Inside of deep web anonymous technology: a short review

H Himawan\textsuperscript{1*}, A Susano\textsuperscript{2}, P Irfansyah\textsuperscript{2}, I Mutia\textsuperscript{2} and P Astuti\textsuperscript{2}

\textsuperscript{1}STMIK Raharja, Tangerang, Indonesia
\textsuperscript{2}Universitas Indraprasta PGRI, Jakarta, Indonesia

\textsuperscript{*}himawanawan10@gmail.com

Abstract. The term "deep web" is a truly interesting point to be the discussion, specifically for researchers or observers of technology in the field of website applications. This is because the world of "underground" websites allows a wider range of website pages than "conventional" website pages such as Facebook, YouTube or Twitter that can be quickly searched or found using various search engines available today. From the 3 levels or layers especially the web surface, deep web and dark web (dark net), this article tries to review the pages of the website in the second layer, especially deep web, where all the web pages that are in the deep web layer cannot accessed using a "traditional" or "conventional" method that simply uses a standard browser or searches using an existing search engine. Furthermore, we also try to explain how the mechanism or mechanism that exists in the pages of the website that goes into the deep web area and learn how the internet users can access the pages of the website.

1. Introduction
The world of internet networks seems to have no limits, this can be understood clearly with the number of expansions of new website pages every time from various parts in this world. The revised data from the internetlivestats.com website page indicates that the number of web pages that can be accessed by internet users has exceeded more than 1.8 billion website pages, this shows the website growth rate has increased highly substantially when seen from the page project the web began first initiated in 1991 by Tim Berners-Lee.

The fascinating information about the number of pages on the website is that only "appears" one part of the website pages that can be accessed online by internet users using "conventional" access methods. In general, when an internet user needs to visit a specific website page, the general thing that will be done is to open the browser then type the website domain name or URL page (Unified Resource Locator) that you choose to reach. The second thing that is more generally made is to open a web page through the help of a search engine such as Google or Bing. But both ways have "weaknesses", where the lack is that these methods "purely" can be applied for website pages that are "recognized" or more precisely indexed in search engine pages [1, 2]. The website pages that are regularly taken by internet users are website pages that are on the surface layer of the internet network, where the pages of this website are also known by various other terms namely surface web, open web, visible web or indexable web [3].

Deep web is web pages that are on the internet network but cannot be accessed quickly by internet users and also by search engines, why is that? As a simple explanation to express the deep web is to use examples of fishermen catching fish in the ocean. If a fisherman wants to catch fish that is
swimming close to the surface of the water, it is enough to use simple equipment and also a simple technique that is to spread the nets and the fish can be captured well. But what if the fish you need to get are not on the surface of the water? Say the fish swim in depths of 100 meters or more? This suggests that the use of ordinary nets only takes time and effort because of the limitations of the tool. Besides that, another factor is if the fish swim in the depths of the sea, then how can the fishermen know or determine the location of the fish? Because the sea has a very large area, therefore ordinary fishing equipment cannot be used. Illustration that can be used to describe the area layer of the deep web, namely "fish" illustrated as web pages. Whereas "fish" that are on the surface are only fish that have small to medium size, then "big fish" are in the depths of the ocean that cannot be reached by ordinary equipment and need to use more sophisticated equipment or modern to be able to reach it. Therefore, website pages that fall into the deep web category can "hide" or not be seen because they are not located on the surface layer of the area. If in the illustration big fish swim into the ocean to hide their existence, then how close you "swim" the website pages into the deep web category so that they can hide their presence from internet visitors and also not detected by various traditional search engines which exists?

We intend to point out how the internet network works in the deep layer of the web which is a hidden page so that it cannot be detected or indexed on the search engine page. In addition, this paper will discuss the functions of the TOR browser (The Onion Router) which is a special open source browser that can be used to access web pages that are included in the category of deep web pages.

2. Deep Web

Deep web has various unique characteristics when connected to ordinary website pages that are usually accessed by internet users. The attributes in search are the contents of dynamic web pages, content that is not associated to one another (did not have back links), contextual web pages and web pages with limited access [4]. In addition, the deep web pages also have other common features such as web pages that are given a password so that it makes it difficult for the search engine to be able to "understand" the contents in it [5]. Deep Web is pulling out of an internet network that is not indexed by search engine pages and is also not connected (linked) to website pages that are in the surface layer. Deep Web includes the characteristics of using special addresses that do not need the standard system Top Level Domains [6].

In practice, several deep web pages are intended to provide or sell illegal items such as weapons and narcotics. All web pages that are in the deep web layer have special domain extensions, namely .onion. Where this special extension indicates that the website page is part of the Deep web. From the operation of the domain, it can be observed clearly the difference when related to ordinary web pages that apply common domains like .com, .org, .edu, .net and others. Deep Web development itself set up in 2002, when the American naval research laboratory developed the TOR internet network which was then continued in 2003 with the development of peer-to-peer (p2p) computer networks.

The number of regular website pages that have been registered online has come to higher than 1.8 billion, then this number is nothing connected to the number of web pages in the Deep Web category or layer. Kristin Finklea (2017) in his report entitled "Dark Web" states that at that time the size of the Deep Web page was 4000 - 5000 times greater than the number of pages on the surface. If the data taken in 2017 is so large, then what about the size in 2018 now? On the other hand, the growth in the number of online websites continues to understand a significant increase every day, this must also show on web pages that are "underground", especially because of their presence which is difficult to detect [1].

A study organized by Brett Hawkins (2016) reports cases of activities on the Deep Web page both legal and illegal, in the writing of the paper also mentioned several examples of Deep Web pages that can be accessed by its users to find various kinds of information. Examples are "duplicates" from Wikipedia (web surface) which are Hidden Wiki that is connected to various services available in the Deep Web layer, where the Hidden Wiki page includes a .onion domain extension and can simply be reached using the TOR browser [5]. Furthermore, Vincenzo Ciancaglini et al (2013) investigated the
operation of networks managed by access the Deep Web page along with the use of statistical data that displays Deep Web page search data and types of services that are often applied in the Deep Web page. [7]. Interesting thing that is also a concern of researchers in writing the paper Vincenzo Ciancaglini et al (2013) is related to various bad things that can be identified on various pages of the Deep Web. Examples are Malware Trading, Crypto Locker (a type of ransomware), illegal drugs (narcotics), money laundering services using Bitcoin, stolen account sales (credit card data, PayPal accounts), sale of fake identities, confidential information related to state documents and others- other. Furthermore Emin Caliskan et al (2015) which details the technical aspects of the use of the TOR browser and all the technology contained therein [8].

3. TOR (The Union Router) Browser
TOR browser is a software or browser application that can be applied to access pages that fall into the category of deep web and dark net. Then what causes the TOR browser special and different from other browser software that is widely used by internet users like Mozilla Firefox, Google Chrome, Opera and others? The initial point that performs the TOR browser unique is that all services that run in the TOR browser network will be as "hidden services". All information sent through the TOR network will be confidential because communication carried out by both the sender and recipient of the information will be encrypted (point-to-point encryption) [6]. Then the protection or protection provided in the TOR network is also layered, making it difficult to track or detect for all activities carried out by its users. In addition to this, the TOR browser has a feature that does not store history or cookies on every web page that is accessed or visited by its users [9], this also makes the search engine difficult to do a "crawling" process or indexes web pages that are accessed using the TOR browser, because there is no "trace" left when the website is accessed or opened for the first time.

The development of the TOR browser itself as written earlier in the section in the deep web subchapter in 2002. Where the TOR network project produces a TOR browser that is open source and also a free network, where users can freely browse or search for various kinds of information there is on the internet network without worrying about being watched or supervised for all activities carried out by other internet users [9]. However, the TOR browser also does not have flaws or weaknesses. Because of its anonymous nature and privacy, forbidden websites or censored by internet service providers can be accessed easily without any restrictions. Thus, users who have limited knowledge about how to work from the TOR browser will also have a level of risk or be vulnerable to various existing digital attacks such as malware, trojans, viruses and so on.

4. How Deep Web Work?
Deep Web is a web page that is "unusual", therefore to be able to access it, internet users need to need a little knowledge of the information in the Deep Web. If in general internet users will open a search engine page then type in keywords or keywords and then the search engine will display search results that are relevant to the given keywords, so this is not the case with Deep Web pages.

![Figure 1. How to Search Website Pages Using Search Engines](image-url)
The workings of ordinary web pages that use the standard http protocol by using a search engine to find web pages based on the query input keyword given by the user (Figure 1). But this is different from deep web (Figure 2), where it works from the deep web page, that is (1) the user opens the TOR browser and then input the keyword search query, (2) the search keyword query will be encrypted when sent to the TOR Nodes (router) run by the contributor from each owner of the Deep Web page, (c) if the search results are not found in a nodes, then the search will continue on the next nodes that match the search results based on the query entered, (4) Deep page The Web will display Deep Web links based on search results and will direct users to the Deep Web page, and (5) all Deep Web pages displayed on the TOR browser page do not have a clear IP address because they are hidden on servers that also have an IP address anonymous (location unknown).

5. Conclusions
The existence of Deep Web shows that actually the "world" of the website that has been seen only a small part of the universe of the digital world. In addition, the Deep Web page is an internet network revolution that shows the existence of a technology that can avoid censorship from government agencies or other parties. Coupled with the existence of digital currencies (cryptocurrency) such as Bitcoins that can be used to make transactions that are difficult to detect or trace.

References
[1] Finklea K 2017 Dark Web. In: Congressional Research Service, (Washington
[2] Yadav A 2014 Insights of Deep Web : What Google can’t search International Journal of Engineering Sciences & Research Technology 3 796-9
[3] Agrawal K, & Agrawal Smita 2013 Deep Web Crawler : A Review International Journal of Innovative Research in Computer Science & Technology (IJIRCST) 1 12-5
[4] Chertoff M and Simon T 2015 The impact of the Dark Web on Internet governance and cyber security
[5] Hawkins B 2016 Under The Ocean of the Internet - The Deep Web SANS Institute InfoSec Reading Room May 1-19
[6] Ciancaglini V, Balduzzi M, Mc Ardle R and Rösler M 2015 Below the surface: Exploring the deep web Trend Micro 1-48
[7] Ciancaglini V, Balduzzi M, Goncharov M and Mc Ardle R 2013 Deepweb and Cybercrime Trend Micro Report 9
[8] Çalışkan E, Minárik T and Osaka A-M 2015 Technical and legal overview of the tor anonymity network NATO Cooperative Cyber Defence Centre of Excellence. Available: https://ccdcoe.org/sites/default/files/multimedia/pdf/TOR_Anonymity_Network.pdf (January 4, 2016)
[9] Naseem I, Kashyap A K and Mandloi D Exploring Anonymous Depths of Invisible Web and the Digi-Underworld