Abstract

Purpose; Internationally increasing terror risks is no longer an issue of particular states, but of all, unsafe from terrorism at any time. At a point which effective countermeasures against incrementing terrors and cyber terrors, the thesis proposes a Big-Data based Anti-terrorism system. The suggested system captures terror cells via surveillance systems, collects relevant data to the cell, terrorists and their supporters online and offline and analyzes the tremendous amount of data via Big Data, thus maintains high level of monitoring on captured terror cells in accordance with the analysis results or blocks pertaining terror networks transfer all data to the terror response center for elimination of the cells. To effective utilize the anti-terror system proposed in the thesis, cooperation for general citizens and legal supports by the government are necessary for reading and collecting private information. Specialists throughout the world have been asserting, civil and institutional cooperations are essentially required to cope with international crimes and terrors. Measures to keep balance between international security and personal information protection are needed.

[Keywords] Crime Prevention, Terror, Big Data, Data Mining, Network

1. Introduction

Use of the Internet is broadly and secretly managed by terror cells and terrorists established from various political, economic, social, historical and religious issues. Typically, the purpose of the Internet use by terrorists are fundraising for cell operations, information delivery, secured communication, political behavior, collection and spread of cyber panic, and recruitment and motivation of terror cell members[1].

Terror groups makes tracking and control by investigative and intelligence agencies through transferring messages via spam, bulk and junk mails in a form of encrypted Steganography. In fact, Al-Qaeda played a hub for 911 terror, and ordered subordinates and affiliated terrorists through unstandardized data files such as particular image messages via emails and Usenet. Terror cells have been not only ordering terror commitments via the Internet, but also conducting recruitments to cultivate potential terrorists[1][2]. For example, in Korea, once a 18 years old high school student, Kim, has been seduced by IS terrorist recruitment via SNS and his departure to Turkey, with his disappearance thereafter, he has been missing until now. Likewise, the characteristic of modern terrorism incorporates the Internet, thus, is a virtual, but organic network combination. To prevent and cope with terrors, that are combined into organic networks, first, as much data as possible relevant to offline terrors must be gathered, analyzed and processed. However, the amount of data being collected in relation to terrors is overwhelming till it cannot be managed with existing data processing method, thus new concepts for data processing and storage have become required. Big Data not only stores, but also process stored data in real time, thus is a notion of future prediction based on processed data, or refers to huge amount of data and its data processing.
The number of days for suspect arrest of Boston terror in the United States in April 2013 were only four days. It was possible owing to thorough investigation on 10TB-scale data, consisting of messages, collected scene screens via SNS, both inside and outside CCTV recordings, and all other materials that could have been useful for the case resolution via Big Data by investigative agencies, immediately after the terror had occurred[3].

The thesis proposes an Anti-terror system based on Big Data, the most effective method for data analysis and on data mining, a type of the Big Data analysis techniques to analyze data being accumulatively collected both online and offline.

2. Related Research

2.1. Definition and types of terrors

Terror and Terrorism are commonly used as synonyms, however have slight difference in their meanings. Terror refers to “behaviors to make an enemies or opposing parties threatened or terrified through violence[4]”, and Terrorism refers to “organized and collective violence commitment, ideology or belief for political purposes”[5]. United Nation Security Council Resolution 1373 defines Terrorism as “criminal behaviors to force specific person, the public, the government and the international organizations to act or to stop them from a particular behavior by involving deaths, serious injuries and harms, including taking hostages, against civilians, causing fear to a specific group or person and the public[6].

Categories of terrorism based on a few standards as followings. First, depending on committing agents: terrorism from the top or the bottom, second whether a certain state is involved, civilians or territories from more than one nation are associated: domestic terrorism, state terrorism, interstate terrorism and transnational terrorism, et al. Lastly, depending on methods and tools of terror, key figure assassination terror, hostage-abduction terror, suicide-bombing terror, aircraft hijacking-bombing terror, ship hijacking-bombing terror, cyber terror and weapons terror of mass destruction are classified[7].

Figure 1. Process of data mining.

Database established from collection of relevant data from various sources including the Internet based on Pinketron Global Intelligence Service Data(PGIS) is referred as Global Terrorism Database(GTD). GTD classification categorizes terror into attack, weapon and target et al. Attack is classified into 9 types – assassination, armed attack, explosive, airborne aircraft hijacking, riot-rebellion, kidnapping-hostage, infrastructure attack, unconventional attack and unidentified. Weapon is categorized into 12 types – biological weapon, chemical weapon, firearms, explosive dynamite, fake weapon, arson-incendiary bomb, scuffle-pandemonium, vehicle, malicious mischief(equipment-device), others and unidentified. Lastly, target is divided into 22 types – business, government, police, military, abortion, air, diplomatic-transnational government, educational institution, water-food-water supply, newspaper-TV-media, non-government, others, civilian-building, religious figure-system, communication, non-militia terror, tourist, transportation, unidentified, water-electricity gas and violence[6].

2.2. Big data and data mining

Global consulting institution, Mckinsey, indicated Big Data as “huge data exceeding the collectable, storage-able, manageable and analyzable capacity with existing data processing technique, hence its definition is subjective and will consistently be changed”. In addition, Big Data defined by IDC as “various data, which next generation technology and architecture to enable extremely high speed collection, discovery and analysis of data via value extraction from data at a low cost”, while Gartner defined as data and phenomenon generating various types of Big Data at a high speed, too fast to deal with[2][8]. Big Data differentiates itself from three aspects
Volume, Velocity and Variety. Volume refers to the physical data amount being simply stored and is the most basic characteristic. Velocity refers to advanced real-time data processing, thus signifies the importance of how fast the visualization process must be executed. Lastly, Variety refers to incorporation of various data sets, thus not only standardized, but also unstandardized data such as pictures, audio, video and social media data, and log files are included[2][8].

Table 1. Data mining phases[2].

<table>
<thead>
<tr>
<th>Phases</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data Selection</td>
<td>Setting a goal for knowledge to discover, create three data set for selection, collection and analysis, then extract and organize variable sets and data sample for specific analytical algorithm to be applied.</td>
</tr>
<tr>
<td>2. Pre processor</td>
<td>Phase to modify data, which revision is required, to discover inherent meaning or pattern and knowledge.</td>
</tr>
<tr>
<td>3. Data Transition</td>
<td>Transition process to increase analysis efficiency and to decrease data complexity, reducing the number of data to be considered, selecting adequate variables based on existing data for more accurate model, then new variables are generated.</td>
</tr>
<tr>
<td>4. Data Mining</td>
<td>If data has no value left via examinations and evaluations, the phase returns to Data Selection again, but if it has, the data is transferred into knowledge.</td>
</tr>
</tbody>
</table>

With 4 values – insight, responsiveness, competitiveness and creativity – and advantages, Big Data has become an appropriate mean to effectively cope with future society, often so-called as uncertainty, risk, smart and convergence[8][9].

Insight is the role of Big Data to prepare for uncertain future – pattern analysis, future outlook and scenario simulation et al. for predictable situations in future societies can be performed on the basis of social phenomena and real world data. Responsiveness plays a role of Big Data against risk of future society to analyze environmental, social and monitoring data and to rapidly capture risk or abnormal signs, thus is a notion emphasizing real-time, in that rapid decision-making and real time responsive supports become available through preliminary recognition and analysis of issues. Competitiveness is the role Big Data to cope with characteristic of future society,

Smart. Big Data performs huge data processing volume – Exabyte – thus, not only enables optimal supports such as context awareness and artificial intelligence services, but also ensures product competitiveness through trade area analysis and flexible and adequate responses to other trend shifts. Lastly, the role of Big Data is Creativity for Convergence. Big Data creates new values through integrations with other areas and utilizes of advanced statistical techniques on convergence area, which has sophisticated causality and correlation, to improve safety and to minimize trial and error[2].

Terminologies for disclosing information from data vary. Knowledge Extraction, Information Harvesting, Information Discovery, Data archaeology and Data Pattern Processing refers to a similar notion[10][11][12]. Data Mining is the most fundamental process to discover knowledge and Knowledge is generated as <Figure 1> illustrates. Data Mining phases as <Table 1> [2].

3. Anti-Terror System in the Proposal

The cyber space has much more advantages to perform various functions and tasks all over the world than closed spaces. Such a systematic disposition if terrorist characteristics are secretly dispersed, thus may develop into Homegrown Terrorism via a number of specific stages. Critical reasons for modern terrorism being active via cyber spaces on the Internet are its anonymity and secrecy. Terror cells in virtual space can access from anywhere, as anyone, with anonymity, and the agents offering terror-related information and educational materials operates through completely irrelevant homepage secretly, thus its function as Cyber Heaven may be sustained. Such stages of terrorism are operated systematically and secretly unlike the reality, and as the stages are not consistent-permanent, accessibility of institutions to control relevant crimes is getting more difficult. Furthermore,
terror cells on cyber spaces make tracking difficult from communication to recruitment ranges. Figure 2 is an phase illustration of terrorism on the Internet[1][2].

Figure 2. Internet terrorism progress phases[1][2].

<table>
<thead>
<tr>
<th>Communication</th>
<th>Internet</th>
<th>Chatting Room</th>
<th>Steganography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Class</td>
<td>Camp</td>
<td></td>
</tr>
<tr>
<td>Planning/Coordination</td>
<td>Haven</td>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>Fund</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruiting</td>
<td>Video</td>
<td>Digital Media</td>
<td></td>
</tr>
</tbody>
</table>

The anti-terror system in the thesis largely consists of 4 phases, and is structured as Figure 3. Having terror signs captured on the Internet networks, via programs, relevant data is collected to grasp which terror cells are associated with. Analyzing the gathered data, the network is consistently monitored or is collapsed by utilizing the analysis result. The phases as follows;

1. Capturing Terror Cells: Using the surveillance program, suspicious groups or suspects are monitored. The program proceeds monitoring on frequently visited websites, blog postings, dialogue forum, search engine queries, SNS posting contents and website histories of the suspicious targets.

2. Collection of Relevant Data: Identifying terror cells, terrorists and their supporters from the analysis result, all relevant relationship and data online and offline are collected. Not only standardized data online, but also non-standardized data – pictures, audio and video, social media data and log files – CCTV recordings, phone-call recordings and even daily material are collected.

3. Big Data Analysis: A. Select data from the data collection for analysis, then extract and organize data sample for specific analytical algorithm to be applied. B. Modify data which are required for revision. C. To increase analysis efficiency and to reduce data complexity, decrease the number of data to be considered, select appropriate variable based on existing data, then generate new variable. D. If illustrations and examinations are not sufficiently significant, the process returns to data selection phase again or move on to the next phase.

4. Surveillance on / Collapse of Terror Cells: Based on the analysis result, the target is constantly monitored, and if there is sufficient probability for actual terror incidence, websites for surveillance are disconnected and the data is transferred to terror response center.

4. Conclusion

Terror risks increasing throughout the world are no longer problems of other countries, but ours insecure from terrors as terrors may occur at anytime. Particularly, in 2014 when Kim collected data about and went to Turkey to join IS Terror groups and became a part of the terror cell, emergence of cyber or international terror establishments has become a problem of all states, not particulars. The thesis has suggested a Big-data based Anti-terror system, at a point, which effective countermeasures against recently increasing terrors and cyber terrors. The proposed system collects relevant data both online and offline to the captured terror cells online, proceeds data-mining to analyze the related data and then, performs consistent monitoring on the cells with terror signs. Moreover,
networks for the high-risk terror cells are entirely blocked, and their associated information are transferred to terror response centers.

Efficacy of social network analysis and Big Data analysis on international crimes and terrorism were acknowledged through practical experiences by foreign institutions. However, to effectively utilize such a system, as reading and collection of personal information are prerequisites, cooperations of general citizens and legal supports by the government are needed. In states, where huge terrors had occurred, infringement of personal freedom and privacy via legislations such as Patriot Act was situationally permitted however, since the disclosure by Snowden, reading and collection of private information have become available only through a warrant. Professionals emphasizes on cooperations among the people and the government to deal with international crimes and terrorism. Measures to balance between national security and privacy protection should be devised.

5. References

5.1. Journal articles


5.2. Additional references


Author
Lee Jae-young/ Semyung University Assistant Professor  
B.A. Semyung University  
M.A. Semyung University  
Ph.D. Chungbuk National University.

Research field

Major career
- 2012~2016.Semyung University, Assistant Professor(Department of Liberal Education)  
- 2016~present. Semyung University, Assistant Professor(School of Information & Communication Systems)