET that targeted Supervisory control and data acquisition systems (SCADA). For the powerful and capable countries, Cyber warfare present an effective strategy to combat, neutralize or disrupt enemy states. It can bring down or topple enemy systems by acquiring control and hijacking critical network data or infrastructure. Cyber warfare has also led to Cyber terrorism coined as the most pressing national security issue facing the US and its allies today (Prichard, J. J., MacDonald, L. E., & Hunt, L., 2004). Similarly, Cyber warfare and information warfare employ information technology as an instrument of war to attack an adversary's critical computer systems (Hirsch, Kett, & Trefil, 2002). Cyber warfare and cyber terrorism are similar. Therefore, they exhibit the same characteristics. (1) premeditated and not merely acts born of rage, (2) political and designed to impact political structure, (3) targeted at civilians and civilian installations, and (4) conducted by ad hoc groups as opposed to national armies (Vatis, 2001).

Table 1. LDAGibbs Documents to Topics

Number	Document	Topic
1	AustraliaCyberSecurity.txt	1
2	SecuringCanada'sITInfrastructure.txt	1
3	JapansApproachCybersecurity.txt	1
4	StateofCybersecurity.txt	1
5	Malaysia'sNationalCyberSecurityPolicy.txt	1
6	ORGANISINGNATIONALCYBERSECURITY.txt	1
7	EnhancingASEAN-wideCybersecurity.txt	1
8	EnhancingInternationalCybersecurity.txt	4
9	CybersecurityinCivil.txt	2
10	ChinaandSoutheastAsia:OfflineInformation.txt	4

11	BeyondApplevsFBI.txt	2
12	DataProtectioninHyper-connectedAge.txt	2
13	social media and"Fake News".txt	2
14	PROACTIVECYBERSECURITY.txt	5
15	CyberattacksintheGulf.txt	1
16	LinkbetweenPiratedSoftwareandCybersecurity.txt	2
17	SecuringASEAN’sCyberDomain.txt	4
18	ThegovernanceofCybersecurity.txt	3
19	SiliconValley:MetaphorforCybersecurity.txt	4
20	NuclearLessonsforCyberSecurity?.txt	4
21	Cyberspace’sDynamicDuo.txt	5

The table above presents the documents collected from the web that was transcribed using LDA to topics. There were twenty-one (21) documents and each document focused on a particular topic which LDA classified accordingly. Like topics that were prominent and commonly discussed. Topic 1, with document 1

to 7 and 15. Topic 2 covers document 9, 11, 12, 13 and 16. Topic 4, comprised of document 8, 10, 17, 19 and 20; Topic 5 consists of documents 14 and 21; while Topic 3 in document 18 remains different from the others.

Table 2: Sample Terms and Frequencies

	Word	Frequency		Word	Frequency		Word	Frequency
1	security	1352	41	Actors	179	81	first	127
2	the	1316	42	Risk	179	82	for	127
3	cyber	1149	43	Sector	178	83	however	127
4	cyber security	948	44	Network	177	84	issues	127
5	information	707	45	Protection	176	85	within	127
6	data	560	46	State	176	86	code	126
7	government	514	47	Strategy	174	87	must	126
8	national	497	48	Countries	173	88	response	126
9	computer	376	49	Australian	172	89	networks	125
10	public	298	50	Yes	169	90	users	125
11	systems	294	51	Active	167	91	number	123
12	private	293	52	Need	166	92	time	123
13	identity	288	53	Measures	162	93	activities	122
14	international	275	54	Criminal	158	94	military	122
15	states	272	55	Industry	158	95	important	121
16	infrastructure	270	56	Level	158	96	incident	121
17	one	260	57	Well	156	97	technical	121

18	act	259	58	Companies	155	98	united	121
19	this	258	59	Part	155	99	person	121
20	attacks	256	60	Cooperation	151	100	threat	119
21	law	247	61	Approach	148	101	available	117
22	access	241	62	Case	148	102	Malaysia	116
23	use	240	63	Provide	148	103	analysis	115
24	defense	239	64	Malware	148	104	relevant	115
25	threats	232	65	Order	147	105	different	114
26	critical	223	66	Research	145	106	focus	114
27	software	223	67	Even	144	107	war	114
28	new	219	68	Attack	143	108	nuclear	113
29	governance	212	69	Report	142	109	two	110
30	internet	212	70	Development	140	110	art	110
31	see	203	71	Example	138	111	costs	109
32	section	197	72	Power	138	112	digital	109
33	technology	197	73	Legal	137	113	global	109
34	services	194	74	Cybercrime	135	114	including	109
35	cyberspace	192	75	Framework	134	115	business	108
36	management	189	76	Online	131	116	pirated	108
37	service	188	77	Regulation	131	117	secure	107
38	system	188	78	World	130	118	infrastructures	107
39	used	186	79	Proactive	129	119	personal	106
40	policy	185	80	Sharing	129	120	role	106

Table 2 above, presents the collection of words from the twenty-one (21) documents with its frequency. The number of times a word (term) appears in the records and is sorted by the most number of occurrences to the least number of occurrences.

The researcher chose to present one hundred twenty (120) words drawn from a sample of 10,109 terms in the 21 documents as shown in Table 1.

Table 3: LDAGibbs Topics to Terms

Document	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	security	software	the	cyber	cyber security
2	cyber	data	data	states	private
3	cyber security	malware	identity	the	defense
4	the	the	security	war	computer
5	information	security	act	nuclear	yes

6	national	pirated	information	military	active
7	government	costs	public	cyberspace	proactive
8	international	idc	section	attacks	companies
9	systems	users	governance	air	law
10	Australian	source	service	attack	data
11	strategy	Singapore	computer	power	access
12	new	government	infrastructure	innovation	see
13	threats	online	protection	strategic	legal
14	research	consumers	cyber security	defense	sector
15	technology	internet	risk	united	note

Latent Themes	Intergovernmental strategies against cyber security threats	Government policies on online consumer protection against Piracy	Protection of computer infrastructures in the government	Military defense against innovative cyber attacks	Legal and ethical issues on cyber security
----------------------	---	--	--	---	--

Table 3 shows the groupings as identified by the algorithm per topic. These were arranged by topics 1 to 5 as prescribed in the algorithm. It also presents the underlying themes drawn from the philosophical views of the researcher derived from the words of each topic.

Table 4: LDAGibbs Topic Probabilities

Document	Topic1	Topic2	Topic3	Topic4	Topic5
1	0.834964	0.077699	0.046412	0.025208	0.015718
2	0.757713	0.032731	0.094808	0.056433	0.058315
3	0.682731	0.063588	0.107764	0.10241	0.043507
4	0.427498	0.424869	0.064237	0.056724	0.026672
5	0.826298	0.030104	0.105882	0.016609	0.021107
6	0.841578	0.01214	0.108649	0.023065	0.014568
7	0.500739	0.159527	0.04579	0.237814	0.05613
8	0.31049	0.057343	0.11049	0.439161	0.082517
9	0.27031	0.474151	0.116691	0.094535	0.044313
10	0.141335	0.18582	0.049583	0.588508	0.034754
11	0.207381	0.462214	0.105448	0.063269	0.161687
12	0.113433	0.48209	0.177612	0.076119	0.150746
13	0.192073	0.515244	0.079268	0.137195	0.07622
14	0.1915	0.057224	0.078861	0.044177	0.628237
15	0.315485	0.083936	0.117221	0.251809	0.231548
16	0.055259	0.889481	0.031628	0.010306	0.013326
17	0.303419	0.119658	0.094017	0.433761	0.049145
18	0.225607	0.029165	0.718011	0.013635	0.013583
19	0.129663	0.043111	0.024871	0.755928	0.046427
20	0.19386	0.038279	0.035247	0.702672	0.029941
21	0.378088	0.060394	0.07986	0.064637	0.41702

Latent Themes	Intergovernmental strategies against cyber security threats	Government policies on online consumer protection against piracy	Protection of computer infrastructures in the government	Military defense against innovative cyber attacks	Legal and ethical issues on cyber security
----------------------	---	--	--	---	--

Table 4 shows the reliability of the topics using Gibbs sampling per documents and topics. It indicates that the identification of topics from table 1 and table 4 was very consistent and illustrated in table 3 showing the collections of words per topic. The consistency and reliability of document 1 to topic 1 had 83.49%, document 2 to topic 1 was 75.77%, document 3 to topic 1 68.27%, document 4 to topic 1 42.74%, document 5, 6, 7 and 15 talk in common to Topic 1 having 82.62%, 84.15%, 50.07% and 31.54%.

From these results, we can surmise that there is a high indication that the latent theme is describing “Intergovernmental strategies against cyber security. These means that the discussions found in the articles in Topic 1 focus on the strategies that governments around the world adopt relative to cyber security. The issue of how to protect national security and economic growth of developing countries using appropriate cyber security strategies is a vital problem that needs a solution. Governments should sustain a coherent cyber security approach using a national strategy, enforceable at a national level and compatible at the international arena (Ghernouti-Hélie, S. 2010). The development of a national cyber security strategy follows a life cycle which divides into two phases; first the development and execution of the plan, then the evaluation and adjustment of the policy. Further, the

strategy follows three approaches; the linear approach where the strategy is developed, implemented, evaluated and eventually terminated (or replaced). The lifecycle approach where the output of the evaluation phase will be used to maintain and adjust the strategy itself; and the Hybrid Approach which ensures continuous improvement cycles on different levels. A sound strategy must define vision, scope, objectives, and priorities (ENISA, 2016).

Similarly, national cyber security policies can take another approach, thru the use of so called Action Themes. Each theme is supported by governmental actions to ensure the growth and prosperity of the Nation. The themes involve Government and business leaders jointly setting strategic agenda through annual Cyber Security meetings. Funding research to detect, deter and respond to cyber security threats. Constructing cyber capacity to prevent and shut down safe havens for criminals. Establishing Cyber Security Center for innovation, Research, and Development. Addressing the critical shortage of skilled cyber security professionals and improving national cyber security awareness. These Strategy initiatives are reviewed and updated annually and are changed every four years (Australia's Cyber Security Strategy, (n.d.).

Taking into account that usually 85–90 percent of the cyber infrastructure is managed by the private sector, it is vital to ensure uninterrupted operation of this infrastructure and its resilience to cyber-attacks. For this reason, the industry should be involved both in strategic planning documents and in the real process of cooperation. Strategic alliances between public and private parties would inevitably be the next necessary step to fight cybercrime and maintain cyber security (Tropina T., Callanan C. 2015).

The unification of national cyber security strategies has become a crucial concern. For instance, if the principles of strategies, in countries differ, it will be hard to find effective mechanisms to fight cyber threats when cooperating among themselves. Therefore, it is necessary to unify the strategies as much as possible so that countries have a similar understanding and similar values, and find a dialogue as soon as possible. Thus, it is vital to cooperate regarding the global phenomenon and unify the strategies to manage global cyber threats (Štivilis, D., Pakutinskas, P., Laurinaitis, M., & de Castel, I. M. V. (2017).

When considering efficient mechanisms, cyber security strategy, and an overview of the national situation, including the cybersecurity situation should be substantial. Such an overview would enable an understanding of a real situation in the country and, based on the presented analysis, would possibly adapt other parts of the strategy according to the specific social-features. These are crucial because transplanting security threats that appear in other strategies but are not germane to the country formulating the strategy may do more harm than good by diverting national resources (Klimburg, A. 2012).

Finally, a robust and functional legislation is necessary to detect crimes that violate human rights and freedoms. Amidst growing security problems of the digital space (computer crime, organized crime, terrorism) and the significance of the global (as well as national) ICI for society it will be necessary to define a legal framework for the protection of digital space both at the international and national level. The interests and rights of various individuals and organizations coincide in this area. For example, proposed measures (legal arrangements) have to take into account not only interests of the state and ICT system owners but also the rights of users and those parties whose data ICT systems contain (Organization for Economic Co-operation and Development., 2002).

Document 9, 11, 12, 13 and 16 have similar topics about (Topic 2) having 47.41%, 46.22%, 48.20%, 51.52% and 88.94% respectively. From these results, we can deduce that the theme is related to “Government policies on online consumer protection against piracy” This refers to the National regulations on how to prevent digital piracy and proprietary infringement thru effective implementation of cyber security laws that ensure online consumer protection. Today, Online piracy of digital content including that of music, movies, software, games, and other products continues to be a huge issue for businesses and public policymakers (Girona, J., Petrescu, M., & Korgaonkar, P. K. 2017). The digitization of media goods effectively weakened copyright laws across the globe by making it easy for ordinary consumers to share media files from computer to computer illegally. Preventing unauthorized downloading and other forms of digital piracy has been a persistent challenge. Global music piracy causes \$12.5 billion of economic losses every year, 71,060 lost American jobs, and a loss of \$422 million in tax revenues (RIAA 2014). Partly these might have been the cause of what the ‘Celestial Jukebox” or the distribution of digital music online and transforming it into cloud-based music services (Burkart, P., 2014).

The companies proposed a new copyright directive and intellectual property enforcement strategies to combat the growing menace. One is the integration of automated anti-piracy systems (‘AAPSs'). AAPS is capable of recognizing content which right-holders have already identified as their own and respond based on standing instructions from the holder whether the platform should permit the use or block it. Another feasible approach is to use cyberlockers these are encrypted cloud storage spaces where people can house their data on remote servers. In other States, the refinement of copyright reform policies and strategies advocated. Through the strong-arm of the law and with the full support of the government, digital piracy sites online are detected and identified then punished. Such was the case of the two popular file sharing sites that were prosecuted and shut down, namely www.Megaupload.com and the Thepiratebay.com; these sites promoted regular download of copyright software for free.

It is worth noting that rights holders have many options for mitigating the impact of piracy on sales, although these strategies often come at a cost to the firms regarding undermining the effectiveness of their existing marketing strategy. Eventually, government interventions can mitigate the impact of piracy on sales. Although there is little evidence that government action reduces the overall range of content available through piracy channels, there is evidence that government anti-piracy interventions can be effective at changing user behavior if they sufficiently increase the search and transactions costs associated with finding piracy content. Specifically, government actions only increase the inconvenience of piracy. Loosely enforced laws have only transient effects on piracy and legal consumption (Danaher, B., Smith, M. D., & Telang, R., 2017).

Whereas, Document 18 had similar Topic about (Topic 3). The theme relates to the Protection of computer infrastructures in the government having a 71.80% reliability rate. These means the Topic is referring to policies, tools or laws that protect ICT infrastructure, and other assets. It regularly pertains to critical infrastructure, physical processes that are controlled by networked computers. Frequently, the failure of these types of systems results in a high socioeconomic impact. Today's critical infrastructures (CI) are systems that are complex interconnected industrial systems that, in recent years, have incorporated information and communications technologies to its operations. The most significant threats to CI's today are Hacking, Ransomware and Cyber terrorism. Critical infrastructures of government agencies, private companies should treat it as a major business or national defense problem and not as a simple IT issue. The growing concern for IT security is so enormous that the present security mechanisms are insufficient. Organizations and device vendors should be creative in developing secure software, and the increased level of protection needs regulation to compel internet users to provide protection. Furthermore, future changes may also include the elimination of password protected sites and replaced by advanced biometric software, like fingerprint readers, iris scans and face recognition. Even more secure may be the implementation of adaptive and behavior-based authentication.

Document 8, 10, 17, 19, 20 had similar topics about (Topic 4) Military defenses against innovative cyber-attacks having 43.91, 58.85%, 43.37, 75.59% and 70.26% reliability rates. This theme refers to the new and ingenious ways of mitigating and destroying cyber-attacks using military assets and resources. For example, when the US Department of Defense (DoD) learns of malicious cyber activities that will affect national and economic security or public safety, DoD coordinates with Department of Homeland Security (DHS) and Federal Bureau of Investigation (FBI) to share threat information and technical indicators of a potential attack. Another way is to establish an enterprise-wide cyber modeling and simulation capability. The DoD works in collaboration

with the intelligence community to develop the data schema, databases, algorithms, and modeling and simulation (M&S) abilities critical to evaluate the adequacy of digital operations. Additionally, the success of military defense does not depend alone on its sheer power and might, but instead the success of these systems lies in maintaining robust worldwide alliances and organizations to discourage shared threats and uplift universal security and stability.

The themes in documents 14 and 21 commonly describe the same topic (Topic 5) Legal and ethical issues on cyber security having a reliability of 62.82% and 41.70%. This theme refers to the issues that affect the implementation of Legal and ethical policies relative to cyber security. Studies reveal that the popularly known legal issues relative to cyber security are freedom of information, Liability concerns, and information sharing. Freedom of Information becomes a legal issue when private sector entities become hesitant to share cybersecurity-related information with the government because of the possible disclosure of this information to the public under the Freedom of Information Act (FOIA). Also, concerns also exist that sharing of cybersecurity information may facilitate access to proprietary and confidential business information by competitors. In some cases, private sectors expressed fears that the government may use information obtained for cybersecurity purposes for non-cyber security reasons, such as regulatory actions.

Another important issue in cyber security is liability concerns. This issue starts when in the context of cyber threats to critical infrastructure, a regulated entity that fails to adequately secure its information infrastructure as required under a federal regulatory scheme would be liable for a cyber-incident that causes harm to customers or other third parties. Second, entities that are not subject to regulation under a federal system may not be subject to negligence per se. However, the performance standards or other requirements imposed under that scheme may still be a liability for negligence if such provisions establish an applicable level of care that the non-regulated entity would abhor against in a private civil suit. Because of the effect that a regulatory scheme can have on civil liability, proposals to regulate the cybersecurity of critical infrastructure may also propose limits on responsibility for regulated entities. The scope of such limits may range from complete immunity from private suits to minor restrictions such as prohibitions against the granting of corrective damages. Such points of confinement on liability may be dependent upon a person's satisfaction of its regulatory obligations, to create a further incentive for compliance.

Lastly, another type of legal issue in cyber security is information sharing. It is an ethical issue when private sector entities may wish to share information with one another about threats they have faced or are currently facing. Organizations that share information are concerned that sharing or receiving such information may lead to civil and criminal liability, or that

shared information may contain proprietary or confidential information disclosed to competitors or government regulators.

CONCLUSION

The corpus of articles collected through Google trends that were mined and analyzed using Latent Dirichlet Allocation Algorithm generated results which implied that the discussion on Cyber security has a significant and profound bearing as it had the potential to be one of the prevalent issues of global importance today, tomorrow and beyond. The underlying themes that were generated by the algorithm revealed vast opinions on different areas encompassing Legal, Ethical, Strategic, Defensive and Governmental Applications of Cybersecurity. It also disclosed ideas from various countries and organizations such as UN, NATO, and ASEAN affiliated countries. Moreover, citizen's views as reflected in written articles suggest that the topic is universal and citizens around the world have wealthy and diverse opinions on what, how, where and when cyber security is needed and applied. Furthermore, this study provides an avenue to discover the inherent qualities of the phenomena used as the basis for future researchers.

ACKNOWLEDGMENT

This paper was made possible through the support of the following faculty in Leyte Normal University, Tacloban City, namely, Dr. Las Johansen Caluza, Dr. Rommel L. Verrecio, and our university president Dr. Jude A. Duarte.

REFERENCES

- [1] Adams, S., Brokx, M., Dalla Corte, L., Galic, M., Kala, K., Koops, B. J., ... Skovránek, I. (2015). The governance of cyber security: A quick comparative scan of approaches in Canada, Estonia, Germany, the Netherlands and the UK. Tilburg University
- [2] Ang, B. (2016). Beyond Apple vs. FBI: Implications for Singapore's Smart Nation Project.
- [3] Backman, S. (2015). Organising National Cybersecurity Centres. *Information & Security*, 32(1)
- [4] BHashim, M. S. (2011, June). Malaysia's national cyber security policy: The country's cyber defense initiatives. In *Cybersecurity Summit (WCS), 2011 Second Worldwide* (pp. 1-7). IEEE.
- [5] Brookes, C. (2015). Cyber security: Time for an integrated whole-of-nation approach in Australia. *Indo-Pacific Strategy*
- [6] Bulmer, M. (1979). Concepts in the analysis of qualitative data. *The Sociological Review*, 27(4), 651-677.
- [7] Burkart, P. (2014). Music in the Cloud and the Digital Sublime. *Popular Music and Society*, 37(4), 393-407.
- [8] Cheong, D. D. (2012). Cyberattacks in the Gulf: lessons for active defense.
- [9] Chong, A. China, and Southeast Asia: Offline Information Penetration and Suspicions of Online Hacking. A View from Singapore
- [10] Choo, K. K. R. (2011). The cyber threat landscape: Challenges and future research directions. *Computers & Security*, 30(8), 719-731.
- [11] Coutts, A. (2012, August 02). Senate kills Cybersecurity Act of 2012. Retrieved May 09, 2017, from <https://www.digitaltrends.com/web/senate-votes-against-cybersecurity-act-of-2012/>
- [12] Craig, A., Shackelford, S., & Hiller, J. S. (2015). Proactive Cybersecurity: A Comparative Industry and Regulatory Analysis
- [13] Craigen, D., & Whyte, D. (2013). Securing Canada's Information-Technology Infrastructure: Conference
- [14] Cullen, F., Henson, B., Reyns, B., Wilcox, P (2010). "Gender, Adolescent Lifestyles, and Violent Victimization: Implications for Routine Activity Theory," *Victims and Offenders* Vol. 5 pg 303-328, 1 October 2010
- [15] Cybersecurity: A global issue demanding a global approach | UN DESA Department of Economic and Social Affairs. (n.d.). Retrieved August 21, 2017, from <http://www.un.org/en/development/desa/news/ecosoc/cybersecurity-demands-global-approach.html>
- [16] Cyber security current and emerging trends for 2017. (n.d.). Retrieved May 10, 2017, from <https://www.globalsecuritymag.com/Cyber-security-current-and,20161220,67853.html>
- [17] Dan Patterson, December 13, 2016. Experts predict 2017's biggest cybersecurity threats. Retrieved May 09, 2017, from <http://www.techrepublic.com/article/experts-predict-2017s-biggest-cybersecurity-threats/>
- [18] Danaher, B., Smith, M. D., & Telang, R. (2017). Copyright enforcement in the digital age: empirical evidence and policy implications. *Communications of the ACM*, 60(2), 68-75.
- [19] Elina, N. (2014). Securing ASEAN's cyber domain: the need for partnership in strategic cyber security.
- [20] Gantz, J. F., Florean, A., Lee, R., Lim, V., Sikdar, B., Lakshmi, S. K. S., ... & Nagappan, M. (2014). The Link Between Pirated Software and Cybersecurity Breaches. International Data Corporation.
- [21] Gironda, J., Petrescu, M., & Korgaonkar, P. K. (2017). Piracy, Price, and Word of Mouth: An Equity Theory Examination of Consumer Digital Piracy Rates—An Abstract. In *Creating Marketing Magic and Innovative Future Marketing Trends* (pp. 1237-1238). Springer, Cham.
- [22] Groff, E. (2008). Adding the temporal and spatial

- aspects of routine activities: A further test of routine activity theory. *SECURITY JOURNAL*, 21(1-2), 95-116. doi:10.1057/palgrave.sj.8350070
- [23] Hathaway, M. E., & Klimburg, A. (2012). Preliminary considerations: On national cyber security. *National Cyber Security Framework Manual*. NATO Cooperative Cyber Defence Centre of Excellence, Tallinn
- [24] Heintz, C. H. (2013). Enhancing ASEAN-wide cybersecurity: time for a hub of excellence?.
- [25] Hirsch, E. D., Kett, J. F., & Trefil, J. S. (2002). *The new dictionary of cultural literacy*. Houghton Mifflin Harcourt.
- [26] Jeffery, C.R. (1993). Obstacles to the development of research in crime and delinquency. *Journal of Research in Crime & Delinquency*, 30:491-497.
- [27] Kao, J. (2014). Silicon Valley: Metaphor for Cybersecurity, Key to Understanding Innovation War. *Cyber Analogies*, 2013-2014.
- [28] Kharouni, L. 2013. "Africa: A New Safe Harbor for Cybercriminals?" Trend Micro Incorporated Research Paper. www.trendmicro.co.uk/media/misc/africanew-safe-harbor-for-cybercriminals-en.pdf. ITU. 2015. ICT Statistics. www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.
- [29] Kissel, R. (Ed.). (2011). *Glossary of key information security terms*. Diane Publishing.
- [30] Lee, S. (2016, April 4). ASEAN Cybersecurity Profile: Finding a Path to a Resilient Regime. Retrieved May 11, 2017, from <https://jsis.washington.edu/news/asean-cybersecurity-profile-finding-path-resilient-regime/>
- [31] Manley, M. (2015). Cyberspace's Dynamic Duo: Forging a Cybersecurity Public-Private Partnership. *Journal of Strategic Security*, 8(5), 85-98.
- [32] Maurer, T. (2011). Cyber Norm Emergence at the United Nations. An Analysis of the Activities of the UN Regarding Cyber-Security. Belfer Center for Science and International Affairs, 6668
- [33] Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (Vol. 41). Sage Publications.
- [34] Muniandy, L., & Muniandy, B. (2012). State of cyber security and the factors governing its protection in Malaysia. *International Journal of Applied Science and Technology*, 2(4).
- [35] Neubronner, S. (2017). Social Media and "Fake News": Impact on Social Cohesion in Singapore.
- [36] Nitta, Y. (2013). Japan's Approach towards International Strategy on Cyber Security Cooperation. Retrieved September, 13, 2014.
- [37] Nye Jr, J. S. (2011). Nuclear lessons for cyber security. *AIR UNIV PRESS MAXWELL AFB AL*.
- [38] Otieno, J. 2014. "Worries over New Avenues of Cybercrime." *The East African*, September 22. www.theeastafrican.co.ke/news/Worries-over-new-avenues-of-cybercrime/-/2558/2461630/-/vs7k0z/-/index.html.
- [40] Prichard, J. J., MacDonald, L. E., & Hunt, L. (2004). Cyberterrorism: A study of the extent of coverage in computer security textbooks. *Journal of Information Technology Education*, 3.
- [41] Rosenzweig, P., & Lieberman, S. J. (2012). Cybersecurity Act of 2012: Revised Cyber Bill Still Has Problems. *Heritage Foundation Issue Brief*, (3675).
- [42] Security in 2017 and Beyond: Symantec's Predictions for the Year Ahead. (n.d.). Retrieved May 10, 2017, from <https://www.symantec.com/connect/blogs/security-2017-and-beyond-symantec-s-predictions-year-ahead>
- [43] Štītīlis, D., Pakutinskas, P., Laurinaitis, M., & de Castel, I. M. V. (2017). A MODEL FOR THE NATIONAL CYBERSECURITY STRATEGY. THE LITHUANIAN CASE. *Journal of Security & Sustainability Issues*, 6(3).
- [44] Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
- [45] Tan, T. B. (2016). We, Citizens of Smart Singapore: Data Protection in Hyper-connected Age.
- [46] Tropina T., Callanan C. 2015. Self- and Co-regulation in Cybercrime. *Cybersecurity and National Security*. Springer, p. 25.
- [47] Waddell, K. (2017, March 05). Is It Wise to Foil North Korea's Nuclear Tests With Cyberattacks? Retrieved May 10, 2017, from
- [48] Yilmaz, S. (2013). Enhancing international cyber security: will the UN reach a deal?.