Critical Analysis of the Emerging Dark Web

A dissertation submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) in Software Engineering

Dissertation Project

By Abdulmuttaleb Abduljalil

Department of Computing & Information Systems

Cardiff School of Management

Cardiff Metropolitan University

April 2017
Declaration

I hereby declare that this dissertation entitled *Critical Analysis of the Emerging Dark Web* is entirely my own work, and it has never been submitted nor is it currently being submitted for any other degree.

Candidate: *Abdulmuttaleb Abduljalil*

Signature:

Date: April 2017

Supervisor: *Tom Crick*

Signature:

Date:

Table of Contents

Critical Analysis of the Emerging Dark Web ................................................................. i
Declaration

Chapter 1: Introduction

1.1 Introduction
1.2 Background
1.3 Rationale of the Research
1.4 Problem Statement
1.5 Research Aim and Objective
1.6 Research Question
1.7 Research Structure
1.8 Conclusion

Chapter 2: Literature Review

2.1 Introduction
2.2 Incremental Forum Spidering
2.3 Multilingual Translation
2.4 Social Network Analysis on Web Forum
2.5 Computational Techniques
2.6 Dark Web Analysis and Visualization
2.7 Dark Web Collection
2.8 Terrorism Informatics
2.9 ISI (Intelligence and Security Information) Research Frameworks
2.10 Focused and Hidden Web Crawlers
2.11 Conclusion

Chapter 3: Research Methodology

3.1 Introduction
3.2 Methods Outline
3.3 Research Onion
3.4 Research Philosophy
3.5 Research Approach
3.6 Research Strategy
3.7 Research Design
3.8 Research Framework
3.9 Working Data Type
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.10 Data Collection</td>
<td>22</td>
</tr>
<tr>
<td>3.11 Research Ethics</td>
<td>22</td>
</tr>
<tr>
<td>3.12 Time Horizon</td>
<td>22</td>
</tr>
<tr>
<td>3.13 Summary</td>
<td>23</td>
</tr>
<tr>
<td>Chapter 4: Data Analysis</td>
<td>24</td>
</tr>
<tr>
<td>4.1 System Design</td>
<td>24</td>
</tr>
<tr>
<td>4.2 Data Set</td>
<td>27</td>
</tr>
<tr>
<td>4.3 System Functionality</td>
<td>28</td>
</tr>
<tr>
<td>4.4 Case Study for Identifying the Active participants in Dark Web Forum</td>
<td>29</td>
</tr>
<tr>
<td>Chapter 5: Conclusion and Recommendations</td>
<td>31</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>31</td>
</tr>
<tr>
<td>5.2 Linking with the objectives</td>
<td>31</td>
</tr>
<tr>
<td>5.3 Recommendation</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>34</td>
</tr>
<tr>
<td>Appendix A</td>
<td>37</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>3</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Introduction
The dark web is depicted as the segment of the deep web which in general terms is described as the hidden websites that users cannot easily get access to without the usage of the special software. The users can access the dark web with the expectation that they are being able to share the information and the file with having a little risk of detection. With the growth of the dark web use, the analogous of the early internet is being indicated, and the adopters of the internet are seen in the curious students and the marginalized groups with the criminals (Bradbury, 2014). The dark web also helps in explaining the common misconceptions that are pressed into the coverage in a vast area. The dark web seems to be confused when it is vastly explained with the help of the internet. The deep web simply explains the firewalls and the paywalls with including the IoT (Internet of Things). The keywords are indexed in the search engines, and also the users can easily view the indexed websites. Apart from this, all the internet traffic is directed between the routers and the networking devices which are depicted for the construction of the networking devices software. It also seems to be comprised of the dark network which is also known as the Tor Network (Cloete, 2012); Tor stands for The Onion Router which helps in adding the encryption to conceal the usage of the web user and makes it harder to be tracked.

1.2 Background
The dark web refers to the assortment of the websites, which present on an encrypted network and cannot be accessed by utilizing the conventional search engines or conventional browsers. Almost the entire websites on dark web use to conceal their identity by utilizing the Tor encryption as the Tor is capable of hiding the identity of the websites as well as the activities performed through the websites. World Wide Web (WWW) is a massive warehouse of the hyperlinked documents, which contains useful information. Currently, the exponential development in the field of information technology provides the opportunities to access a large amount of information via WWW. The WWW is broadly divided into two segments; surface web and deep web or dark web as per the depth of the information provided by the data from the website (Chen, 2012). The content
of the dark web is dynamically produced by the web server as well as a return to the users throughout the online query. To access the dark web, the users should request from a particular database through a search interface. Currently, it is observed that a large amount of data on the WWW is obtainable only through the search interface. Therefore, large amounts of information are still not visible to the users. As for example, while a user needs to search data of few train routes, they have to fill the details in a search form in the railway reservation site. Consequently, they obtain the dynamically-produced page including the details of the available trains. Hence, the portion of the web that is accessible only through the search interface is called the dark web. The research is essentiality of the dark web, specifically on the emerging importance of the dark web in present situations.

1.3 Rationale of the Research

Today, billions of people are using computer and internet facilities for their work and entertainment. The internet facilities have allowed the users to visit websites, conduct online business and communicate with others without leaving a trace of their location or identity. The technology experts, who have invented the internet, have developed an anonymized technology known as the Dark Web which is not visible to the users. The supports say that Dark Web helps to protect the dissidents in the repressive regimes, allowing military personal and police to conduct secret operations and increasing security for the users on the web (Gehl, 2016). On the other hand, the Dark Web has been criticized as a pathway for cybercrimes, e.g., used by sex traffickers, drug dealers, and child pornographers. But the advocate of civil liberties counters that the online anonymity can change the world scenario. It is the place where stolen documents and personal information are dumped/led or auctioned. The hackers are also using Dark web to expose or sell the data and information causing damage to the organizations and individuals. The UK government has also introduced a cybercrime unit to deal with dark Web with a focus on breaking down crime rings.

1.4 Problem Statement

The deep web/dark web is not referred to as the dark web. Anonymous Anonymity technology and hidden service sites are being widely used, and unsafe misconceptions are being repeatedly spread. The UK government has focused on how to tackle the dark web crimes and has also taken important steps to deal with it. However, the dark web is the part of the deep web/dark web that has been
hidden intentionally and is inaccessible through the web browsers. The dark website provides a platform for the users on the internet for whom the anonymity is necessary as they not only provide protection from the unauthorized users but it also includes encryption to obstruct monitoring. A known source for the contents that resides on the dark web can be found in the Tor network (Lowery, 2017). Thus, Tor network is the anonymous network that is accessed with the special browser referred to as the Tor browser. There is no difference between the real-world crime and virtual-world crime. The advancement of technology has also led to the emergence of crimes that has affected the society. However, the government has introduced cybercrime cell to deal with serious crimes.

1.5 Research Aim and Objective
The main aim of the research study is to recognize the importance of dark web in the current situation and to find out the positive and negative impact of dark web.

The core objectives of the research study are:

- To find out the activities that helped by the dark web in a positive way.
- To find out the activities that helped by the dark web in a negative way.
- To identify the ways those reduce the negative impacts of the dark web.
- To find out the processes that help in using the dark web in a way that is more constructive.

1.6 Research Question
The research question helped in providing adequate emphasis to the core issue of the research and based on the issue the research study use to be conducted. Therefore, as the research question focus on the core issue of the research study, it plays important role in dissertation study.

- What are the positive and negative impacts of dark web on the society?
- How the negative impact of the dark web can be controlled?

1.7 Research Structure
The research structure helps in depicting the explanation of the topic with an in-depth analysis of the dark web. The research structure provides the information related to the overview of the research which is considered for the explanation of the research process. The key contributions of the five chapters is also provided in this research structure which helps in illustrating the overview of the research.
Chapter 2: Literature Review
This chapter is conducted for gaining the information regarding the past researchers, and thereby the researcher can easily gain the resources for carrying out the analysis of the research on the dark web. The explanations are provided regarding the past researcher made by different researchers who help the researcher to gain the detailed information regarding the research. The importance of this chapter is being explained in an appropriate way which helps the researcher to gain the fixed goals and the target. The vital nature of this chapter is depicted in this structure of the research. The appropriate information can be gained by the researcher for carrying out the further research.

Chapter 3: Research Methodology
This chapter is used for gathering the raw data which are used for gathering the raw data which is being explained and thereby it also helps the researcher to gain the data for the purpose of structuring the research. It also helps the researcher for the selection of the tools and the equipment that are necessary for the purpose of constructing the research. It thereby also helps the researcher to structure the research. It also helps the researcher to recognize the crucial elements and thereby the appropriate construction with depicting the purpose of conducting the research. It also helps in structuring the purpose of the study with depicting the optimum usage of the tools and the techniques used for carrying out the research study.

Chapter 4: Data Analysis
The analysis of the data is being undertaken by the conduct of the data analysis which is undertaken by the researcher for the collection of the data and thereby the appropriate analysis can be undertaken by the researcher. It also helps in analyzing the information’s gathered for the study, and thus the overall research can be easily strengthened by the researcher. The gathered information and the data are appropriately utilized for the collection of research and thereby the appropriate illustration of the research work is being gathered.

Chapter 5: Conclusion
The conclusion chapter of the dissertation will provide the overall view and information of the research topic. The chapter will also link the objective of the research with the analysis of the
research study. The researcher will be able to extract the result of the research study by data, information, and findings. Apart from this, by the nature of the research study, the researcher will determine and evaluate the recommendations.

1.8 Conclusion
The technological advancements have several good effects, and at the same time, these show some negative effects on the society. Dark Web is an exceptional web technology (since it cannot be easily accessed), which supports in hiding the information from the end users. It is also observed that the dark web has a massive amount of quality information than the surfaced web and extraction of the dark web is highly effective for the general users as well as the specific users. The conventional web crawlers have a constriction in searching the deep web information (Weimann, 2015). Therefore, few of the web crawlers are specially made for crawling into the deep web information so that a large amount of essential data can be obtained. The dark web has positive as well as negative impact on the society. The UK authority applies it in right ways, whereas the criminals use it in negative ways.
Chapter 2: Literature Review

2.1 Introduction
The Dark Web is defined as the particular subsection of the deep web which is related to the collection of the websites. This process is continuing the hiding of the IP address, and thereby the server process can normally be detected by using encryption. The deliberation of the hiding section is being used for the purpose of detecting the buildings, and thereby it can be used both for the good and the bad. As the chapter literature review is being concerned, the importance and the essentiality of the project are being undertaken for the betterment of the surveillance regarding the state (Arora, Kanjilal and Varshney, 2012). The exchange of the weapons and the occurrence of the child pornography are easily conducted by the help of this Dark Web. The distribution of the network is being analyzed by the help of Tor network, and the users can easily afford the anonymous anonymity of the process.

Therefore for the conduct of the in-depth analysis, the various works of literature are provided for the enhancement of the research, and thereby the Tor routing with the other principles are provided by the help of the various US intelligence systems. It not only enables the Dark web process for the licit purpose but the illicit purpose also. The privacy of the system with depicting the appropriate analysis of the trackers of the network is being easily depicted for the purpose of analyzing the fact and also the research is being continued by the help of the ISI research frameworks (Wu, 2011). The conduct of the literature review is based on the detailed research on the various parts of the dark web which is being explained in an appropriate way. The research also helps in depicting the appropriate facts regarding the research which was conducted by the researcher.

2.2 Incremental Forum Spidering
Spiders are defined as the software programs that are used to transverse the World Wide Web information. The web documents can be easily retrieved by the process of the HTTP protocols and thereby the six characteristics features can be depicted for the process which is the accessibility, URL ordering features, collection type, collection update and the content richness (Barnett and Jiang, 2016). The function of the spider program is that the easy accessibility can be achieved by the process of handling the registration and thereby the exact and
the desired information from the various types can be easily collected (Barnett and Jiang, 2016). For the purpose of providing the comprehensive characteristics, the forum of the dark web is constructed with the various sources of the data in a sequential and the timely manner. This helps in the arrangement of the data, and thereby the appropriate collection process can be continued by the users. Therefore, the accessibility and the collection update issues can be easily removed with depicting the important characteristics. The accessibility and the collection process also consist of two vital issues which are appropriately explained in the following points:

- The accessibility problems can be resolved by the process of the human-assisted approaches and thereby for the purpose of the collection approach, the Incremental Spidering and the period Spidering approaches are undertaken. The proper process is being adopted for the purpose of conducting the appropriately targeted forums with depicting the betterment of the system.

- Due to the collection size of the forum portal of the Dark web, the Spidering of the entire collection can be easily indicated for the purpose of making the appropriate attempts. While the attempts are made, the problems in the downloading URLs are being depicted which is seemed to starting with the link of URL and thereby the potential threats can be depicted with the malicious attacks by the malicious links (Bayati et al., 2010).

With the adoption of the incremental spidering process, the minimization of the problems can be seen with depicting the appropriate, feasible solution for the collection with Up-to-date. For the purpose of collection updating, the incremental Spidering is made to be focused on the downloading process, and thereby the appropriately targeted forums can be constructed with security. It also helps in constructing the new contents that are important for the targeted pages with making it malicious free links.

2.3 Multilingual Translation

The Multilingual issue is the most vital issue which is indicated in the research of the Dark web and thereby the various forms of the research is being carried out by the appropriate Jihadist content. The variation of the languages helps in indicating the appropriate issue which is being clearly indicated in the languages such as the Arabic, German, French, etc. In this, the multilingual context seems to be adopting the different methods that are crucial for the enhancement of the
research regarding the translation tasks. The approaches that are undertaken for the translation process are the translation-based approach, machine translation based approach, corpus-based approach and the dictionary-based approach (Berendt, Hotho and Stumme, 2010). The machine-based approach is showing the existing machine based language that is essential for the purpose of providing the automatic translation, and thereby the approaches that are undertaken seems to be very much similar in the context. In the case of the dictionary-based approach, the bilingual dictionary is being constructed which helps in obtaining the translation based results in an appropriate way. The translation result seems to be obtained by the process of looking into the dictionary, and thereby the term can be easily translated if seems to be present in the dictionary. In the case of the corpus-based approach, the analysis of the large documents based research is being carried out, and thereby the translations seem to be completed in a statistical manner (Brugger, 2012). The process of mixing these approaches is carried for the appropriate construction of Google translate and thereby the popular machine translation tool is being defined in the Google translation services. The appropriate development of the services is defined for the purpose of providing the eighty languages which can be easily translated for the benefit of the user and thereby the proper translation of the English can be done in any of the languages.

2.4 Social Network Analysis on Web Forum
The social network analysis is being conducted for the purpose of obtaining the graph-based methods, and thereby the analysis of the network group becomes easier with depicting the group or the population strength. The impact of the social interactions is widely depicted by the usage of the social networks, and thereby the real world networks can be easily identified. The social networks are formed by the usage of the Dark networks which are crucial for building the sites of the social networking. The Dark networks are either real or virtual, and thereby the criminal activities can be conducted in an easy way (Egoryan, 2015). The analysis and the study clearly depict the criminal activities in the social networks which are harmful to the society. The terrorist activities are organized by the help of the Dark network, and thereby the religious opinions are targeted in the issues with creating the religion discrimination. Thus the terrorist web forums are created by the process of the important information gathering regarding the weapons used in the world of social networking. The important resources are also collected to defense the counter-terrorism groups constructed in the world.
In the form of the web forums, the threads are easily posted which are displayed automatically from time to time. The starter of the post and the other threads clearly represents the thread starter and therefore the social network in the web forums is seemed to be based on the thread structure as it is known in the practical world is reply structure (Gehl, 2016). The reply network is being depicted as the bipartite graph structure that consists of the various nodes and the links. Henceforth each node or the every link seems to be representing the thread starter and thereby the author starts replying to that provided thread.

The metrics of the social networking is measured by carrying out various measurements regarding the reflection and the importance of the connectivity of the node as measured by the SNA (Systems Network Architecture). Various several types of the metrics are constructed for the purpose of having the better connectivity, and thereby the inclusion of the centrality, cohesion and the reachability is being depicted. The appropriate advancement of the research of the social networking is done on the basis of the measurement metrics, and thereby the appropriate approaches can be easily explained. The centrality is a set of the metrics that are described as the importance of the nodes reflecting the connection of the network (Fersini, Pozzi and Messina, 2016). Therefore the inclusion of the distance, degree and the closeness clearly constructs the centrality of the network. The cohesion is described as the performance seen at the nodes that are connected to each other and thereby the term is defined as the clique. The measurement of the cohesion is done by clustering the nodes coefficient, and therefore the higher value obtained indicates the node and the first neighbor. The reachability clearly describes the connectivity of the two nodes that are formed together to have the appropriate construction of the connection with depicting the number of edges in the path which connects the two nodes. In addition to the other classic metrics, the closeness and the distance can be easily depicted for the purpose of assuming the relationship between the two nodes. Therefore, it provides an easy way for the purpose of estimating the importance of the web page network (Harrison, Roberts and Hernandez-Castro, 2016). Therefore the presentation of the importance clearly states the appropriate values in the form of the response as gathered in this case. The web page networks which are defined in this case are the Page Rank scores.
2.5 Computational Techniques

The computational techniques include data mining, text mining and Web mining. The communicational tools are grouped into the collection, analysis and visualization, significant deep web, dark web, spidering and the computational linguistic. The visualization of the research is being conducted for the development of the social media analysis, and thereby the visualization of the research is being depicted by the researchers. The development of the deep web, dark web content is being made by the spiders/ crawlers which collect the deep web, dark web content from the research of the library contents. The spiders are seemed to be designed for fetching textual documents: HTML, PDF and Word files with the links that are the PHP, CGI and the files of ASP (Jo and Im, 2010). The spidering tool helps in recognizing the forum hosting software with their individual formats, and thereby the reconstruction is being continued by the help of the posting the figures useful for the interaction. The periodic forum functioning is being recorded for the collection and the processing of the forum contents that are used for the purpose of incremental updates which are used for depicting the needs of the research. The languages are being selected for the purpose of carrying out the computational linguistic techniques and thereby the development is seemed to be successful with depicting starting of the structure. The development of the specialized techniques for the collection of the multimedia files and the attachments from the forum are also constructed for the purpose of performing the stenographic research (Weimann, 2015). This also helps in identifying the images with depicting the collection and the multimedia analysis in the segment of the video representation and the image recognition. This also helps in identifying the unique terrorist activities which are the recorded video contents and the styles. The organization of the criminal activities includes the terrorist that often seems to be forming the connected covert networks which shows the illegal activities. Therefore often the dark web is being constructed for the purpose of exhibiting the unique structures in the organization and thereby the communication channels and the resilience to work are being attacked for the purpose of disruption (Jo and Im, 2011). Therefore the new computational techniques that are implemented for the betterment of the social network analysis include the network learning and the network topological analysis which are included for the purpose of having the random network. This will also help in minimizing the terrorist attacks, and the systematic study is being undertaken for the purpose of constructing the systematic and the criminal network analysis in an appropriate way. Thus, the
success can be easily measured for the counter-terrorist squad and thereby the minimization of the terrorism can be seen with implementing the complex network structure.

2.6 Dark web Analysis and Visualization

Different SNA techniques have developed for examining forum posting and website relationships. The topological metrics such as degree, betweenness, etc. and properties such as growth, preferential attachment, etc. have also been developed and used to model the terrorist site and terrorist interactions. Several clustering such as block modeling and projections such as spring embedder, multidimensional scaling techniques have been developed to visualize their relationships. The main focus is on understanding the “dark networks” and their unique properties (Davies, 2008). Detailed coding schemes have been developed to evaluate the extremist websites and terrorist contents. The categories of the content include training, recruiting, communication, propaganda, sharing ideology, etc. The computer programs help to identify automatically the selected categories such as forum availability, web master information, etc. The web metrics analysis evaluates the media richness, technical sophistication, terrorist websites and web interactivity of the extremist. The technical capabilities and features have been examined to determine and evaluate the web saviness level. The web metrics gives a measure of resources and capability of the extremists/terrorists. All the website metrics of the terrorists are computed and extracted using the computer programs. However, not all the sites are violent or radical equally. The sentiment and affect analysis allow determining violent and radical sites that impose significant threats (Clifton, 2012). The research also shows how the radical ideas and techniques become infectious on the basis of the sender and content interactions. The visualization techniques have been developed to examine the affect/sentiment changes among people and in time. The research includes different probabilistic multilingual affects lexicons and projection and dimension reduction techniques.

The technique (https://www.scss.tcd.ie/Khurshid.Ahmad/Research/Sentiments/K_/_a7-abbasi.pdf) has been developed to determine and analyze the anonymous senders on the basis of the signature association with the forum messages. The syntactic and lexical features of the traditional analysis expanded to include systems such as font size, web links, color and semantic such as racism, violence features of the relevance to the online texts of terrorists and extremists. The visualization and advanced Ink blot have developed to identify the web...
signatures. The write print technique has been developed for English, Chinese and Arabic language (Moran and Hunt, 2009). The write print in Arabic consisted of more than four hundred features and extracted automatically from the messages online using the computer programs. The write print technique can achieve an accuracy level of around 95%. A unique coding scheme has been developed in order to analyze the videos generated by the terrorists on the basis of production characteristic, metadata, and contents associated with the videos. The semi-automated tools allow the investigators to accurately and quickly code and analyze the videos.

The analysts have also developed programs to determine IED (improvised explosive device) related contents generated by insurgency groups and terrorist groups in Dark Web. There are small numbers of sites that disturb large percentage of web pages that are related to IED, forum posting, explosive videos, training materials, etc. Signatures have also been developed for the IED sites on the basis of contents, multimedia and content characteristics (Willard, 2011). Training materials are used for the identification of the troops before their any illegal activity occurs. The online platform provides a wide opportunity to the terrorists for exchanging their information and data. The Data Web portal forum provides access to 30 forums that together comprises more than 14,000,000 messages from 350,000 participants in different languages. The portal provides download, statistical analysis, social network visualization and translation functions for the selected forum.

2.7 Dark Web Collection

The crawlers/spiders have been developed to collect deep web/dark web contents. The spiders provide access to the password protected sites and also perform randomized fetching. The spiders are provided with adequate training to fetch all pdf, Word files, HTML, PHP, CGI, links, images, ASP files, videos and audios in a web site. The spidering forum tool determines forum hosting software around 15 and their format. The complete forum includes headings, authors, threads, posting, time tags that allow reconstructing the interactions of the participants. The periodic spidering forum and incremental updates have been perfumed on the basis of the research needs. The forum contents are collected and processes in English, Arabic, French, Chinese and Spanish using the linguistics computational techniques. Specialized techniques have been developed for collecting multimedia attachments and files from forum and web sites and spidering (Willard, 2011). The stenography research helps to determine encrypted images in the multimedia analysis and collection such as videos generated by terrorists. The first step to study the presence of the
terrorists and extremists on the web is to capture the web sites operated by the terrorists and storing them in the repository for future analysis.

The attributes of the web interactivity can be summarized into the following categories: transaction interactivity level, community interactivity level and one to one interactivity level. The one to one interactivity level attributes measure how well the websites support users to give their opinion to the owner of the web site. The community interactivity level attributes measure how well the website supports the two-way interactions between multiple users and site owners. The transaction interactivity level measures how well the users are allowed to accomplish tasks on web sites. The first step to study the use of the web by a terrorist is to build a high quality and effective Dark Web Collection. The semi-automated approach has been developed to collect the contents of Dark Web.

The approach of the collection contains the following steps: identifying the extremist/terrorist groups, continuing the collection process by identifying the groups through the authoritative source (Chen, 2008). The sources include the government report, organization report and studies published by the researchers. Once the websites of the extremist/terrorist are identified, then a program is used to download automatically all the contents. The automatic method of downloading allows building an effective dark web collection with huge numbers of documents.

2.8 Terrorism Informatics

Terrorism informatics is referred to as the application of advanced information fusion, analysis techniques and methodologies to acquire, process, integrate, manage and analyze the diversity of the information related to terrorism for international/national security applications. The technique is derived from the disciplines such as informatics, mathematics, science, statistics, social sciences, public policy, and linguistics. The research shows terrorism involves a huge amount of information from different sources, languages, data types, information fusion and analysis such as text mining, data mining, language translation, data integration, video and image processing helps to detect and prevent terrorism (Chen and Yang, 2008). The investigators and analysts are analyzing and gathering information on modern terrorism, and it has become important to understand the terror network. The Dark Web research program of Arizona Artificial Intelligence Lab University has made complementary efforts in evaluating the emerging discipline of the treason informatics. The

1 www.eisic.eu/eisic2011/hchen.aspx
Dark Web program adopts the system, computational approach, and data to study terrorism and terror on the internet. The collection and analysis of a longitudinal, fluid and large-scale collection of content generated by terrorists using the computer programs are determined closely. The terrorism informatics heavily depends on terrorism domain database and knowledge. On the basis of the Dark Web, the terrorism centers are grouped in following categories: terrorism online sources and database, higher education institutes and think intelligence sources, this is clearly evidenced in the book Gal, Kantor and Shapira, (2008). Internet Haganah is involved highly in disabling and monitoring websites of terrorists on the web. It alerts any counterterrorism and fosters indirectly information “warfare”.

The terrorism informatics is growing rapidly, and it has been found that there is a broader collaboration between computational researchers and social scientists who are interested in the counterterrorism research. It is also expected that the new methodologies, computational approaches and terrorism databases should be based on the complex interactions and rich contents produced by the extremists and terrorists on the internet (Yang, 2010). The international community is also focusing on the terrorism, and it has affected the society significantly. The emergence of terrorism has been gathering popularity in the world. The investigations are still going on the web sites that are operated by the terrorists.

2.9 ISI (Intelligence and Security Research Information) Research Frameworks

Crime is an act that is forbidden or prohibited by the law. The offender is punished by the law considering the act as an illegal act. The more threat that a crime poses on the common people the more it is likely to be of a security concern to the country. The crimes such as theft, homicide, and traffic violations are in the jurisdiction of the local law enforcement organizations. The other types of crimes need to be dealt with both national security and local law enforcement authorities depend on the country UK. The identification of fraud and theft are relevant at both the national and local level- the criminals may escape by using false identities, and the smugglers can also enter the country by holding fake visas or passports (Yang, 2013). Internet fraud, network hacking, intrusion, illegal trading, hate crimes, virus spreading, cyber pornography, cyber privacy, theft of the confidential information and cyber terrorism, narcotics trafficking and terrorism are diffuse often geographically resulting in the security concern all across states, cities, and countries. The cybercrimes impose significant threats to the public safety across all the jurisdictional areas. Domestic and international terrorism involves multiple types of crime and causes great damage.
The knowledge discovery in databases techniques play a significant role in improving the counterterrorism and fighting capabilities of security, intelligence and enforcement of the law by decreasing the information and cognitive overload by the ISI.

The traditional discovery knowledge techniques include prediction and classification, rules mining, outlier analysis and cluster analysis. The research advances, text mining approaches and natural language processing that automatically summarize, extract, translate and categorize text documents that have been used widely. On the basis of the special characteristics of criminals, security data, and crimes, the existing ISI technology was divided into six categories: crime association mining, information collaboration and sharing, crime clustering and classification, intelligence text mining, temporal and spatial crime mining and network mining. The six categories are based on the traditional discovery knowledge technologies and new approaches have been added that includes temporal and spatial crime mining and the criminal network analysis (Yang, 2010). Several unique data mining techniques classes are of great relevance to the ISI research. The text mining is crucial for extracting the key entities (places, people, weapons, narcotics, time, etc. and the relationship presented in the voluminous police reports, open source clips, intelligence reports, etc. (Qaiser Saleem, 2014). Some of the techniques should be multilingual in nature which includes cross-lingual information retrieval and abilities for the machine translation.

Temporal and spatial mining and visualization are needed for the geographic information system and analysis of terrorist and criminal events. Terrorists and organized criminals form interconnected networks for the illegal activities. It is often referred to as the dark networks, and computational techniques include network learning, network topological analysis, and social network analysis are needed for the systematic study of the covert and complex networks. There are many regulations, agreements and regulations governing confidentiality, data collection, and reporting that exists which could have a direct impact on the application and development of ISI technologies

2.10 Focused and Hidden Web Crawlers
Focused crawlers are defined as seek, index, maintain and acquire pages on a specific topic that represents the narrow segment of the web. The need to gather high quality, specific domain contents results in significant characteristics for the crawlers that are relevant for the collection of the Dark Web forums. Some of the characteristics are specific to hidden and/or focused web
crawling, and others are relevant to the types of spiders. Most of the search engines cover that are referred to as "publicly indexable web." this is the part of the web that is accessible easily with the traditional web crawlers. A large portion of the internet is generated dynamically, such contents require users typically to have prior fill out forms, authorization or to register (Schäfer and Bildhauer, 2013). The covert side of the internet is referred to as the deep/invisible/deep web dark web. The hidden web contents are stored often in the specialized databases. For example, the IMBD movie review databases contain plethora of information, yet the standard crawlers are not able to access the information. According to a research study, conducted in the year 2000, it has been found that invisible web contains approximately 500 times more information that is presented in the traditional surface web. The general categories have been introduced to access hidden web through automated web crawlers. Thus, the first approach shows the use of the automated form filing methods. Different query generation automated approaches for querying such as fetching dynamically generated contents and hidden web databases have been proposed. Other methods keep the index of the hidden search engines and redirecting the queries of the users, however, different automated approaches exclude/ignores querying or collection of the page needed to login. The automated form filling method is a problem for the Dark Web forums as it requires a login. The second approach is to access the hidden web, and it provides a semi-automated framework allowing the experts to help the crawlers in achieving access to the hidden contents. The involvement of human interaction is depended on the complexity of the accessibility problems faced, an example of this would be that simple forums ask for an email address and the name as well as other information needed to login, this can be automated with the standardize responses. The approach is suitable for the Dark Web this is due to the fact that the access process complexity varies significantly. There are also many types of research on collecting weblogs. The Blog Pulse portal\(^2\) is the blog portal analysis, the site contains an examination of discussion on trends/topics for around 150,000 spidered weblogs (Willard, 2011). The blog can be useful for the marketing intelligence. The blog contains the review of the products, and it is analyzed using the sentiment technique analysis. The web crawling forum represents a unique set of problems. The discovery of the web forums is considered to be the challenging task due to the lack of the centralized index. The web forum needs extraction of information for the derivation of the metadata. The wrappers are significant for incremental analysis and data analysis. The BoardPulse is the system to harvest

---

\(^2\) https://help.twitch.tv/customer/portal/articles/2752590-how-to-use-pulse
the messages from the online forums. Wrapper and crawler are the two components. The web includes multimedia and indexable files. The indexable files include static text files and dynamic text files. The multimedia files include animation, images, video files, and audio.

2.11 Conclusion

The overall literature review consists of the various components which are essential for the purpose of describing the structure of “the dark web.” The dark web analysis with the visualization and the dark web collection is being depicted which shows the enhancement of the research. The construction of the research seems to be fruitful which covers each and every point of the analysis requisite for the construction of the network system. The terrorism informatics is being provided with the ISI research that provides the appropriate structure of the research. The significant previous works are described in the context of the Dark Web Collection by the help of the Incremental spidering analysis and the multilingual component that helps in the construction of the language translation system (Sadan and Schwartz, 2011). The Dark Web seems to be providing a heterogeneous structure which is useful for the terrorists to carry out the criminal activities and the necessary changes can be indicated for the purpose of banning the terrorist activities. Apart from this, the focus on the Hidden web crawlers is depicted for the purpose of enhancing the sentiment analysis and thus the future work can be constructed for making a pause on the networking system with securitizing the network in a more complex way. The deep mining and the analysis of the interface of the user are being described in this literature review which helps the research to carry out the appropriate and the further research with accomplishing the prefixed aims and the objectives of the research.
Chapter 3: Research Methodology

3.1 Introduction

The research methodology chapter presents methods, tools and techniques, to help in selecting the most appropriate research tools and techniques as per the objectives of the dissertation and the characteristics of the research study. The implementation of the appropriate research philosophy, approaches, strategy, design and framework help in collecting the necessary data and the research methodology also supports the data analysis procedures (Willard and Bellamy, 2012). The undertaken research study is on dark web and the objective of the dissertation study is to study the impact of dark web on society; positive as well as negative and how to reduce the negative impact of the dark web. The tools such as the research philosophy, research approach, research design, research strategy, time horizon and data analysis helps in conducting the dissertation study in proper way. The dissertation study provides great emphasis on the different aspects of the dark web; it has positive and negative impacts on the society. The research methodology chapter provides clear and precise justification of the various methods applied in this research study and depict different methods of data collection for the specific research study. The research study includes of systematic as well as logical procedures of collecting relevant data and information about the undertaken dissertation study so that help in completing the research project in a proper way and thus accomplishing the undertaken research objectives (Jha, 2008). The research procedures consist of the systematic ways of discovering and revising theories, facts and applications that support the dissertation study to accomplish its primary objectives.

3.2 Methods Outline

In order to conduct research on the dark web a positivism philosophy has been adopted to examine the information, which is used in the dissertation study. The deductive approach is a selected research approach as the deductive approach is a bottom-up approach. The descriptive research design has been chosen for this particular research to conduct a deep detail research on the specific subject. Besides this, the researcher has selected the non-experimental research strategy in order to conduct the research as the data and information are collected for this particular research from the secondary resources. The necessary data for the research has been collected from the secondary resources and analyzed to reach the conclusion of the dissertation study.
3.3 Research Onion
The research onion is one of the vital tools of research methodology and it provides significant contribution to the dissertation study. With the help of the research onion, the researchers can be able to select the most appropriate tools and techniques according to the subject of the dissertation study. Each layer of the research onion allows the researchers to choose their preferred tools and techniques from a range of different tools and techniques. Research onion provides comprehensive research process and in this helps the researchers to understand the research processes in ways that are more convenient. It helps in understanding as well as addressing the pattern of the research study in proper way. Moreover, with the help of the research onion the researcher can conduct the research in more perspective way (Ketchen and Bergh, 2006). By applying the techniques of the research onion, the researchers conduct the research on the impact of dark web on the society in a way that is appropriate. It helps in achieving superior quality of research methodology for the specific research.

3.4 Research Philosophy
The research philosophy helps in defining the development of the knowledge that achieved during the research study. Besides this, it helps in understand the distinguished nature of information and knowledge. In this way, help in better description of the research study, and also help in taking better approach for the particular dissertation study. There are mainly three types of research philosophy applied in the dissertation study; positivism, realism and interpretivism. In positivism research philosophy, the positive verification use to be done through the research study. The realism research philosophy intends to interpret the reality present around the world. It underpins the thought that the human mind is influential by effective reasoning as well as clarification. Lastly, the interpretivism intends to define the world, where the external elements are unavailable as it is a place of non-objectivity. Depending on the appropriateness of the dissertation subject, the researchers select the philosophy of their research so that the subject can be investigated more proper way (Kothari, 2004). For the undertaken dissertation study, the researcher has adopted positivism research philosophy as it provides adequate scope to the researcher to study the research subject.
3.5 Research Approach

The research approach support the researchers in defining the development pattern of the undertaken research study. The research approach establishes a model of study, which help in observing the progress of the research study as well as verifying the content of the research study. There are two types of research approaches; deductive approach and inductive approach. The deductive research approach helps in observing different source of information. On the other hand, the inductive approach intends to explain a topic that deficient in any early research. It helps in building a new theory with the help of the research findings and result. For this particular research, the researcher needs help of the various theories that explored from the secondary research. The application of the deductive approaches match the study pattern of the specific dissertation study thus, the researcher select the deductive approach for this particular dissertation study. Apart from this, as the inductive approach avoid the cross-sectional study it is not appropriate for the study.

3.6 Research Strategy

The research strategy helps in choosing different method during the research process so that the research study can be conducted in a proper way. The method is composed of various types of investigation like experimental as well as non-experimental research study. The scientific research studies are mainly conducted with the experimental research strategy, where the scientific techniques help in dealing with different problematic scenarios. The intelligence and security research are mainly non-experimental research where the researchers mainly depend on the data and information from secondary resources. The data from the various secondary resources are used in this particular research in order to study the pros and cons off the dark web technology and the impact of the dark web on the society (Rabinovich and Cheon, 2011). Therefore, the researcher adopts the non-experimental research strategy for this specific dissertation study.

3.7 Research Design

The research design play vital role as it helps making the blueprint for recording the contents of the research procedures during the research study. There are primarily three types of research design applied in dissertation study: explanatory, descriptive, and exploratory research design. The explanatory research design use to be applied to follow a pre-planned procedure of research study (Vartanian, 2011). Alternatively, the descriptive research design intends to extract the reason behind the research findings by testifying with several theories of the research study. Furthermore,
the exploratory design is a mixed application that support both the explanatory and descriptive design. For this study of the dark web, the descriptive research design has been adopted as it supports in deep detail analysis of the undertaken dissertation study.

3.8 Research Framework
The intelligence and security area does not have well defined data sources like any other domain of research such as marketing, finance, as well as medicine. The sales records from the business organizations provides adequate data for the marketing research as well as medical history of the patients can be collected from the hospital that play vital role in medicine research. Therefore, in order to conduct a research study on the domain of intelligence and security both types of information; authoritative information and open source of information can be used. The examples of the authoritative information are telephone record, reports of crime incidents, financial statements, immigration, and customs records. Similarly, the examples of the open sources of information are books, journal articles, news articles, websites etc. The data for the dissertation study on the intelligence and security domain use to be collected from these various sources and the sources are often in various range of different formats such as structured database records, image file, audio and video file (Scruggs and Mastropieri, 2006). These sources may contain important information like criminal association however, the contents are often unstructured, multilingual text and difficult to access as well as retrieve. Besides this, often because of massive volume of data it is difficult to extract important data with credible intelligence and knowledge. Crime is a behavior of committing a forbidden act or neglecting of a duty or responsibility, which is defined by a public law. It makes the offender liable to punishment under the specific law. It is often noticed that the criminals use the dark web to disseminate their criminal activities such as to lure the small criminals to be associated with them or spreading terror.

3.9 Working Data Type
Data plays an important role in the dissertation study and the result of the dissertation study mostly depend on the collection and analysis of the data. Moreover, depending on the characteristics of the data and the topic of the research study the researchers adopt the methods of data collection. The data are mainly two types’ primary data and secondary data. The primary data are mainly first hand data collected by the researchers by their own efforts. On the other hand, the secondary data are mainly second hand data, extracted from the previous research work conducted by former
researchers on the same topic in which the researcher want to conduct dissertation study. The primary data delivers most relevant facts and current information. The secondary data can be collected from various books, academic journals and conferences, new articles and websites that contains information about the undertaken topic of the dissertation study. The primary data can be again segmented into two parts quantitative data and qualitative data. The quantitative data are short form of data mostly numeric data that help in gaining widespread knowledge about the research topic (Kothari, 2004). The qualitative data are descriptive data that provides deep detail knowledge about the dissertation study. The secondary data help in gaining the basic idea about the research study. This particular study is conducted with the secondary data as the primary data is unavailable for this particular research topic.

3.10 Data Collection
The necessary data for the dissertation study were collected from the secondary resources. Moreover, for this purpose, several books, academic journal articles and conference papers, new articles, websites has been explored and reviewed by the researcher.

3.11 Research Ethics
The research ethics comprise with the certain rules and regulations that the researchers require to maintain in order to conduct a dissertation study. The ethical consideration supports the researcher in maintaining right code of conduct during the research study so that help in avoiding the different types of issues that may occur during the research study (Welman et al., 2005). The research work done by the researcher is for only academic use and it is strictly maintained that the findings of the research will not use for any commercial purpose.

3.12 Time Horizon

<table>
<thead>
<tr>
<th>Primary Activities/ Stages</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 5</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic is chosen and defining its scope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition of secondary data sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of research proposal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of literature review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.13 Summary

The research methodology helps in selecting the most appropriate research tools and techniques for the dissertation study of dark web. The chosen research tools help to examine the research topic in deep detail way for a better result. By the chosen research, techniques the data analysis would be conducted that help in understanding the undertaken subject in a better way. The undertaken techniques and the structures enables to have an appropriate explanation of the research.
Chapter 4: Data Analysis

4.1 System Design

The Dark Web Forum Portal system (link.springer.com/chapter/10.1007%2F978-1-4614-1557-2_13) consists of the following three components which are the data acquisition, data preparation, and the system functionality. The overall system design is indicated as same as the previous designs, therefore the increment of the Spidering component is seen with regularly updating the collection.

![System Design of Dark Web Forum Portal](source)

**Figure 1:** System Design of Dark Web Forum Portal

Source: Abraham and Hassani, (2012)
The detailing of the components is provided below which focuses on the following sections as listed below:

- **Data Acquisition:** In the case of the spidering process, the specific programs are developed for the enhancement of the online forums that contains the Jihadist and other related contents that are identified by the domain experts. The component of the spidering is composed together for the purpose of the completing the process of spidering. The forums generally consist of the process that exists in the collection and when the forum is being added to the spidering, the complete spidering process is being applied for the collection of the available postings and thereby, the incremental spiders are designed for the collection of postings (Abraham, 2012). The incremental spidering is designed for identifying the collection of the postings and thereby the last update is being posted in the forum with depicting the small portion of the forum. In order to achieve the small portion of the forum data, the spidering process is being processed which will create a much more efficient goal for the development of the collection in each forum. The increment of the process of spidering consists of the three main steps which are the sub forum list page spidering, incremental spidering and the thread list page spidering. The incremental spidering helps in depicting the first spider and the parse sub forum.
space pages which are represented in the various themes. The thread list s\$pidering consists of the list of the pages related to the metadata regarding the discussion and thereby the discussion threads are sorted for the betterment of the updates. For each sub-forum, the threads are being discussed and thereby the date of the latest postings is interrelated for the construction of the databases and thereby the new threads are created with the collection of the databases. If the thread is being listed on the first page of the thread page, the s\$pidering of the forum is shifted to the next forum. The collection of the URL is depicted for the new threads which will help in the enhancement of the postings (Abraham and Hassanien, 2012). The conduct of the experiment is being made by the Incremental s\$pidering using the Hanin Net Forum. The forum is depicted as most active which is being identified by the domain experts and thereby the experiment regarding the results are depicted to be near the upper bounded case regarding the postings made in the forum. Henceforth, this process clearly indicates the domain experts regarding the forum list and thereby the collections of the postings are depicted for the purpose of providing promising solution regarding the forum update.

- **Data Preparation:** This component is used for forum parsing the programs which are developed for extracting the detailed forum data from the raw HTML, Web Pages and store it in the local database. For each forum, the detailed extraction of the threads is very much requisite for the extracting the thread names, messages, post dates and the member names.

- **System Functionality:** Various and the different set of the functions are developed and incorporated in the system with indicating the real-time services, including single and the multiple forums and the detailed database is being structured for the statistical analysis. The development of the forums, browsing and searching with including the multilingual translations. The Dark Web Forum portal is being included in the database for the implementation of the Apache Tomcat and the data base is being implemented with using the Microsoft SQL (Abraham, Hassanien and Snášel, 2010). The function of the multilingual translations is depicted for the implementation of the Google Translate service which seems to be able to detect the Non-English texts. The social network virtualization is being used to visualize the interactions among the forum members. Therefore the interactions are being made by using the JUNG source for the purpose of
visualizing the interactions among the forum members. The JUNG source is defined as the java universal network or the graph framework which is being used an open source library.

4.2 Data Set

The forums are appropriately incorporated in the system which consists of the total 29 forums. Out of the 29 forums, there are 17 Arabic forums, 7 are English forums and the 3 are the French forums. There are also two Germany and Russian forums which are provided in the table provided below:-

<table>
<thead>
<tr>
<th>Name</th>
<th>Language</th>
<th>Time Span</th>
<th>No of Members</th>
<th>No of Threads</th>
<th>No of Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-I3oraq</td>
<td>Arabic</td>
<td>01/08/2006 - 1/2/2010</td>
<td>3,503</td>
<td>52,322</td>
<td>223,648</td>
</tr>
<tr>
<td>Al-Fallujah</td>
<td>Arabic</td>
<td>09/19/2006 - 006- 2/2005</td>
<td>5,853</td>
<td>74,899</td>
<td>547,712</td>
</tr>
<tr>
<td>Al-FM:laws</td>
<td>Arabic</td>
<td>01 - 2/2005</td>
<td>2,187</td>
<td>9,359</td>
<td>38,382</td>
</tr>
<tr>
<td>Hfidad al-Suyuf</td>
<td>Arabic</td>
<td>03/18/2006-</td>
<td>1,597</td>
<td>8,096</td>
<td>55,947</td>
</tr>
<tr>
<td>Alokab</td>
<td>Arabic</td>
<td>04/08/2005-</td>
<td>1,547</td>
<td>12,097</td>
<td>23,709</td>
</tr>
<tr>
<td>Al-Qimmah</td>
<td>Arabic</td>
<td>11/23/2007-</td>
<td>287</td>
<td>147,598</td>
<td>1,227,207</td>
</tr>
<tr>
<td>Alsayra</td>
<td>Arabic</td>
<td>04/05/2001-</td>
<td>66,703</td>
<td>1,224</td>
<td>46,928</td>
</tr>
<tr>
<td>Ansar</td>
<td>Arabic</td>
<td>11/07/2008-</td>
<td>313</td>
<td>12,041</td>
<td>1,227,207</td>
</tr>
<tr>
<td>At-tahadi</td>
<td>Arabic</td>
<td>04/14/2008-</td>
<td>12,215</td>
<td>113,579</td>
<td>2,251,553</td>
</tr>
<tr>
<td>Hawaa World</td>
<td>Arabic</td>
<td>01/01/2001-</td>
<td>113,579</td>
<td>40,501</td>
<td>2,251,553</td>
</tr>
<tr>
<td>Hadramout</td>
<td>Arabic</td>
<td>11/25/2000-</td>
<td>19,491</td>
<td>151,694</td>
<td>1,552,227</td>
</tr>
<tr>
<td>Ma'an</td>
<td>Arabic</td>
<td>07/29/2007-</td>
<td>1,880</td>
<td>57,047</td>
<td>1,552,227</td>
</tr>
<tr>
<td>Al-Mujahidin</td>
<td>Arabic</td>
<td>11/09/2007-</td>
<td>4,259</td>
<td>29,980</td>
<td>140,930</td>
</tr>
<tr>
<td>Montada</td>
<td>Arabic</td>
<td>10/08/1985-</td>
<td>40,291</td>
<td>120,181</td>
<td>1,412,028</td>
</tr>
<tr>
<td>Ana al-Muslim</td>
<td>Arabic</td>
<td>03/21/2007-</td>
<td>12,215</td>
<td>179,791</td>
<td>1,343,370</td>
</tr>
<tr>
<td>Shumukh</td>
<td>Arabic</td>
<td>12/08/2008-</td>
<td>3,938</td>
<td>46,666</td>
<td>289,201</td>
</tr>
<tr>
<td>Ansar</td>
<td>English</td>
<td>04/28/2004-</td>
<td>377</td>
<td>129,056</td>
<td>29,056</td>
</tr>
<tr>
<td>Gawaher</td>
<td>English</td>
<td>06/09/2004-</td>
<td>6,790</td>
<td>210,656</td>
<td>569,709</td>
</tr>
<tr>
<td>akening</td>
<td>English</td>
<td>11/14/2000-</td>
<td>2,361</td>
<td>25,112</td>
<td>116,009</td>
</tr>
<tr>
<td>Islamic Network</td>
<td>English</td>
<td>06/01/2008-</td>
<td>1,573</td>
<td>25,112</td>
<td>116,009</td>
</tr>
<tr>
<td>Islamic Web</td>
<td>English</td>
<td>10/06/2004-</td>
<td>745</td>
<td>25,112</td>
<td>116,009</td>
</tr>
<tr>
<td>Tim To Islam</td>
<td>English</td>
<td>05/05/2002-</td>
<td>9,926</td>
<td>38,702</td>
<td>308,970</td>
</tr>
<tr>
<td>Ummah</td>
<td>English</td>
<td>02/27/2009-</td>
<td>14,349</td>
<td>71,218</td>
<td>1,192,583</td>
</tr>
<tr>
<td>Al Minha Dj</td>
<td>French</td>
<td>03/21/2003-</td>
<td>313</td>
<td>2,007</td>
<td>6,421</td>
</tr>
<tr>
<td>Forums S' aslama</td>
<td>French</td>
<td>10/12/2004-</td>
<td>2,665</td>
<td>20,468</td>
<td>131,559</td>
</tr>
<tr>
<td>Language</td>
<td>Forum</td>
<td>Start Date</td>
<td>Messages</td>
<td>Posts</td>
<td>Users</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>------------</td>
<td>----------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>French</td>
<td>Al-Mourabitoune</td>
<td>06/08/2006-</td>
<td>3,198</td>
<td>7,905</td>
<td>72,140</td>
</tr>
<tr>
<td>German</td>
<td>Ansar</td>
<td>11/12/2009-</td>
<td>62</td>
<td>726</td>
<td>1,645</td>
</tr>
<tr>
<td>Russian</td>
<td>Kayli7rhat</td>
<td>12/13/2005-</td>
<td>5,634</td>
<td>6,144</td>
<td>558,042</td>
</tr>
</tbody>
</table>

These forums are appropriately selected for the purpose of getting the appropriate output from the significant input regarding the terrorism researchers. The selected input consists of the security and the military educators with the other experts who are responsible for the creation of the terrorism through the Dark Web. The analysis helps in depicting the selected languages which includes the major Jihadist Websites and thereby the English section is being included for the enhancement of the extractions regarding the web. Therefore the forums regarding the English language are constructed which helps in depicting the extremist and the more moderate groups are created for the purpose of decoding the language. The French, Russian and the German forums are created who seems to be using the mode of the language for the purpose of carrying out the communication process. An additional opportunity is being created which helps in the evaluation of the multilingual translations and thereby the appropriate evaluations of the languages are depicted for the purpose of depicting the mode of communication (Buscema and Tastle, 2013). The total number of messages that are depicted to be about 13M and near about 3M is added annually in the form of the postings to the forum which is indicated by the help of incremental spidering.

4.3 System Functionality

The system consists of four types of functions which are the single multiple forum browsing, multilingual translations, forum statistics analysis and the social network visualization. The description of the types is provided below which are as follows:

**Forum Browsing and Searching**

- Single forum browsing and searching: This search function provides the users to search the messages with the use of the titles with the other bodies by using the multiple keywords. The user chooses the Boolean algebra for the purpose of conducting the operation with the help of the keywords such as the AND or OR. The users can also easily express the searched items in English even when the forum consists of the Arabic Language. The search will return the matches for both the English and the Arabic translations and the terms can be easily chosen.
Multiple forums browsing and searching: In addition to the browsing, the particular forum consists of the portal that supports the multiple forum supporting and thereby the threads can be easily retrieved across all the forums. The detailed search can be easily carried out by searching the keywords with clicking row of the particular forum.

Social network visualization
The interface regarding the SNA function consists of the three parts which are the search panel, analysis panel and the visualization panel. The search panel helps in allowing the user to choose the related searches by the help of criteria which are the forum, keyword and the time period. The threads that meet the criteria are depicted as the related threads and are used for constructing the social network. Any of the forums that are listed in the portal can be selected by the help of the SNA. The keywords help the users to select the language and the topics with the help of comma for the purpose of separating the topic. The analysis panel allows the user to select the various metrics and also helps to set the parameters for the purpose of graph visualization. Each and every node of social network metrics is used for enclosing the number of the postings and various social network metrics (Wu and Li 2011). This result in social networking consists of the graphs that are filled with analysis. The virtualization is the process of displaying the graphs based on the settings depicted in the analysis panel with including the thickness of the link proportionate to the intensity of the interactions undertaken. Therefore, the node can be easily dragged to any position in the panel and thereby the connection between the nodes and the corresponding links can be easily highlighted at the tip of the mouse button.

4.4 Case Study for Identifying the Active participants in Dark Web Forum
A case study has been undertaken to show a scenario on how to apply the SNA element to detect active users on a specific subject of interest. The case study is on religious beliefs. An Islamic awakening case study has been selected. By searching with the following keywords, "Islam, Muslim, Shia and Sunni" a subject-based social network has been generated. From 01/01/2009 to 12/31/2009 has been selected as the undertaken time period. As illustrated in the following figure "Bint ul Islam," "I love Islam," and "Abuhanah" were the most active three participants among the numerous participants on the above-mentioned topics. It is recognized by calculating their numbers of messages on this religious topic. The researcher can demonstrate the SNA based on other diverse graph metrics like the "betweenness," along with "out degree."
By examining the “Link to user post” box, the user can observe the detailed messages. The messages about this particular religious subject add the following:

"Sister, since you are a new Muslim, Allah will test you as He Says in the Qur’an: ‘Do men think that they will be left alone on Saying, "We believe," and that they will not be tested?’ The researcher tested those before them as well as Allah will know those, who are true in comparison to the other, who are false. [Al-Qur’an 29:2-3] . You have to remain strong” by: I love Islam. “Between The Past and the Future Imam Ibn ul Qayyim al Jawziyyah al-Fawaid, pp 151-152 Al-Istiqaamah, No.2 Your life in the current moment is in among the previous as well as the future. As a result, It can be stated that what has preceded can be fixed by Tawbah (repentance), nadam, (regret) and istighfar (seeking Allah’s forgiveness”) by "Bint Ul Islam".

Figure 3: Active Participants in the Forum of Islamic Awakening
Chapter 5: Conclusion and Recommendations

5.1 Introduction
This chapter shows the result of the dissertation study means success or failure of the dissertation study. The final verdict of the dissertation study used to be obtained in this chapter. The researchers after completion of the data analysis and discussion reach to the final opinion about the issue of undertaken dissertation topic. In this particular dissertation study, the researcher has tried to identify the impact of the dark web on the modern society. The dark web technologies play important role in society as it improves the security levels, thus important data can be transferred in confidential way. However, the terrorist and militant groups use these same technologies by malicious way to spread hatred and violence in the society.

This project has adopted several tools and techniques, which support in the process of the data analysis and provide adequate knowledge and understanding about the importance of the dark web in modern society. Apart from this, it helps in recognize the impact of the dark web in the modern society and at the end of this chapter a list of recommendation has been formulated, which help in reducing the negative impact of the dark web technologies.

5.2 Linking with the objectives
In this specific part of the concluding chapter, we have tried to establish a link between the empirical experience of the researcher during the investigation process and the theoretical knowledge, which was obtained from the literature review chapter. The success or failure of the dissertation study is largely dependent on the alignment or linking of the theoretical knowledge and the practical experiences. The researcher has decided some research objectives in the very first introduction chapter of the dissertation study and this final chapter, a link has been established between the objectives and the practical experiences of the researcher.

Objective 1: To find out the activities that helped by the Dark Web in a positive way
According to the collected information, it is observed that the defense and armies mainly use the dark web technologies to protect the confidential information of national security. In these ways, helps in transiting secure information between the different branches and level of the national domestic and border security forces. Besides this, the dark web technologies help in transferring
particular information to the right person. Such as without the PNR (Passenger Number Record) number rail will not provide any status of the passenger list, position, or sitting arrangements.

Objective 2: To find out the activities that helped by the Dark Web in a negative way
The terrorism informatics field is growing continuously, and the investigators are very much interested in the counterterrorism research. It is also expected to introduce new methodologies, computational approaches and terrorism databases based on the complex interactions and rich content produced by the extremist and terrorists on the internet. It is a challenge for the researchers and investigators to collect and analyze the Dark Web information and data as because terrorists can hide easily their identities and also removes traces of their activities on the web. A huge flow of information on the web has made difficult for the investigators to obtain traces of the activities of terrorists. Advanced content analysis, domain knowledge, visualization, and web mining can be used for determining and evaluating information on the web. Information visualization is used to identify terrorist activities on the web.

Crawler to collect Dark Web Forum has been focused, and accessibility mechanism has been used to access the determinate forums within the success rate of over ninety percent. Language independent features have been used by the crawlers which include anchor text, level features, and URL tokens in order allow effective content collection in multiple languages. Specific software transversal wrappers and strategies are being used to support the incremental crawling. The system uses the incremental crawling techniques couples with improvement mechanism that respiders the collected pages.

Dark networks such as narcotics trafficking networks and terrorist networks impose significant impact on the society and economy. The evaluation of the topology of the dark network can reveal a wide range of hidden information of secret organizations as well as help to develop effective strategies to disrupt their activities. However, it is very much difficult to obtain information from dark networks. It has been found that the covert networks share many topological properties that common with other types of networks. An efficient flow of commands, goods, and information can be connected with small structures having high clustering coefficient and average small path length. The properties of small world structures allow the dark networks to attack the bridges related to
different communities rather than attacking the hubs. The human networks have high clustering coefficients that are not always presented in other experimental networks and difficult to regenerate by known network effects like small world and special attachment effects. Alternative mechanisms such as the Highly Optimized Tolerance (used for determining the robustness of the system) can govern the evolution of complex systems in uncertainty and high-risk environment.

5.3 Recommendation

The investigators and researchers are very much careful while determining and evaluating Dark networks as because Dark networks are hidden networks, and actual networks are commonly unknown. The links can be missing or hidden in the elicited networks. The hidden links can play vital roles in maintaining the functions and operations of the secret organizations. Therefore, extreme caution should be taken while disrupting these networks.

A hybrid approach can be developed that uses both linguistic features, and CMC (Computer Medicated Communication) system features to construct interaction patterns from the web discourse. The researchers are working on analyzing the role of the users in the web forums by network interactions produced by the HIC algorithm. The entire interactions all across the web should be monitored by the security agencies all over the world. Moreover, in the case of the security agencies observe any discrepancies, the incident must be disseminated in the web and must communicate with the entire important security agencies to short out the issue in time. All the developed nations and the important security agencies should work on this particular subject in a joint way and should make effective communications among them. In these ways, the impact of the dark web can be minimized from the society, and the violators can be identified more fast, which will minimize the negative utilizations of the dark web technologies.
References
Dordrecht: Springer.


### Appendix - A
Project Ethics Number: 2016D0460

<table>
<thead>
<tr>
<th>Name of applicant:</th>
<th>Abdulmuttaleb Abduljalil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor (if student project):</td>
<td>Tom Crick</td>
</tr>
<tr>
<td>School / Unit:</td>
<td>School of Management</td>
</tr>
<tr>
<td>Student number (if applicable):</td>
<td>St20063132</td>
</tr>
<tr>
<td>Programme enrolled on (if applicable):</td>
<td>BSc (Hons) Software Engineering</td>
</tr>
<tr>
<td>Project Title:</td>
<td>ANALYSIS OF DARK WEB</td>
</tr>
<tr>
<td>Expected start date of data collection:</td>
<td>14/11/2016</td>
</tr>
<tr>
<td>Approximate duration of data collection:</td>
<td>3 Months</td>
</tr>
<tr>
<td>Funding Body (if applicable):</td>
<td>N/A</td>
</tr>
<tr>
<td>Other researcher(s) working on the project:</td>
<td>N/A</td>
</tr>
<tr>
<td>Will the study involve NHS patients or staff?</td>
<td>No</td>
</tr>
<tr>
<td>Will the study involve human samples and/or human cell lines?</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Does your project fall entirely within one of the following categories:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper based, involving only documents in the public domain</td>
</tr>
<tr>
<td>Laboratory based, not involving human participants or human samples</td>
</tr>
<tr>
<td>Practice based not involving human participants (e.g. curatorial, practice audit)</td>
</tr>
<tr>
<td>Compulsory projects in professional practice (e.g. Initial Teacher Education)</td>
</tr>
<tr>
<td>A project for which external approval has been obtained (e.g., NHS)</td>
</tr>
</tbody>
</table>

If you have answered YES to any of these questions, expand on your answer in the non-technical summary. If you have answered NO to all of these questions, you must complete Part 2 of this form.

**In no more than 150 words, give a non-technical summary of the project**
The aim of the project is to have a fully understandable research and awareness about the topic. And at the end create a simple guide for unsuspicious people that don’t know what Deep Web is and to explain all the necessary information that we need to know before accessing the Tor network and how one can have safe and legal access without any risk.

DECLARATION:
I confirm that this project conforms with the Cardiff Met Research Governance Framework
I confirm that I will abide by the Cardiff Met requirements regarding confidentiality and anonymity when conducting this project.

STUDENTS: I confirm that I will not disclose any information about this project without the prior approval of my supervisor.

Signature of the applicant: Date: 21/11/2016

FOR STUDENT PROJECTS ONLY

Name of supervisor: Tom Crick

Signature of supervisor:

Research Ethics Committee use only

Decision reached: Project approved
Project approved in principle
Decision deferred
Project not approved
Project rejected

Project reference number: Click here to enter text.

Name: Click here to enter text. Date: Click here to enter a date.

Signature:

Details of any conditions upon which approval is dependant: Click here to enter text.