International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

Identification of Illegal Forum Activities inside the Dark Net

Narayan Bhatkande[1], Varsha Rotti [2], Cinderella Coutinho[3], Swanand Patil[4], Mahantesh Laddi[5]

Student, Department of Computer Engineering, S.G.Balekundri Institute of Technology, Nehru Nagar, Karnataka, India[1]

Student, Department of Computer Engineering, S.G.Balekundri Institute of Technology, Nehru Nagar, Karnataka, India[2]

Student, Department of Computer Engineering, S.G.Balekundri Institute of Technology, Nehru Nagar, Karnataka, India[3]

Student, Department of Computer Engineering, S.G.Balekundri Institute of Technology, Nehru Nagar, Karnataka, India[4]

Assistant Professor, Department of Computer Engineering, S.G.Balekundri Institute of Technology, Nehru Nagar, Karnataka, India[5]

ABSTRACT: Dark Net is commonly used independently to examine users' experiences while surfing on websites. This project explores the value of both methods when used separately through running three independent studies, on website. The freedom of the Deep Web offers a safe place where people can express themselves anonymously but they also can conduct illegal activities. The first phase involved live Dark Net data collection and analysis; the second phase involved controlled collection and analysis of Dark Net data and the final phase involved running and analyzing evaluation sessions. The results of all three sessions were then analyzed to determine the varying types of findings which are gained from each method. The conclusion of this project presents recommendations of how both Dark Net and website evaluation may be used together to overcome limitations of each of the methods and to strengthen insight into user experience. This includes guidelines proposing where both Dark Net and think aloud evaluation may be used throughout the User Centered Design process, to help usability practitioners to better understand user experience.

KEYWORDS*:* Website tracking, Real time data updates, Geolocation, Visitor log.

I. INTRODUCTION

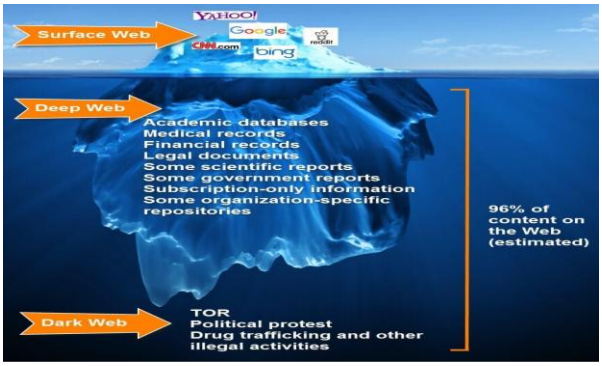

Fig 1: The deep web

The Silk Road's popularity has sparked many Dark Web marketplaces to develop. This explosive growth has created new platforms for criminal organizations to market illegal goods. Thus, the Dark Web has generated great interest from academics and governments seeking to reveal participants' identities in these highly lucrative, yet illegal, marketplaces. Traditional Web scraping methodologies and investigative techniques have proven to be inept at unmasking these

marketplace participants. This research provides an analytical framework for automating Dark Web scraping and analysis with free tools found on the World Wide Web. We successfully tested a Web crawler, built using AppleScript, using a case study marketplace to obtain account information for thousands of vendors and their respective marketplace listings. This project specifically explains why AppleScript has become the most effective and productive form of scraping marketplaces in the Dark Web. Our case study findings support the validity of our proposed theoretical method, which is important for researchers researching this growing trend and for investigators investigating illegal activity on the Dark Web.

Open-source intelligence provides value in decision-making about information security through knowledge of threats and malicious activities that could potentially affect business. Open-source information is popular using the internet; however, the average cybersecurity researcher uses the darknet less frequently. The complexities of using the darknet for open-source intelligence include the use of advanced methods for capturing, analyzing and analyzing. Darknet is an intelligence platform built in python which is open source. The main goal of this project is to collect open data from the deep web (aka dark web) to collect as much information as possible to create an interactive tree graph with the aid of data mining algorithms. The interactive tree graph module will show the relationships of the intelligence data gathered.

Cybercrime activities in the dark web can be seen as one of the most important problems for communities around the world. Cyber criminals may use the Internet to engage in criminal activities such as: trading and buying drugs, pedophilia, hiring hitmen, forgery, piracy and terrorism. In this project, we highlight the illegal activities that are happening on the Internet's dark side, particularly the dark web fora. We used the Onion Router to search the web content posted and discussed by users of dark web fora. This project also looks at the most common activities in the illicit topics that users chat about in the dark web fora.

We performed affect analysis on the dark web forums, which is useful for measuring the presence of illicit subjects such as drugs, violence, forgery and piracy. The results show that there are similarities in the subjects between the dark web forums.

Standard Quality Assurance (QA) is not sufficient in ensuring that all visitors to a website are able to complete their objectives through using an interface with good usability, comprising; efficiency, effectiveness and satisfaction (Add wise, 2001). There are therefore a number of further methods which may be used to evaluate the usability of an interactive device, including think aloud, web analytics, cognitive walkthrough, heuristic evaluation and many more. However, there is an understanding within the field that testing using only one technique is insufficient (Nielsen, 2002a). Rosenfeld & Moville (2002) have learned that to get the most value out of user evaluation, integration of multiple research methods is required.

The following application critically explores specifically Dark Net and thinks aloud evaluation methods. The aim of the study was to determine the effectiveness of each of these methods when used independently. Using findings from this evaluation I devised guidelines on how both Dark Net and think aloud may be best applied and possibly combined throughout the User Centered Design (UCD) process to gain a deeper understanding of user experience. Dark Nets are concerned with collecting, analyzing and interpreting web metrics. The focus is on improving web communication by mining and analyzing web usage data to discover interesting metrics and usage patterns.

The goal of Dark Net is to assist companies in improving the quality of their website – including the usability and performance of the site. Data analytics service can help businesses identify aspects of e-business which are working well through reports and then devise recommendations of how they may work better.

Poor website usability, results from designers' lack of understanding of how users surf for information. However, correct analysis of Dark Net data can present this type of information. Data can be tracked relating to: how long people stay on a website, which areas they visit, where they came from and where they went next. This data is generally presented in reports which may be used to examine the efficiency of regularly used aspects of a website, as well as reviewing rarely used facilities to understand why users may avoid such features. The usefulness of such web analytics reports in corporate decision making depends on the report viewer; web designers are interested in detailed reports, whereas strategic managers only require summaries which show the number of visitors to each page, adjusted to either explore fine details of certain aspects of a website or present broad information on site usage, to meet the requirements of the person using the details.There is some literature which details the process of collecting and analyzing Dark Net data, along with critical discussion of the method. Much of this literature details specific software which is used to collect and analyze Dark Net data.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

II. RELATED WORK

Abbasi and Chen (2007) presented affect analysis of Middle-Eastern extremist group forums comparing with American extremist group forums, which is useful for measuring the presence of violence and hate speech by extremist groups. The authors used lexicon to evaluate the level of hate speech and violence. The lexicon facilitates in- depth analysis of multilingual content. The proposed approach has been evaluated by applying it across sixteen Middle-Eastern and American extremist group forums. The analysis shows that the Middle-Eastern testbed forums have a higher degree of violence than the American extremist group forums. By using the testbed, the authors found that the American and Middle-Eastern extremists on deep web forums have the same level of hate-related affect [1]. The authors built a system for affect analysis across the deep web forum. Their techniques can present the affect intensity with the relationship analysis in order to provide a better analysis of multilingual discourse affect usage trends. These techniques can be used for further analysis of extremists' groups across the deep web [1].

Zhang et al. (2009) presented a web forum data integration framework for searching and analyzing international Jihadist forums, specifically, within a web-based knowledge portal. The web-based, multilingual portal and the deep web forums portal were developed based on the authors' methodology. These forums were selected from seven Jihadist forums. Many functions were applied by the integrated portal, including multilingual translation, social network visualization, forums statistical analysis, forum searching and forum browsing. The forum applied functions can support a user to understand and utilize the data easily and quickly [2].For setting up an anonymous virtual machine, the user needs to have a virtualization tool, for research purposes using Oracle's Virtual Box is recommended as it is open source and supports every type of OS extension. After downloading and installing Virtual Box, it's time to choose an anonymizing OS. Currently there is only one operating system which has surpassed all other operating systems in anonymity and security known as WHONIX Operating system [3]. Whonix is not actually an operating system but a virtual appliance which comprises of two appliances known as Whonix Gateway (manages the anonymous and relay-based network connectivity), and Whonix Workstation (in which the user will do his/her actual work and run TOR for browsing deep web).

In order to solve the problem of anonymous communication, the US Naval Research Institute Goldschlag, Syverson and Reed jointly proposed the onion routing technology, which has found impact on the anonymous communication network. The Second Onion Router is established. The second generation of onion routing anonymous communication system on the first generation of onion routing technology. The main purpose is to enable normal network communication on the Internet, while the identity information and network relationship between the two parties are not observed by third parties. The Tor network is an Internet-based distributed network [4]. Its network topology is contained in the Internet topology. The Tor system consists of multiple components, including the Onion Proxy (OP). Onion Router (OR), Directory Server and Application Server.

It was through a forum post on March 2011 that the first large dark web hosted market place was publicly announced. The Silk Road was presented as an anonymous online market, still under development in hope to receive feedback from the bitcoin community. While starting with drugs, the staff was aiming to offer weapons and other products that may be difficult to find on the surface web market places. With the combined efforts of FBI and Interpol, the site was seized on February 2013, ending with the arrest of Ross William Ulbricht, who was charged with engaging in a money laundering and narcotics trafficking conspiracy as well as computer hacking [5].

Agora is (or possibly was) a popular "dark web" marketplace. Agora does have characteristics that are similar to other black-market operations [6] and it's based on behavior that doesn't comply with an organizational set of rules. It is fairly obvious that entities that fail to comply with the set of rules do so in order to obtain some sort of benefit. Generally speaking, Agora was selling both products and services, with a minimal set of rules. At the time of our research the only items that couldn't be sold were body parts, and the only service that was forbidden to sell was assassination. It is possible that the very limited set of rules might have been some sort of market positioning statement - in practice the key areas of business within the market where drugs, fake ids and weapons (but only for a limited time). As for all black-market operations, operations on Agora were not taxed, neither directly nor indirectly, and Agora offered sellers the possibility for sellers to place products that could not be typically sold legally. The key aspects of Agora are largely similar to the ones of other illegal operations: protection of the identity for the members, exchange of money, illicit profits. The main source of a large amount of threat intelligence is the dark web forum, and the deep dark forum can better understand the motives and activities of illegal elements. A close understanding of these communities will greatly help maintain cyber security and enable cyber security practitioners to better understand their opponents. Through the visualization of the Dark Web Forum, valuable insights into the structure and culture of the forum can be obtained. Many aspects of the Dark Web Forum have been the subject of academic research, such as their

direct and anonymous communication as a person. Observe the communication methods of the forum, find that the means and tools of communication are public threads or private messages, and sometimes provide plug-ins for instant chat platforms such as Jabber and ICQ. In addition, entering a dark web forum often requires the user to register personally.

The Dark Web represents an additional tool of the trade for traditional criminals. In November 2014, police from the European Union and the United States seized hundreds of Tor." Onion" domains utilized by numerous drug marketplaces—including Silk Road 2.0. Criminals on the Dark Web today not only sell drugs, weapons, counterfeit currency and documents, but also use this anonymized network for other nefarious dealings. Thus, academics and investigative Road 2.0. A number of studies have attempted to quantify the growth and membership of these popular and successful online communities, and also to analyze their business models. One researcher, for example, examined the growth of 16 different marketplaces after the demise of Silk Road, and analyzed how vendor security practices have improved over time. In February 2015, the FBI seized a server on Tor, called Playpen, which was used to distribute child abuse images. The FBI successfully deployed tracking software, referred to as a "NIT", to unwitting client computers, visiting the Playpen server, and then successfully identified the real IP addresses for 215,000 users. There are also several examples of the Dark Web being used by terrorist organizations. The greatest challenge to analyzing the Dark Web, and investigating associated illegal activities, has been the lack of transparency and effective analytical tools because of encryption techniques and the anonymity of users. These layers of encryption have provided anonymity to both vendors and their customers. Scholars and law enforcement must often rely on old-fashioned investigative techniques, which are available to the public. This approach involves taking data from the Dark Web and attempting to correlate that data, for example usernames, to make a positive identification on the World Wide Web. Alternatively, an investigator may run the data captured against an internal database of suspects or known criminals. Dark Web marketplaces have thousands of product listings and user accounts, and therefore it is virtually impossible for manual investigative and analytical techniques to yield actionable data expeditiously. In fact, identifying marketplace owners, customers and products remains challenging. Therefore, the aim of this research is to provide an analytical framework for automated scraping and analysis that can be more accessible to scholars and practitioners.

Deep Web is a hidden web database (~95% of the total internet) and was initiated for anonymous communications with American personnel and embassy stationed abroad, over time as the idea about anonymity popularized, criminals adopted deep web as a hub for carrying out illegal activities like selling credit cards, weapons, drugs & where one could hire a hitman, hacker etc. But deep web also provides knowledge about government and the secrets that the general public should be aware of. Moreover, it has information about secret services worldwide, deep web has the largest library and collection of over a hundred thousand books. The way in which a user uses deep web is purely based on his/her conscience and ethics [10]. The internet is a big collection of data and how much information we provide them with lies within a user's hand. A major part of security is compromised when a user shares his/her details online on a public platform or social media so it's important for a user to limit the information being transmitted. This project covers every aspect about staying anonymous and secure identity of an internet user online, but general awareness and commonsense plays a much important role than any preventive measure and is the key role to be safe online.

In the current era of the internet of things, a Dark Web has been developing under the surface. The Dark Web usually relies on the combination of crypto currencies such as bitcoins and anonymized access as the foundations in creating a market place for dealing illegal drugs, weapons and other illegal contrabands. In recent years, the Dark Web has been in extreme scrutiny and investigations from legal authorities around the globe. The clamp down of these have come with mixed success. Whilst many of these websites have been successfully shut down; others resurface or re-migrate soon after. The most significant case is the Silk Road, which was shut down in October 2013 and then resurface as Silk Road 2.0 in a month later. Then, soon after the shutdown of Silk Road 2.0 in 2014, many of the vendors just simple migrate to other marketplaces such as Evolution and Agora. Although a number of prominent web sites in the Dark Web have been shut down thanks to intervention from police forces, the main characteristics of these Dark Web sites are still unknown and we didn't know enough of how these websites operate. The term Dark Web was first introduced in the year 2000s and has been in use both in the media and in academia since then. Most research done are focusing in identifying extremist views and identifying extremist groups. Only recently there interest has started to surge in others areas, including drug trafficking [11], looking at alias classification and authorship attribution.

In research on improvised explosive device webpages, Chen (2008) presented a classifying framework. Since extremists can spread information about improvised explosive devices (IEDs) on the Internet, these webpages interest the police and the intelligence agencies for national security reasons. Security organizations should be able to check and identify the suspicious web sites. The classification of IED's webpages should identify IED webpages with a low false positive rate. The authors used a Complex Feature Extractor, Extended Feature Representation, and Support

Vector Machine (SVM) learning algorithm. The web pages containing information about IEDs were collected from the deep web. The authors examined two classification models: the 1's classification model used the Extended Feature Set; the 2's classification model included a group of both the Extended Feature Set and Information Gain Heuristic. The accuracy of the approach was around 88%. The testbed of over two thousand and five hundred webpages has been formed using Focused Crawling Technique. The Information Gain Heuristic has been utilized in Feature Selection. The Accuracy result of IEDs webpages was around eighty eight percent by using Genre classification. The Genre classification has been implemented through the testbed on over two thousand and five hundred webpages, forty material pages with around two thousand and five hundred discussions have been identified by manual categorization using a domain expert [12].

III. PROBLEM STATEMENT

Whenever we start any tracking project for a new darknet or an existing application, the first thing we do is a complexity scan of the complete machine details of the client. We go through basic and advanced settings, check the data quality and consistency – we try to find the week spots and make suggestions on how they can be improved. Most of the time we encounter a number of recurring problems that appear in nearly every new account on which we work. In this application we are going to track some information of client.

IV. PROPOSED WORK

Following are the list of objectives of the application:

1. Real time data updates
2. Customizable dashboard
3. All websites dashboard
4. Goal Conversion Tracking
5. Event Tracking
6. Geolocation
7. Pages Transitions
8. Track different user interaction, Automatic tracking of file downloads, clicks on external website links
9. Accurately measure the time spent by visitors on your website.
10. Visitor Log
11. No Websites and data limit

V. FEASIBILITY ANALYSIS

A feasibility study is conducted to select the best system that meets performance requirement. This entails an identification description, an evaluation of candidate system and the selection of best system for the job. The system required performance is defined by a statement of constraints, the identification of specific system objective and a description of outputs.

The key considerations in feasibility analysis are:

1. Economic Feasibility
2. Technical Feasibility
3. Operational Feasibility

5.1 Economic Feasibility

It looks at the financial aspect of the project. It determines whether the management has enough resources and budget to invest in the proposed system and the estimated time for the recovery of cost incurred. It also determines whether it

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

is worthwhile to invest the money in the proposed project. Economic feasibility is determined by the means of cost benefit analysis. The proposed system is economically feasible because the cost involved in purchasing the hardware and the software are within approachable. The personal cost like salaries of employees hired are also nominal, because working in this system need not required a highly qualified professional. The operating-environment costs are marginal. The less time involved also helped in its economic feasibility.

5.2 Technical Feasibility

It is a measure of the practically of specific technical solution and the availability of technical resources and expertise.

- The proposed system uses ASP.NET WITH C#.NET as front-end and SQL server 2008 as back-end tool.
- Oracle is a popular tool used to design and develop database objects such as table views, indexes.
- The above tools are readily available, easy to work with and widely used for developing commercial application.

5.3 Operational Feasibility

The system will be used if it is developed well then be resistance for users that undetermined:

- No major training and new skills are required as it is based on SPIRAL model.
- It will help in the time saving and fast processing and dispersal of user request and application.
- New product will provide all the benefits of present system with better performance.
- User involvement in the building of present system is sought to keep in mind the user specific requirement and needs.
- User will have control over their own information. Important information such as pay slip can be generated at the click of a button.
- Faster and systematic processing of user application approval, allocation of IDs payments, etc. used had greater chances of error due to wrong information entered mistake.

VI. METHODOLOGY

A data-flow diagram (DFD) is a graphical representation of the "flow" of data through an information system. DFDs can also be used for the visualization of data processing (structured design).
On a DFD, data items flow from an external data source or an internal data store to an internal data store or an external data sink, via an internal process.

A DFD provides no information about the timing or ordering of processes, or about whether processes will operate in sequence or in parallel. It is therefore quite different from a flowchart, which shows the flow of control through an algorithm, allowing a reader to determine what operations will be performed, in what order, and under what circumstances, but not what kinds of data will be input to and output from the system, nor where the data will come from and go to, nor where the data will be stored (all of which are shown on a DFD).

DFDs help system designers and others during initial analysis stages visualize a current system or one that may be necessary to meet new requirements. Systems analysts prefer working with DFDs, particularly when they require a clear understanding of the boundary between existing systems and postulated systems. DFDs represent the following:

1. External devices sending and receiving data
2. Processes that change that data
3. Data flows themselves
4. Data storage locations

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

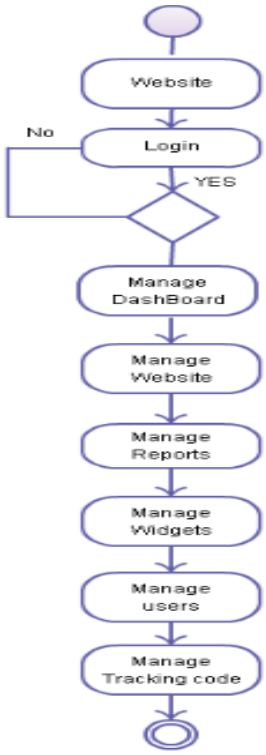

Fig 6.1: Dataflow diagram of the project

VII. RESULTS AND ANALYSIS

Darknet Tracker is an application that does mainly two things:
• Collect and store analytics data.
• Provide reports of the stored data.

To achieve that result, several parts of Darknet come into play:
• The JavaScript tracker is served by Darknet Application via HTTP so that websites can include it in their pages
• The tracker collects data on the web page it's included in and sends it to Darknet by calling the HTTP tracking.
• The archiving task runs and pre-processes data (either on the fly or via a cron task)
• Data is exposed in reports, which are accessible through the web interface or the HTTP reporting.

The plug-in architecture
Piwik's codebase is composed of:
• Matomo Core, which provides the base of the application along with extension points.
• Plugins, which use the extension points to add behavior and content to the application.

VIII. CONCLUSION

The conclusion of this project presents recommendations of how both Dark Net and website evaluation may be used together to overcome limitations of each of the methods and to strengthen insight into user experience. This includes guidelines proposing where both Dark Net and think aloud evaluation may be used throughout the User Centered Design process, to help usability practitioners to better understand user experience.

REFERENCES

[1] Walsh, D., A Beginner's Guide to Exploring the Darknet, https://turbofuture.com/internet/A-Beginners-Guide-to-Exploring-theDarknet, May 2018
[2] Christos Iliou, George Kalpakis, Theodora Tsikrika, Stefanos Vrochidis, Ioannis Kompatsiaris, Hybrid Focused Crawling for Homemade Explosives Discovery on Surface and Dark Web.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 6, June 2020 ||

[3] Whonix Wiki Contributors, "Comparisons of Whonix with other OS," 3 July 2017. [Online]. Available: https://www.whonix.org/wiki/Comparison_with_Others [Accessed 3 July 2017].
[4] Tong Yuen Yum, Explore the Dark Web. The University of Hong Kong, 2018.
[5] A. Greenberg, "End Of The Silk Road: FBI Says It's Busted The Web's Biggest Anonymous Drug Black Market," 2 10 2013. [Online]. Available: http://www.forbes.com/sites/andygreenberg/2013/10/02/end-of-the-silk-road-fbi-buststhe-webs- biggestanonymous-drug-black-market. [Accessed 16 6 2016].
[6] E. L. Feige, "Reflections on the Meaning and Measurement of Unobserved Economies: What Do We Really Know About the 'Shadow Economy'," Journal of Tax Administration, vol. 2, no. 6, 2016.
[7] Buxton, J., Bingham, T., the Rise and Challenge of Dark Net Drug Markets, January 2015.
[8] Al Nabki, M. W., Fidalgo, E., Alegre, E., and de Paz, I., "Classifying Illegal Activities on Tor Network Based on Web Textual Contents", Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 1, pp. 35–43, Valencia, Spain, April, 2017.
[9] Walsh, D., A Beginner's Guide to Exploring the Darknet, https://turbofuture.com/internet/A-Beginners-Guide-to-Exploring-theDarknet, May 2018
[10] K.Y.A, Bargh and J. McKenna, "Plan 9 from cyberspace: The implications of the internet for personality and social psychology," Personality and Social Psychology review, pp. 57-75, 4 January 2000.
[11] M. Splitters, F. Klaver, G. Koot and M. Van Staalduinen, "Authorship Analysis on Dark Marketplace Forums," in roceeding of Intelligence and Security Informatics Conference (EISIC), Manchester, 2015.
[12] Hsinchun Chen, IEDs in the Dark Web: Genre Classification of Improvised Explosive Device Web Pages, IEEE, 2008.