# ANTICIPATORY CYBER SECURITY RESEARCH: AN ULTIMATE TECHNIQUE FOR THE FIRST-MOVE ADVANTAGE

[1] Bharat S.Rawal, [2] Songjie Liang, [3] Alae Loukili, [4] Qiang Duan

[1&4]*Department of Information Sciences and Technology. Penn State Abington, Abington, PA 19001 USA.*
[2]*SoroTek Consulting, Inc Potomac, Maryland, MD 20854, USA*
[3]*Department of Mathematics and Computer Science, Ohio Dominican University 1216 Sunbury Road, Columbus, OH 43219, USA.*

***Abstract*: Across all industry segments, 96 percent of systems could be breached on average. In the game of cyber security, every moment a new player (attacker) is entering the game with new skill sets. An attacker only needs to be effective once while defenders of cyberspace have to be successful all of the time. There will be a first-mover advantage in such a chasing game, which means that the first move often wins. In this paper, in order to face the security challenges brought in by attacker's first move advantage, we analyzed the past ten years of cyber-attacks, studied the immediate attack's pattern and offer the tools to predict the next move of the cyber attacker.**

***Keywords:* Cyber security; Vulnerability; State Based Stochastic Model; Anticipatory Research**

## 1. Introduction

Across all industry segments, 96 percent of systems have been breached on average [1]. Attacks against U.S. businesses and governments are commonplace, with an estimated 100 million attempts each day [2]. Cyber security has become a daunting problem for businesses, government administrators, and millions of end users.

DOI: 10.18421/TEM51-01
https://dx.doi.org/10.18421/TEM51-01

**Corresponding author: Bharat S. Rawal**.
**Qiang Duan:** Department of Information Sciences and Technology. Penn State Abington Abington, PA 19001 USA.
**Songjie Liang:** SoroTek Consulting, Inc Potomac, Maryland, MD 20854, USA.
**Alae Loukili:** Department of Mathematics and Computer Science, Ohio Dominican University 1216 Sunbury Road, Columbus, OH 43219, USA.



As new techniques are developed to prevent or mitigate cyber security attacks, attackers are persistently evolving new methods to circumvent these new measures. The reality is that there are no easy or perfect answers to a cyber security threat. Cyber security as an issue is too comprehensive; there are too many variables in security, too many known and unknown vulnerabilities in hardware and software.

We propose a mathematical model for a prediction of future cyber attacks. We will use the most prevalent sales forecasting models Monte Carlo. Models are at the heart of economics, the driver of many downstream business decisions. We have conducted research to analyze hundreds of the thousand past events of cyber-attacks (from 2000 to 2015), collected from various databases, research publications and information posted on the web. Also, we will mine recent tweets posted on Twitter (including tweets from leading security experts), and compare with data sets for the past fifteen years. We also included the "inverse problems method" by combining prior information with new data obtained by measuring some observable parameters (from the recent stream of social media and the other security threat reporting systems).

Our proposed prediction system will increase cyber security awareness and vigilance of private public, or government organizations. In monetary terms, it is estimated that the cyber security market will be worth $170.21 Billion by 2020 [3]. If our model can offer just only one percent savings, it will save $1.7.Billion. Existing forecasting models: Most of the current approaches are based on the time series and Moving Average, Hidden Markov Model and State Based Stochastic Model.

The remainder of the paper is organized as follows. Section II presents related work; Section III describes various vulnerability taxonomies; Section IV discusses the attacker types and their characteristics; Section V represents the threat from insiders; Section

VI shows past data breaches reported from Jan 2005 to August 2015; Section VII represents the study of recent attacks; Section VIII discusses a study of recent tweets; IX The cyber threat anticipatory system; X presents the anticipation for cyber threats; XI Illustrates the cyber threat anticipatory system; XII discusses the results obtained; and Section XII concludes the paper.

## 2. Related work

In recent years, researchers have studied the numbers of empirical and theoretical aspects of data breaches, such as the effect of breaches on a firm's stock market price [4]. The 2015 Data Breach Investigations Report investigates nine common patterns and analyzes all possible types of data breaches quantitatively. This effort has been led by Verizon alone with 70 organizations of around the world [5]. The survey was conducted by Ponemon for 567 executives in the United States on how a company prepares itself for the moment when the big data breach occurs [6]. The Breach Portal Notice to the Secretary of HHS Breach of the Unsecured Protected Health Information U.S. Department of Health and Human Services Office for the Civil Rights report records each incident of the data breach affecting >=500 people among Health Services since 10/21/2009. Ocrportal describes information about breach types and devices [7]; and the Symantec reports the lists of various security issues and hacker's attacks comprehensively [8].

The numbers of attacks such as phishing are astonished high with 5 out of every 6 for big firms with more than 2500 people. It is 40% increase in 2014 compared to the previous year. For the small and medium companies, the attacks increase 26% and 30% respectively [9]. Numerous recent national policy documents have stressed the standing of cyber security to the wellbeing of the contemporary world [10, 11]. The leader's nationwide system to protect cyberspace [10] describes the urgencies for an answer, the decrease of threats and weaknesses, consciousness and training, and general security and global cooperation. The vital constituent of information and communications influences all of the other critical US organizations [12]. The open web application security project also lists the top ten vulnerabilities of the year for web-based applications [13]. Jagannathan et al. discusses risk management and early detection of risks based on insider threat [14]. Tittel [15] clarifies about integrated threat management and why it is vital to discuss it in his paper. Wu et al. [16] recognize a forecast model built on assimilating environmental features and attack diagrams in a Bayesian system. Anusha et al. [17] examine numerous techniques like unimodal and multimodal for improved security. They deliberated round validation at the beginning of the examination and the system user checks. Axelrad et al. [18] presented a Bayesian network model for the inspiration and psychology of the malicious insider. Khan and Hussain [19] recognized associations between attack probability and vulnerability. Moore et al. [20] presented a modeling and simulation groundwork, built on the system dynamics approach to testing the efficiency of insider threat discovery controls.

## 3. Vulnerability taxonomies

A vulnerability is defined as the absenteeism or flaws of a protection that could be exploited. A vulnerability resulting from development is a fault, bug or defect [21]. However, vulnerabilities, which result from design or configuration mistakes are not faults. Ozment and Andy note that not all vulnerabilities are faults, and not all faults are vulnerabilities [22]. Weber, Karger, and Partaker claim that the significant difference between faults and vulnerabilities is that, vulnerabilities are exploitable, but faults are not necessarily exploitable [23].

In the field of computer security, a vulnerability when exploited permits an attacker to decrease a system's information assurance. Prakash, Atul, and Rudrapatna, state that vulnerabilities are made up of three principal components; a system weakness or defect, invader access to the fault, and invader capability to exploit the flaw[24]. Vulnerability classifications can be viewed from different angles. Howard et al. describe traditional vulnerability classifications, and state that they can be useful for descriptions of what went wrong [25]. There are several classification techniques as described below:

1. Classification by Software Development Life Cycle (SDLC) Phase: This method attempts to classify vulnerabilities according to in what phase they were introduced to the software product [26].
2. Classification by Genesis: Vulnerabilities, as flaws, can be classified as intentional or unintentional faults. Intentional defects are further classified as malicious or non-malicious ones [27, 28].
3. Classification by Location in Object Models: This classification tries to group vulnerabilities according to which model entity they belong to. Classifying vulnerabilities using the OSI reference model in networking [29] is an example of this type of classification.

4. Errors or Mistakes: The classification of the error is based on how it triggered the vulnerabilities, the nature of their effect and the type of alteration or solution made to remove the error [30].
5. Classification by Enabled Attack Scenario: A particular type of vulnerabilities that are very descriptive and accurate enough to allow a set of attacks. For example, "Cross Site Scripting" ("XSS") attack allows injection of malicious code into the content available on the Web browser [31].

## 4. Attacker types and their characteristics

We reviewed the literature on cyber-attackers based on a search that included the following keywords: "cyber criminals", "insiders", and "hackers". Adams, Mackenzie, and Maged Makramalla [32] expand keyword hunt to support the terminology variations in existing literature when representing individuals or groups that commit cyber-attacks. Here we focus on cyber-attackers to identify attacker types and their motivations, resources, and knowledge/skills. Recognizing the attacker types is vital in developing more precise profiles when making and implementing solutions planned to decrease cybercrimes [33]. Adams, Mackenzie, and Maged Makramalla classify the following eight types of cyber-attackers [32].

Attacker types and their characteristics based on an exhaustive search of the available resources on the web and published studies, we could not recognize a single application of cyber-attacker characteristics used in gamified cyber security skills training for employees. As a result, we reviewed the literature on cyber- attackers by searching for some familiar words such as: "cyber criminals", "crackers", and "hackers". We looked at differences in terminology definitions and tried to give “unified” definitions. Various literature describes the following eight classifications of cyber-attackers:

1. Script kiddies/rogue hackers: attackers who do not necessarily have the skills to carry out special attacks without tools provided for them from Internet and friends. Since this group does not necessarily understand how the attacks are carried out, they most likely have no clue about the extent of the damage they can cause [32], [34,35].
2. Cyber-punks (including virus writers): a more experienced group than Script kiddies, these attackers produce and use programs to cause mayhem and to gain recognition. Since this group often looks for fame, it has little to no respect for ethical norms [36].
3. Insiders: invaders who are surrounded by the organization,whoseattacks cause deliberate or unintended harm because of their lawful access [36]. Because access is not a challenge they face, most insider attackers have minimal technical skills [37], which make them relaxed goals for offenders who influence them to perform an action that exposes the organization infrastructure [38].
4. Petty thieves: attackers who seek financial gain from these illegal activities such as stealing person’s identity for profit, and taking over an organizational system for ransoms [33, 36] qualify this group’s activities as not sophisticated, and this group is not entirely dependent on the return of their hacking activities. Most of the individuals of this group are attracted to the finance sector, credit card and banks [33].
5. Grey hats: attackers who are a mix-breed between ethical hackers–white hats and dirty pirates–black hats. Their motive at times is to expose vulnerabilities in an organization’s system and to let them know about it; other times once they found a weakness they can exploit it to gain something [36]. Often highly skilled, they write scripts that cyberpunks and script kiddies typically employ [33].
6. Professional criminals: attackers who are hired to penetrate an organization system. They are also known as cyber-mercenaries [36]. Sometimes the cyber-attackers are employed to attack business competitors for financial gain. Group of attackers operates in the secretive world so that they wouldn’t be identified [33].
7. Hacktivists: invaders who are encouraged by philosophy and belief. This type can consist of terrorist groups [36].
8. The nation’s state attackers who are assumed to be working for a governmental body. Most of their activities are targeted towards the disruption of the enemy’s systems or the protection of the nation state's systems. This assembly comprises guerrilla administrations and freedom fighters, and their goals are not dissimilar to those of recognized governments [33, 36].

Regardless of the type of the attacker, there is a great deal of knowledge sharing in the hacking community. [39] states that sharing information helps build stronger bonds within the group while encouraging and challenging others to learn and engage more.

## 5. Threat from insiders

Insider threats represent an especially deceptive threat to organizations. As committed employees, they are given access to information that could compromise the organization if it falls into the wrong hands [40].

Despite much research into psychology and motivation of insiders, the fact remains that it is tough to predict insider attacks [41]. The following Table summarizes the nine events published on the Chinese website [42] and reveals the importance of protection of the data/business secrets from insiders.

*Table 1.Important Cyber-attacks from insides in China occurring during 2010 to 2013.*

| Company | Internal tech info and commercial secrets | Person's Rank to leak info | Damage | Causes | When |
|---|---|---|---|---|---|
| Foxconn | 3D data diagram for iPad 2 | Technical staff | Massive losses | Unknown | 2010 |
| Samsung Electronics | technology development strategy and other | Technician | Unknown | Not renew contract | 2011 |
| Taiwan's AU Optronics | AU AMOLED and other sensitive technology | Former executives | Unknown | promised annual million dollar salary by the competitors | 2011 |
| Neusoft Group | Trade secrets and core technical information | manager, the deputy in R & D CT machine, and procurement executive | $2.5 million and research development delay | Huge money return | 2012 |
| China National Bureau of Statistics (NBS) | National unpublished economical data | Deputy Director of the Office of the Secretary Office of the Secretary NBS | Reputation and trust from people | Huge benefits | 2012 |
| China Merchants Bank | Customers' personal information | Internal staff | Damage reputation/lost customers | Huge benefits | 2012 |
| Industrial and Commercial Bank of China | Customers' personal information | Internal staff | Damage reputation/lost customers | Huge benefits | 2012 |
| Store No 1 | 900,000 customer's info | Internal staff, previous staff, and external | Lost customer and profit | Huge benefits | 2012 |
| High Tech Computer Corporation(HTC) | New HTC cabinet design and UI for HTC SENSE 6.0 | Vice president and chief archtect, Director, and an | Unknown | The stolen UI for their own company after leaving HTC. | 2013 |

## 6. Past data breaches reported from january 2005 to august 2015.

Figure 1. extrapolates the security threats to the year 2016. We have used multivariate forecasting methods based on modern statistical models, and we will attempt to generalize the extrapolation methods to the multivariate case, using time-series models or techniques that are structural or theory-based. There have been 858,403,517 records breached, and 4,599 data breaches made public since 2005 [43].

The Figure 2. lists top Attack-Methods used between years 1999-2011. Of the total attacks, 40.8% were Denial of Service, 11.3% were Cross Site Scripting (XSS), and 7.5% were Brute Force.

Table 2. illustrates the worst corporate hack from Jun 6, 2005, to March 18, 2015. The majority of organizations compromised were engaged in financial service, government agencies, and retailer. Court Venture (Experian) compromised its 200,000,000 records. Adobe software has compromised 152,000,000 records, and Nationwide has compromised 1 million records.

Figure 3. The hack attack trend from 2005-2015 for Group A consists of Non-Governmental Organization (NGO), Business Services (BSO) and Education (EDU). Educational institutes got maximum hit and NGOs are least affected by the hack.

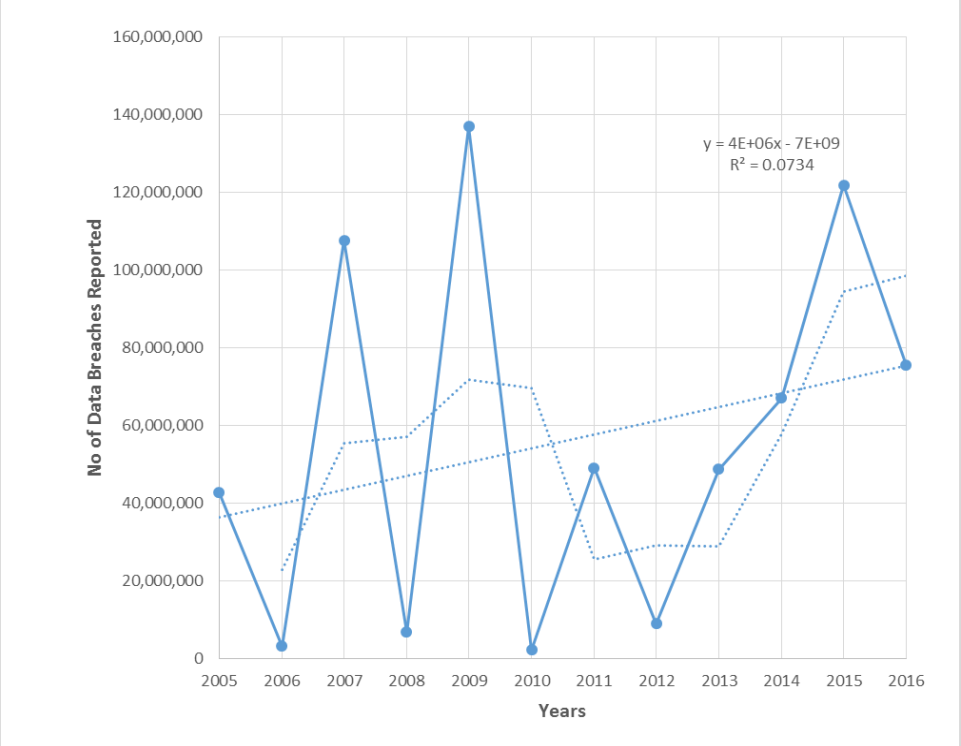


*Figure 1. No. of Data Breaches Reported From January 2005 to August 2015.*

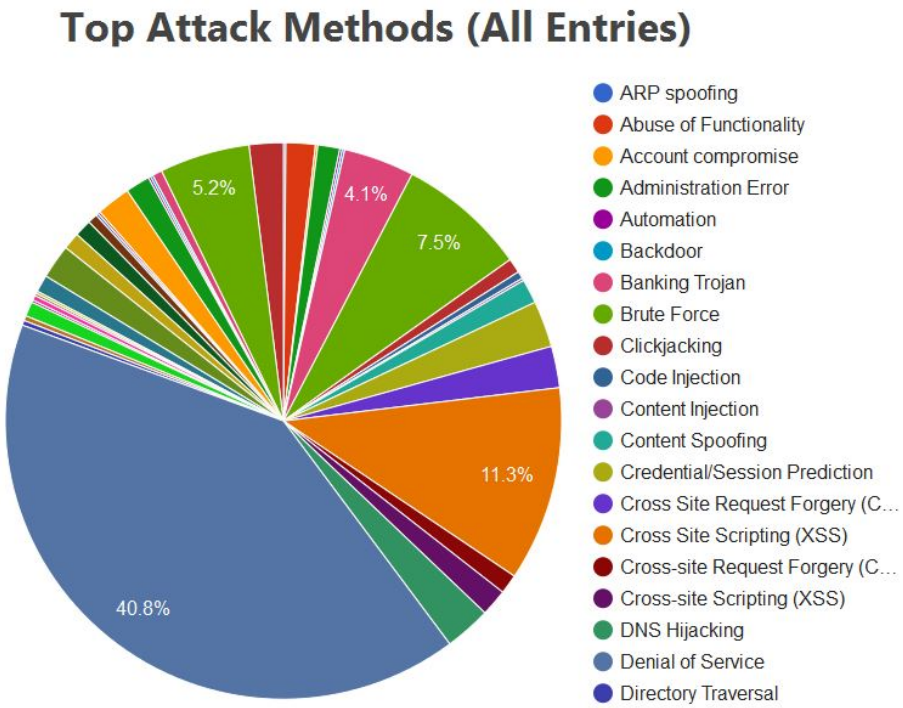


*Figure 2. Top Attack Methods in the years 1999-2011[56]*

*Table 2. A quick guide to the worst corporate hack attacks [43]*

| DATE | COMPANY | TYPE | RECORDS |
|---|---|---|---|
| 3/18/2015 | Premiere Blue Cross | MED | 11,000,000 |
| 2/5/2015 | Anthem | Health Insurance | 80,000,000 |
| 9/2/2014 | Home Depot | Retail | 109,000,000 |
| 8/27/2014 | JPMorgan | Finance | 83,000,000 |
| 5/21/2014 | Ebay | Travel Agency | 145,000,000 |
| 12/13/2013 | Target | Retail | 110,000,000 |
| 10/18/2013 | Adobe | IT (Software) | 152,000,000 |
| 9/14/2013 | Court ventures (Experian) | Collecting court records | 200,000,000 |
| 10/26/2012 | South Carolina Department of REV | Gov | 6,400,000 |
| 11/16/2012 | Nationwide Mutual Insurance | Finance | 1,000,000 |
| 11/10/2011 | Steam (Value Corporation) | game | 35,000,000 |
| 10/30/2011 | Tri Care Management Activity | Insurance (Med) | 5,117,799 |
| 3/26/2010 | Education | education credit | 3,300,000 |
| 2/6/2010 | AvMed Health Plan | Insurance (Med) | 1,220,000 |
| 8/2/2010 | US Military Veteran | Gov | 76,000,000 |
| 1/20/2009 | Hartland Payment System | Insurance (Med) | 130,000,000 |
| 3/23/2009 | Oklahoma Department of Human Service | Gov | 1,000,000 |
| 8/2/2008 | Countrywide Financial Service | Finance | 17,000,000 |
| 7/9/2008 | Division of Motor Vehicle Colorado | Gov | 3,400,000 |
| 9/14/2007 | T.D. Ameritrade | Finance | 6,300,000 |
| 7/3/2007 | Fidelity national Information Service | Finance | 8,500,000 |
| 1/17/2007 | TJX Retail Stores | Retail | 100,000,000 |
| 9/7/2006 | Circuit City and Chase Card System | Retail | 2,600,000 |
| 5/22/2006 | US Department of Veteran service | Gov | 26,500,000 |
| 6/6/2005 | Card System | Finance | 40,000,000 |
| 6/16/2005 | City group ,UPS | Finance | 3,900,000 |

In the year 2006,maximum number of attacks was placed on the educational institutions and there was a gradual reduction in the consecutive years. The lowest hit was in the year 2015. The attack on BSO was the peak in the year 2013. NGO got the highest setback in the year 2013 though, NGO was the lowest hit segment compared to other industries.

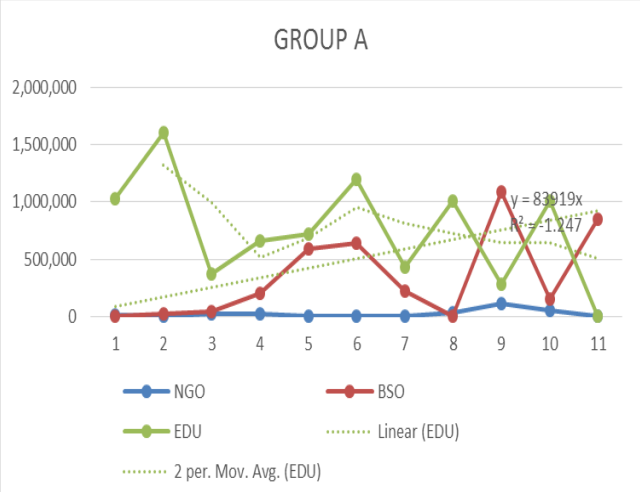


*Figure 3. The hack attacks trend from 2005-2015 for Group A consists of NGO, BSO and EDU agencies*

Figure 4., the hack attack trend from 2005-2015 for group B consists of FIS, R/M, GOV, and MED agencies. The financial institutions were affected badly in the year 2010, the trend showed reduction until 2014 and started increasing in the year 2015. In the group B, the second highest attack impact was on retail merchant agencies, they got the maximum hit in the year 2007, and there is an adverse trend in the year 2015.

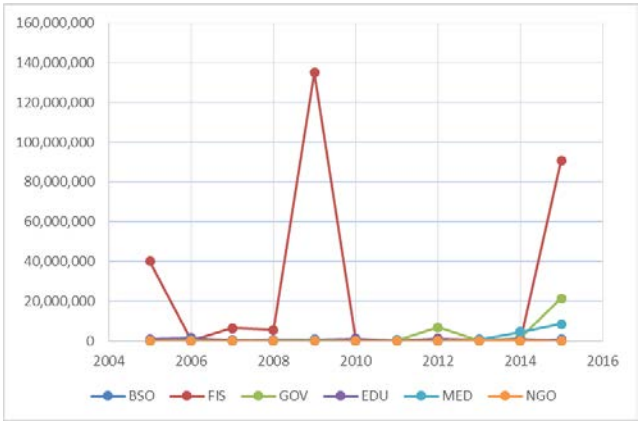


*Figure 4. The Hack Attacks trend from 2005-2015 for Group B consists of FIS, R/M, GOV, and MED agencies*

## 7. Study of recent attacks

From Figure 5. the analysis of cyber-attack from Sept 24, 2015 to Oct 24, 2015, the educational institutions are the highest target of attacks and contribute 28% to the total cyber-attacks. Figure 6. illustrates the overall rate of cyber-attacks in a typicalhour [1]. One can see that every three seconds one cyber-attack is taking place in our digital world.

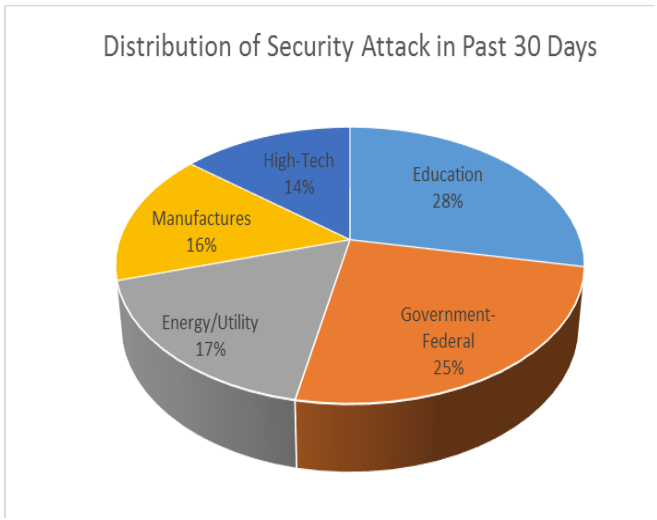


Figure 5. *The percentages of cyber-attacks reported in different industries from September 24 2015 to October 24, 2015.*

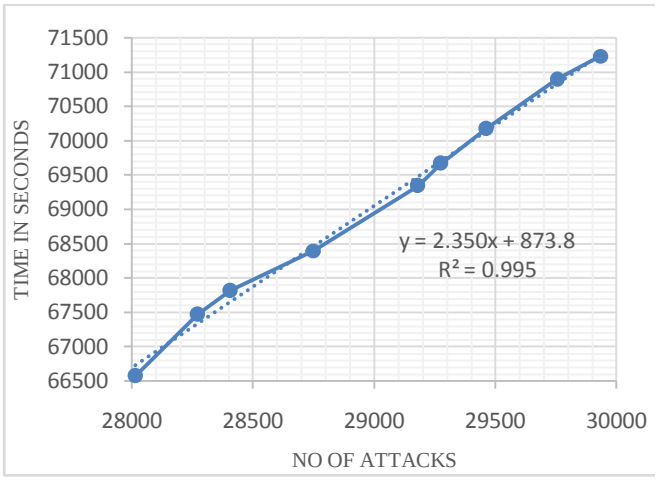


*Figure 6. The typical rate of cyber-attacks reported all over the world from 18:29:34 -19:41:33 Eastern Time on October 24, 2015.*

Figure 7. describes the historical number of data breaches reported in different industries from January 2005 to August 2015[43]. Table II gives a quick guide to the worst corporate hack attacks from 6/16/2005 to 3/18/2015[44].

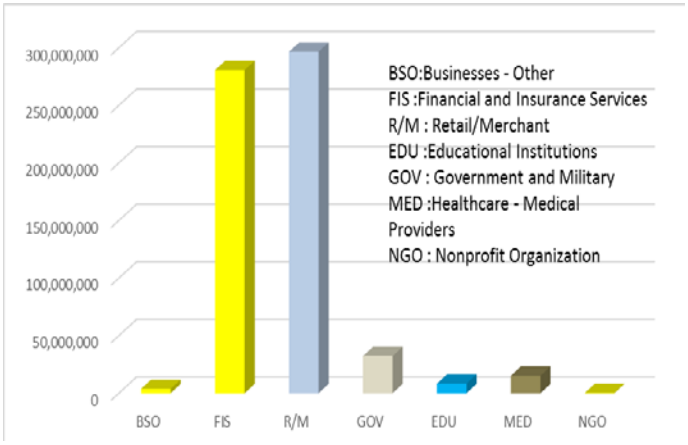


*Figure 7. No. of data breaches reported in different industries from January 2005 to August 2015.*

Table 3. describes a snapshot of approximately one minute on 12/08/2015. During this interval, maximum attacks were originated from the United State, and at the same time, maximum attacks were targeted from other countries like Saudi Arabia, China, Singapore, Germany, United Kingdom, Australia and Russia. The rank of country and number varies according to a time of the typical day.

*Table 3. A quick study of corporate attacks on 12/08/2015 10.14.49.001PM-10.15.49.779 PM [45]*

| Attack Origins Country | | Attack Target | | Attack Types | | |
|---|---|---|---|---|---|---|
| Number | Country | Number | Country | Number | Port | Service Type |
| 944 | United States | 1256 | United States | 485 | 25 | smtp |
| 308 | Saudi Arabia | 480 | France | 469 | 2222 | EthernetIP-1 |
| 278 | China | 315 | Saudi Arabia | 250 | 137 | unknown |
| 170 | Singapore | 204 | United Arb Emirats | 214 | 23 | telnet |
| 128 | Germany | 151 | Splain | 159 | 5060 | sip |
| 127 | United Kingdome | 56 | Russia | 112 | 50864 | unknown |
| 99 | Australia | 25 | Liechtenstein | 88 | 445 | microsoft-ds |
| 74 | Russia | 19 | Cyprus | 71 | 50856 | unknown |
| 58 | Netherlands | 9 | Maxico | 59 | 1500 | vlsi-lm |
| 36 | France | 8 | Taiwan | 31 | 16286 | unknown |

Table 4. describes the snapshot at 10.14.49.779 PM on 12/08/2015. The table gives description of the attacker and its geographical location, IP information, the attack type and port number information. During this interval, typical tools were SMTP, telnet, squid http, black ice-icecap, and unknown tools

*Table 4. A quick guide tocorporate attacks on 11/24/2015 10.14.49.001PM to 10.14.49.779 PM[45].*

| Time Stamp | Attacker | Attacker IP | Attacker GEO | Target GEO | Attack Type | Port No |
|---|---|---|---|---|---|---|
| 10.14.48.896 | Macro Zebut | 194.187.140.61 | Meschede,DE | Aix-En-Prove | smtp | 25 |
| 10.14.49.056 | Unit-ls Ltd. | 195.211.154… | Odessa,USA | De Kabl Juncti | unknown | 21320 |
| 10.14.49.338 | Ovh Sas | 37.187.106.163 | Roubaix,FR | Aix-En-Prove | smtp | 25 |
| 10.14.49.474 | Chinet Shanghai Broadband Adsl 168 | 58.33.40.40. | Shanghai,CN | Lynnwood,US | unknown | 50864 |
| 10.14.49.697 | Taipei Taiwan | 61.231.5.197 | Taipei, TW | Marseille,FR | unknown | 8118 |
| 10.14.49.779 | Taipei Taiwan | 61.231.5.197 | Taipei, TW | Marseille,FR | squid-http | 3128 |
| 10.14.50.037 | Websense Hosted Security | 85.115.54.190 | Southall,GB | Aix-En-Prove | smtp | 25 |
| 10.14.50.156 | Microsoft Corporation | 157.56.112.254 | Amsterdam,NL | Aix-En-Prove | smtp | 25 |
| 10.14.50.440 | M1 Ltd | 175.156.153.55 | Singapore,SG | Dubai,AE | telnet | 23 |
| 10.14.50.607 | Singlehop Inc. | 198.20.70.114 | Chicago,US | Lynnwood,US | blackice-icecap | 8081 |

## 8. Twitter

Twitter is an online social networking website and microblogging service that allows users to post and read text-based messages of up to 140 characters, known as "tweets." It was launched by Jack Dorsey in July of 2006. Twitter is now in the top 10 most visited internet sites [34]. According to Statistic Brain as of dated September 25th, 2015 there are 58 million tweets per day (672/sec) [34]. Also, we have presented mining from recent tweets posted on Twitter's database and compared the data sets collected from the past fifteen years.

During one hour period on 11/24/2015 at 8.00PM-9.00PM,we have collected 36724 KB Twitter streaming data from the Potomac Maryland location and typically we found those were 16 incident of Hacker, five security, four Threat, nine Breach and one Attack. We have developed a small Perl-batch script for mining the Twitter's streamed data, which is much faster than standard Tweeter's API. We have observed that the rate of Tweets depends on the time and a typical day of popular cyber-attacks. Also, users may use their own language for discussion of cyber-attacks.

As shown in Table 5.,samples data collected from Twitter in Potomac, MD for 8.00 PM- 9.00 PM from Nov 18 to Nov 24, 2015. If we consider the uninterrupted cases only, we found the average size of raw data was 1331.8 MB, data with general messages only 40393.8KB and, the data with security keyword messages are 36KB.

*Table 5.Sample of Twitters data collected in Potomac, MD for 8.00 PM- 9.00 PM from Nov18 to Nov24, 2015.*

| Twitter Streamed Data | | | |
|---|---|---|---|
| Date | Raw Data (MB) | Data with only twitter messages (KB) | Messages with only key words (KB) |
| 11/18/2015 | 596* | 16890 | 21 |
| 11/19/2015 | 1381 | 40074 | 46 |
| 11/20/2015 | 1403 | 44997 | 40 |
| 11/21/2015 | 1380 | 43722 | 30 |
| 11/22/2015 | 1349 | 36452 | 28 |
| 11/23/2015 | 770* | 22265 | 30 |
| 11/24/2015 | 1146 | 36724 | 36 |
| Note: | | | |
| 1) *During sampling hour, the program sometimes stopped | | | |
| while collecting the data due to a network issue. | | | |
| 2) Location for data collection: Potomac, MD | | | |
| 3) Time: from 8 pm to 9 pm every day during Nov. 18 to Nov. 24, 2015 | | | |
| 4) Desktop: Optiplex960 with Intel Core Duo CPU 3.00GHz and | | | |
| RAM 8.00GB as well as Fios network connection | | | |

## 9. Samples of typical tweets on the twitter

Following are some samples of messages tweets with key security words during11/19-23/2015.

"Surprise I hacked Yik Yak"
"Successfully Hacked Thousands of ISIS Social Media Accounts"
"my account was hacked, but getting it squared away"
"Whoever hacked my phone so John cena would appear every time I hit the home button, fu… you."
"Someone be hacked Nub Wub on xbox and give me the gamertag"
'Wow, Georgia. That's not so much a data breach as a data ticker tape parade, raining down everywhere."
"His Own Father HACKED His Hand Off, The Reason Why Will Shock You"
"one of our Dropbox accounts must've been hacked"
"i thought u hated her u must be hacked"
"Hacking Secret Ciphers with Python Hacking Secret Ciphers wi"
"Congrats to HeartbeatTVProstaffer Jake Hacker on his Clermont county OH stud. Well done, Brother!"

## 10. Anticipation for cyber threats

As shown in Figure 1., according to moving average model for overall industry there will be 38% reduction in data breaches in the year 2016. Table VI describes the prediction (based on Monte Carlo model) for the year 2016 for the individual industry.

Table 6. illustrates the forecast for 2016 for various industry segments. Monte Carlo method is used for 2016 prediction; we can notice that all individual industry is showing a drastic decrease in cyber threat compared to the year 2015. The biggest (60%) percentage drop is in MED, and lowest rate (33%) drop is in the GOV segment. The NGO and EDU's data were not available for the year 2015. We believe that the average of Monte Carlo and Extrapolated value will produce the most accurate prediction.

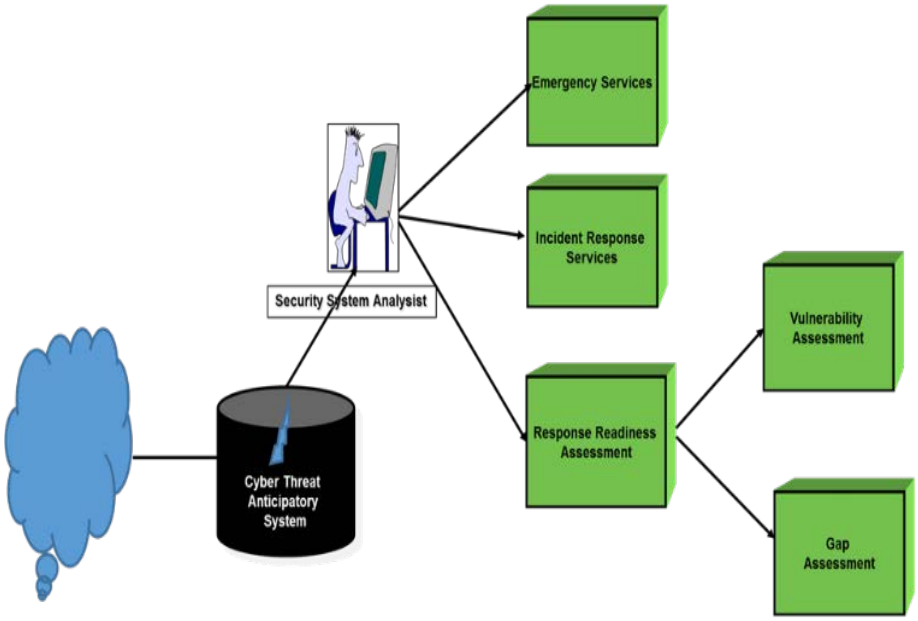


*Figure 8. The cyber threat anticipatory system.*

*Table 6.Cyber threat prediction for the year 2016*

| | Nominal | Min | Max | Number of Cyber Attack in 2015 | Stochastic( Monte Carlo) Prediction for 2016 | Extrapolation for 2016 | Mean of Monte Carlo and Extrapola |
|---|---|---|---|---|---|---|---|
| BSO | 347946 | 2274 | 1086854 | 847,200 | 458375 | 733727 | 596051 |
| FIS | 25518610 | 51236 | 135000550 | 91,000,000 | 83011517 | 10657338 | 93668855 |
| GOV | 21500000 | 48045 | 21500000 | 21,500,000 | 12225994 | 1758873 | 13984867 |
| NGO | 36243 | 1272 | 110000 | N/A | 52824 | 36587 | 44705 |
| MED | 1687119 | 15967 | 8551626 | 8,551,626 | 3627097 | 3910632 | 3768865 |
| EDU | 755848 | 915 | 1609402 | 611,130 | 611130 | 875761 | 743446 |

## 11. The cyber threat anticipatory system

As shown in Figure 8., whenever a cyber-threat anticipatory system encounters the sign of a possible threat in the streaming data from various online sources, it will send an alert to the security analyst. An alarm can be triggered by system vulnerability and network traffic. Based on the typical threat level, the security analyst can investigate the effect and take required protective actions.

After receiving a security alert, the system analyst will activate emergency services, incident response services and response readiness assessment system. The response readiness evaluation system will enable vulnerability and gap assessment systems. The anticipatory threat system is thus generalized and can be tailored to the requirements of the individual organization.

The cyber threat anticipatory system identifies the high-threat events that have the potential to interrupt operations seriously. To counteract or mitigate cyber threat we have developed Split-protocol based Attack Counter Tree and various protection mechanisms. We have demonstrated the variety of security applications based on the split protocol concept. The useful design features of the Split-server concept offers development of unique firewall architecture.

Split-servers' ability to work in a many-to-many configuration with role changeover and migration capability. In Split-server, the CS and DS units are identical in design, and one can be substituted to take on the role of the other [46-54]. Also, one CS can interact with many DS devices; and, one DS can work on behalf of many CS devices. And our power series based encoding technique allows compaction and anonymity of information [50, 55]. Also, from Figure 1., the extrapolation for the year 2016, there will be 38% reduction in data breaches. And most vulnerability to data breach will persist in the business of Financial, Insurance Services, and Retail/Merchant. NGO, and Other Business and Educational Industry will have the least vulnerability for the data breaches.

## 12. Discussion

We have used multivariate forecasting methods based on modern statistical models and made some attempts at generalizing the extrapolation methods to the multivariate case, such as time-series models. There were 858,403,517 records breached, and 4,599 data breaches made public since 2005. Also, we have presented mining from recent tweets posted on Tweeter's database and compared the data sets collected from the past fifteen years. If we consider the uninterrupted cases, we found the average size of raw data was 1331.8 MB, data with general messages only 40393.8KB and, the data with security keyword messages are 36KB.

As per our forecast for the year 2016, there will be 38% reduction in data breaches. And most vulnerability to data breach will persist in the business of Financial, Insurance Services, and Retail/Merchant. NGO, and Other Business and Educational Industry will be least vulnerable to data breach.

## 13. Conclusion

This paper discussed the problem of predicting cyber-attacks and where they will occur in different segments of business or government. We proposed that data-driven models can be adapted from economics research to predict where future cyber-attacks will occur. We believe that if this approach has promised in identifying patterns in cyber threats to make predictions, and that this information can provide a significant benefit. We presented the analysis of 595,529,851 past events of cyber-attacks (from 2000 to 2015) collected from various databases, research publications and information posted on the web. Also, we have presented mining from recent tweets posted on Tweeter's database and compared the data sets with the last fifteen years collection of data. In atypical example, we have collected 36724 KB Tweeter streaming data from the Potomac Maryland location and we found there was 16 incident of Hacker, five security, 4 Threat, 9 Breach and one Attack. We have developed a small Perl-batch script for mining the Tweeter's streamed data. Perl-batch script is much faster than Tweeter API. During our study over one year period, we have observed that the rate of tweets depended on the time and a typical day of popular cyber-attacks. Also, users may use their language for discussion of cyber-attacks.

## References


[1] FireEye Cyber Threat Map, https://www.fireeye.com/cyber-map/threat-map.html (accessed on 28 Nov 2015)

[2] Taylor, Robert W., Eric J. Fritsch, and John Liederbach. Digital crime and digital terrorism. Prentice Hall Press, 2014.

[3] http://www.lyncmigration.com/news/215/10/28/8268137.html accessed on 28 Nov 2015

[4] Romanosky, Sasha, David Hoffman, and Alessandro Acquisti. "Empirical analysis of data breach litigation." Journal of Empirical Legal Studies 11.1 (2014): 74-104.

[5] http://www.verizonenterprise.com/DBIR/2015 accessed on 28 Nov 2015

[6] http://www.experian.com/assets/data-breach/brochures/2014-ponemon-2nd-annual-preparedness.pdf

[7] https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf)

[8] http://www.symantec.com/security_response/publications/threatreport.jsp

[9] http://www.statisticbrain.com/Twitter-statistics/ (accessed January 02, 2016).

[10] G. W. Bush. National strategy to secure cyberspace, office of the president. 2003.

[11] President's Information Technology Advisory Committee, Cyber Security: A crisis of prioritization, 2005.

[12] The National Strategy for Homeland Security, http://www.dhs.gov/interweb/assetlibrary/nat-strat-hls.pdf, 2002.

[13] Roy, Sankardas, Charles Ellis, Sajjan Shiva, DipankarDasgupta, VivekShandilya, and Qishi Wu. "A survey of game theory as applied to network security." In System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1-10. IEEE, 2010.

[14] Jaganathan, Venkatesh, PriyeshCherurveettil, and PremapriyaMuthuSivashanmugam. "Using a Prediction Model to Manage Cyber Security Threats." The Scientific World Journal (2015).

[15] E. Tittle, Preventing and Avoiding Network Security Threats and Vulnerabilities, 2013 http://www.tomsitpro.com/articles/threat management-it security-firewall-it certification-network security,2-477.html.

[16] J. Wu, L. Yin, and Y. Guo, "Cyber attacks prediction model based on Bayesian network," in Proceedings of the 18th IEEE International Conference on Parallel and Distributed Systems (ICPADS '12), pp. 730–731, Singapore, December 2012.

[17] N. S. Anusha, T. S. Soujanya, and D. S. Vasavi, "Study on techniques for providing enhanced security during online exams," International Journal of Engineering Inventions, vol. 1, no. 1, pp. 32–37, 2012.

[18] E. T. Axelrad, P. J. Sticha, O. Brdiczka, and J. Shen, "A Bayesian network model for predicting insider threats," in Proceedings of the 2nd IEEE Security and Privacy Workshops (SPW '13), pp. 82– 89, May 2013.

[19] M. A. Khan and M. Hussain, “Cyber security quantification model,” Bahria University Journal of Information and Communication Technology, vol. 3, no. 1, pp. 23–27, 2010.
[20] A. P. Moore, D. A. Mundie, and M. L. Collins, “A system dynamics model for investigating early detection of insider threat risk,” Tech. Rep. DM-0000143, Program Software Engineering Institute, Carnegie Mellon University, 2013.
[21] Bharat S. Rawal and Anthony Tsetse, “Analysis of Bugs in Google Security Research Project Database.” IEEE RAICS-2015 is the Third edition of the International Conference on Recent Advances in Computational Systems, December 10-12, 2015.
[22] Ozment, Andy. "Improving vulnerability discovery models." Proceedings of the 2007 ACM workshop on Quality of protection. ACM, 2007.
[23] Weber S, Karger PA and Partaker A (2005) A software flaw taxonomy: Aiming tools at security. Software Engineering for Secure Systems (SESS’05).
[24] Prakash, Atul, and RudrapatnaShyamasundar, eds. Information Systems Security: 10th International Conference, ICISS 2014, Hyderabad, India, December 16-20, 2014. Proceedings. Vol. 8880. Springer, 2014.
[25] Howard, M, LeBlanc, D and Viega, J (2005) 19 Deadly Sins of Software Security.Emeryville, C A: McGraw-Hill/Osborne
[26] Dowd, M, McDonald, J and Schuh, J (2006) the Art of Software Security Assessment: Identifying and Preventing Software Vulnerabilities. Boston,MA:Addison Wesley Professional
[27] Weber S, Karger PA and Partaker A (2005) software flaw taxonomy: Aiming tools at security. Software Engineering for Secure Systems (SESS’05).
[28] Landwehr CE, Bull AR, McDermott JP and Choi WS (1994) A taxonomy of computer program security flaws. ACM Computer Surveys 26(3):211–254.
[29] ISO 7498:1984 Open Systems Interconnection - Basic Reference Model.
[30] Du, W, Mathur, AP (1998) Categorization of Software Errors that led to Security Breaches, In Proceeding of the 21st National Information Systems Security Conference (NISSC'98), Crystal City, VA.
[31] Meunier, Pascal. "Classes of vulnerabilities and attacks." Wiley Handbook of Science and Technology for Homeland Security (2008)
[32] Adams, Mackenzie, and MagedMakramalla. "Cybersecurity Skills Training: An Attacker-Centric Gamified Approach." Technology Innovation Management Review 5.1 (2015).
[33] Rogers, M. K. 2011. The Psyche of Cybercriminals: A Psycho-Social Perspective. In S. Ghosh& E. Turrini (Eds.), Cybercrimes: A Multidisciplinary Analysis: 217–235. New York, NY: Springer.
[34] Aggarwal, P., Arora, P., &Ghai, R. 2014. Review on Cyber Crime and Security. International Journal of Research in Engineering and Applied Sciences, 2(1): 48–51.
[35] Roy, Sankardas, Charles Ellis, Sajjan Shiva, DipankarDasgupta, VivekShandilya, and Qishi Wu. "A survey of game theory as applied to network security." In System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1-10. IEEE, 2010.
[36] Hald, S.L.N.; Pedersen, J.M., "An updated taxonomy for characterizing hackers according to their threat properties," in Advanced Communication Technology (ICACT), 2012 14th International Conference on , vol., no., pp.81-86, 19-22 Feb. 2012
[37] Williams, P. A. H. "In a ‘trusting’ environment, everyone is responsible for information security" Information Security Technical Report, Volume 13, Issue 4, November 2008, Pages 207-215
[38] Parmar, B. "Employee Negligence: The Most Overlooked Vulnerability." Computer Fraud & Security, March 2013, Pages 18–20.
[39] Arief, B., &Besnard, D. "Technical and Human Issues in Computer-Based Systems Security". Technical Report Series: University of Newcastle upon Tyne Computing Science. Newcastle, UK: Newcastle University. 2003
[40] Greitzer, Frank L.,and Deborah A. Frincke. "Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation." In Insider Threats in Cyber Security, pp. 85-113. Springer US, 2010.
[41] Kramer LA, RJ Heuer, Jr, and KS Crawford. 2005. Technological, Social, and Economic Trends that are Increasing U.S. Vulnerability to Insider Espionage. Technical Report 05- 10, Personnel Security Research Center (PERSEREC), Monterey, California.)
[42] http://ja.house.sina.com.cn/news/2013-09-11/09163559541_2.shtml accessed on 01/01/2016
[43] http://www.privacyrights.org/data-breach accessed on 01/01/2016
[44] http://list25.com/25-biggest-cyber attacks-in-history/ accessed on 28 Nov 2015
[45] http://map.norsecorp.com accessed on Jan 05, 2015.
[46] B. Rawal, R. Karne, and A. L. Wijesinha. “Mini Web Server Clusters based on HTTP Request Splitting” HPCC 2011: The 13th IEEE International Conference on High Performance Computing and Communications, Sep 2-4, 2011, Banff, Canada.
[47] B. Rawal, R. Karne, and A. L. Wijesinha, “Splitting HTTP Requests on Two Servers,” 3rd International Conference on Communication Systems and Networks (COMSNETS), 2011.
[48] B. Rawal, R. Karne, and A. L.Wijesinha, H.Ramcharan and Songjie Liang. “A Split Protocol Technique for Web Server Migration,” The 2012 International workshop on Core Network Architecture and protocols for Internet (ICNA-2012) October 8-11, 2012, Las Vegas, Nevada, USA.
[49] Bharat Rawal, Harold Ramcharan and Anthony Tsetse, “Paper 68 Emergent of DDoS Resistant Augmented Split Architecture,” IEEE 10th HONET-CNS, EMU, Famagusta, Cyprus 2013.
[50] Bharat S.Rawal, Sonjie Liang, Anthony Tsetse and Harold Ramcharan, “Split-encoding: The next frontier tool for Big Data,” 2nd International Conference on Advanced Computing, Networking, and Informatics [ ICACNI-2014 ] 24 - 26 June 2014 Kolkata, India.
[51] Bharat S. Rawal, Lewis I. Berman and H.Ramcharan, “Multi-Client/Multi-Server Split Architecture,” The International Conference on Information Networking (ICOIN 2013), Jan 28-30, 2013 Bangkok, Thailand.

[52] B. Rawal, R. Karne, and A. L. Wijesinha. Sonjie Liang. “Applications of Split Protocol Paradigm” International Journal IJCA.
[53] Bharat Rawal, “ACT Meets with Split-protocol” International Journal on Computer Networks & Communications (IJCNC) 2015.
[54] Bharat Rawal, Ramesh Karne and QiangDuan "Split-system: The New Frontier of Cloud Computing" accepted by IEEE CSCloud Nov. 3-5, 2015, New York, USA.
[55] Bharat S Rawal and Songjie Liang. “Nth Order Binary Encoding with Split-protocol.” (Under review)
[56] Web-Hacking-Incident-Database, http://projects.webappsec.org/w/page/13246995/ (accessed on 28 Nov 2015).