Application of Cryptography and Graph in Onion Routing

Naufal Alexander Suryasumirat 13519135
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13519135@std.stei.itb.ac.id

***Abstract*—In an age where humans have been more connected than ever before, it is hard to retain anonymity, especially online when browsing the web. Whether it be the surface web, the deep web, or the dark web, personal data and privacy is something individual people must maintain carefully and keep from prying eyes. It is important to protect your privacy because people might use your data for malicious reasons, yet most web browsers lack this awareness in keeping your data private and secure. This is where onion routing might help in keeping you anonymous when browsing the internet.**

***Keywords*—Onion Routing, Privacy, Anonymity, Graph, Number Theory.**

I. Introduction

Browsing the internet is an act of looking through the internet or as more commonly known as surfing the internet for information, entertainment, education, and much more. To browse the internet, one needs a software application to get access of information on the World Wide Web known as a web browser. When accessing the internet using a normal web browser, most websites now require you to send an HTTPs request to the server, which means that on some level, the message you send using HTTPs requests is encrypted. Though, for some cases to some people, this encryption method might not be enough. Using this method of encryption when browsing the internet is still susceptible to interception when using public internet and many other cases.

The internet as we know it, also called as the surface web, though seems infinite with no end in sight, is actually only the tip of the iceberg of the whole that is the internet. Surface web only contributes about 5% to the World Wide Web, whereas the rest are what we know as deep web and dark web. Both terms are different and each one describes different parts of the World Wide Web and they are not one and the same.

The deep web is a part of the World Wide Web that are not indexed by the common search engines such as Google, Yahoo, Bing, and more. The contents of deep web are behind the surface web and requires forms to fill to access such as username and passwords. For example, the contents of Facebook are mostly the deep web because to access the different pages of Facebook would require someone to fill out their username and passwords, these are called the deep web. These non-indexed contents of the internet can also include a live private server that is not published through normal means so the common search engines would not be able to index those pages. Search engines such as Google would not have the means to index these pages because they are private with no means to access them without direct access to the server, unless they publish their websites or put the link on another indexed page, most search engines would not be able to index them.

On the other hand, the dark web is a small part of the World Wide Web that requires specific software or configurations to access. This part of the web requires web browsers that implements onion routing method that anonymizes the client when surfing the web. The web browsers are, though not limited to, Tor, Freenet, and more. This method of surfing the web is not only limited to accessing the dark web, but can also be used to surf the internet in general. Using onion routing benefits the client because it anonymizes them and makes them safer when browsing the internet.

The dark web and onion routing are often associated with illegal activities and most people relate users of the dark web as criminals. Anonymity invites those kinds of activities, but the dark web was not created for that reason. It is up to the people who use it whether they use it for good or bad reasons. The dark web is only a tool for people to use, a powerful tool to be exact, but in the end, the decisions are still put upon those who use the tool.

Onion routing is an internet browsing method that allows for anonymous communication on a computer network. As the name suggests, an onion network is built to encapsulate a message over multiple layers of encryption to keep the message safe from unwanted eyes, similar to how an onion is encapsulated with multiple layers.

Onion routing works by encrypting the message from the client's computer with multiple layers of encryption before sending the message through multiple intermediary networks. These intermediary computers then later pass on the message while simultaneously decrypting the message once. Analogous to peeling an onion layer by layer. By the end of the route, the destination then decrypts the last layer of the message and follows the instruction of the message before sending it back to the client using the same method, except instead of decrypting the message, each intermediary computers' task is now to encrypt the result of the message and pass it on to the client.

II. Number Theory

A. Integers

Integers are numbers that contains no decimals. Members of integer includes -3, -2, 1, 0, 1, 2, 3, and many more. There are many properties of an integer, including its divisibility. The divisibility of an integer is dependent on the integer itself. There exist integers that cannot be divided by another integer without resulting in real numbers, except if it is divided by itself or by one. This kind of integer is what we call prime numbers.

B. Properties of Integers

An integer has many properties, one of the properties of an integer to be discussed in number theory is its divisibility properties. Suppose a and b are integers, with a being a non-zero integer. If *a* divides *b* with no remainder left, there exists an integer *c* resulting in $b = ac$. Using formal notation: $a \mid b$ if and only if $b = ac$, with c being an integer, and a is a non-zero integer.

C. Euclidean Theorem

Euclidean theorem is a theorem contained within number theory. Suppose *m* and *n* are integers, with n being a non-zero positive integer, there exists a quotient *q* with the remainder *r*.

$$m = nq + r$$

D. Euclidean Algorithm

Before going deep into discussing the Euclidean algorithm, there is a term known as Greatest Common Divisor of two integer that are not zero. Greatest Common Divisor or GCD for short is as the name suggest the largest integer that divides two non-zero integers. The objective of the Euclidean algorithm is to find the GCD of two numbers. The Euclidean algorithm can also be expanded to find the GCD of more than two numbers, while still following the same steps of the algorithm. Suppose there exists two positive integers *m* and *n*, the steps of the algorithm to find the GCD of the two numbers are as follows:

$$r_0 = r_1q_1 + r_2 \quad 0 <= r_2 <= r_1,$$
$$r_1 = r_2q_2 + r_3 \quad 0 <= r_3 <= r_2,$$
$$\dots$$
$$r_{n-2} = r_{n-1}\, q_{n-1} + r_n \; 0 <= r_n <= r_{n-1},$$
$$r_{n-1} = -1\, r_n q_n + 0$$

When the algorithm has reached zero, then the Greatest Common Divisor of the two integers have been found, the GCD of found by the algorithm is the latest non-zero remainder from the algorithm. If the latest non-zero remainder from the algorithm results in one, then the two numbers calculated by the algorithm are called as relative primes of each other. A prime number is an integer that can only be divided by one and itself, whereas relative primes are two or more numbers that cannot divide each other that results with a remainder of zero and does not have a common factor apart from one. Relative primes are used by many other theorems such as Fermat's Little Theorem, and many more.

E. Modular Arithmetic

Modular arithmetic is a subject in Discrete Mathematics that focuses on the remainder when two whole numbers or integers are divided, often visualized as a clock. This analogy is perfect for setting up an example of how modular arithmetic works, for example two hours after 11 o'clock is 1 o'clock which is perfect for describing eleven added by two and taking the remainder when it is divided by twelve, or the maximum number of a clock, which results in one. That result is what we call thirteen modulo twelve, and it resulting in one. This is also used in congruency of two numbers. Two non-zero whole numbers are congruent of each other modulo *X* if and only if the two numbers' remainder when divided by *X* is the same.

F. Chinese Remainder Theorem

The objective of the Chinese Remainder is to find the congruency of a system of linear congruence. For this theorem and algorithm to work each of the modulo needs to be relative primes of each other or as most commonly known co-primes of each other.

$$x \equiv a_1 \pmod{m_1}$$
$$x \equiv a_2 \pmod{m_2}$$
$$\dots$$
$$x \equiv a_n \pmod{m_n}$$

This linear system of congruency has a unique solution in a modulus of the multiplication of each of its equations. For example, the illustrated equation above has a solution in modulus of

$$m = m_1 \cdot m_2 \cdot \dots \cdot m_n$$

As have been stated above for this to get the right solution is to have the modulo of each equation to be relative primes of each other, for example two, four, and six wouldn't work, as they each have a common factor of two.

III. Cryptography in Onion Routing

Cryptography is the theory of preserving information from unwanted attention through algorithms and encryption. In general, it secures the information by masking it or changing the information which can later be reverted back to the original information. There are many methods of cryptography, each to its own benefits.

Onion routing uses its own encryption method called onion encryption. It is a symmetric onion encryption scheme which is a set of four algorithms, the encryption scheme is as follows:

$$OE = (G, E, D_1, D_2)$$

The scheme describes each of its algorithms, G represents the abstraction or visualization of each nodes in the onion network or cloud, much like an implemented graph. E represents the connection used by the directory server to relay the message along the graph much like an edge of a graph. Whereas D_1 and D_2 represents the processing of incoming message and outgoing

message, in its encrypted state or decrypted state. Onion routing uses symmetric encryption which is an encryption method where a key is used to both encrypt and decrypt the information. In its implementation in onion routing, there must be a minimum of three keys used to both encrypt and decrypt the sent message. The client has all of the keys, whereas each intermediary node uses parts of the keys to decrypt and encrypt the message. Cryptography uses number theory to implement its algorithms to encrypt and decrypt the information.

IV. GRAPH

Graph is a subject of discrete mathematics which are structures that are made up of discrete nodes that are mostly known as vertices and the relation between those vertices, which can be called as a connection between the nodes or vertices, known as edges. The edges are used to describe the connection between the vertices or nodes which can describe any relation between discrete objects whether in theory or in its applications.

A normal graph with no pointing arrows used as its edges is called as an undirected graph to differentiate between ones using arrows as its edges, known as directed graphs. An undirected graph has a formal notation of $G = (V, E)$.

The V symbolizes the set of vertices that are not empty, whereas the E is used to describe the set of edges or the connection between the graph's vertices, and the set of edges of a graph might be empty, which may be called as an unconnected null graph. Graphs has many types, as has been mentioned before there are directed graphs and undirected graphs, both of those graphs also have types that differentiates the type of graph is being shown.

Besides directed graphs and undirected graphs, there are also mixed graphs which are the type of graph that is a combination between an undirected graph and a directed graph. There is no limitation between the number of directed and undirected edges, but it has to contain the two types of edges to be categorized as a mixed graph. After that, there is also weighted graph. A weighted graph might be a directed or undirected graph that contains a number for each edges of the graph that represents its weight.

A. Graph Types

Graph has many types for describing different relations between each of its nodes or vertices. There are different types of graphs based on different categorization. Those categorizations are whether if the graph has loops, rings, multiple edges, the type of edge it has, and more.

The types of graph categorized by its number of edges or the number of loops or rings it has is as follows:

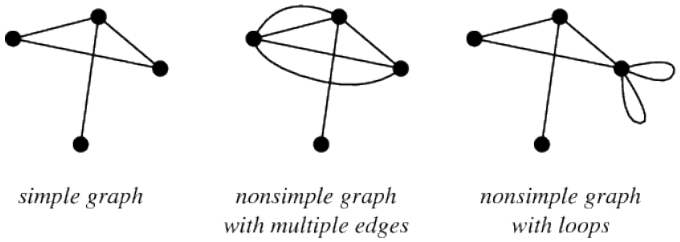

Figure 1. Example of types of graphs based on whether it has multiple edges or loops.
Source: https://mathworld.wolfram.com/SimpleGraph.html

1. Simple Graph
   Simple graphs are graphs without multiple edges connecting its vertices and graphs without containing any loop. A loop is an edge that connects a vertex to itself, whereas multiple edges means that a vertex does not have multiple edges connecting to the same other vertex.
2. Non-simple graph
   A non-simple graph is a graph that either has one or multiple loops or multiple edges connecting its vertices. A non-simple graph with multiple edges is called a multigraph, whereas a graph that contains a loop or a ring is called a pseudograph.

From the leftmost of Figure 1, the types of graphs are simple graph, multigraph, and pseudograph. The simple graph is indicated by no multiple edges and no loops, the multigraph is indicated by multiple parallel edges connecting two vertices, whereas the last is a pseudograph indicated by the loops on the rightmost vertex of the graph.

The types of graph categorized by its edges' orientation are undirected graphs and directed graphs. A combination between the two that is directed and undirected graphs is called a mixed graph which contains edges that are directed and also edges that are undirected. The types of graphs are as follows:

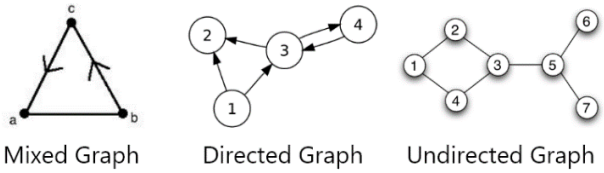

Figure 2. Example of types of graphs based on the orientation of its edges.
Source: https://www.researchgate.net/figure/An-undirected-graph-with-7-nodes-and-7-edges_fig3_265428782, https://upload.wikimedia.org/wikipedia/commons/5/51/Directed_graph.svg, https://en.wikipedia.org/wiki/Mixed_graph#/media/File:Mixed_Graph_Example.jpg

1. Directed Graph
   This type of graph is also called a digraph is a graph with edges that has orientation to indicate where it is pointing usually using arrows. This type of graph has a different type of degree, which is degree in and degree out for each of its vertices.

2. Undirected Graph
   Undirected graph is a graph with edges that is unoriented, which means that a connection of vertex a and b is the same as the connection of vertex b and a.
3. Mixed Graph
   A mixed graph as shown on Figure 2 is a graph that has oriented and unoriented edges. For the example given in Figure 2, the edge connecting vertices a and b is unoriented, whereas the edge connecting vertices b and c is oriented.

V. Orion Routing

As has been mentioned before, onion routing is a method of communication or internet browsing method that is anonymous, and differs from the common method of browsing the internet. There are many implementations of onion routing, but most of its implementations apply the same principle. That is to encapsulate the message sent by the client in multiple layers of encryption that only the client has all of the keys to decrypt while the message is sent through multiple computers that have parts of the client's keys to decrypt the message while passing on the message to another router so that an attacker wouldn't know where to begin to intercept the message. This is an oversimplification of how onion routing works, yet it is needed to paint the picture for the next part of the paper.

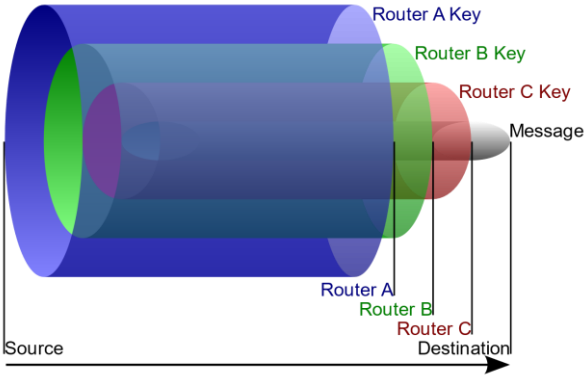

Figure 3. Structure of sent message using onion routing method
Source: https://upload.wikimedia.org/wikipedia/commons/e/e1/Onion_diagram.svg

From Figure 3, it can be seen that a message sent using onion routing method is much like an onion with multiple layers, hence, the name inherited for this type of anonymous communication is onion routing. The example given by Figure 3 uses 3 intermediary routers before reaching its destination. In its implementation there can be much more intermediary routers between the client and the destination, but the minimum number of routers that should be used for an onion routing method to be efficient and effective is three. Any more than three would be more secure although would cost the user's experience in getting the message back as quickly as possible.

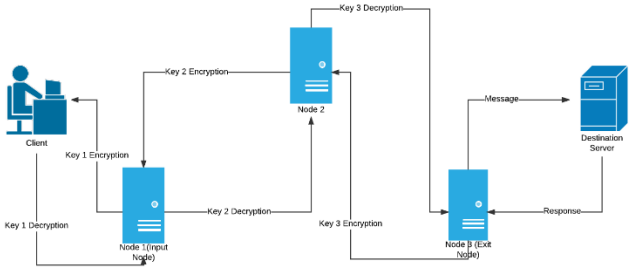

Figure 4. Example of an onion network
Source: https://media.geeksforgeeks.org/wp-content/uploads/Onion-Routing-Page-1.png

Figure 4 explains how a message would be sent through the onion network through intermediary nodes that can be represented as a graph as shown above. A graph of an onion network would be a directed graph from the client through the intermediary nodes, specifically it would be a directed weighted graph representing the minimum distance between each node for the message to be quickly sent while maintaining the quality of security of its encryption.

To determine the most efficient route through the minimum amounts of intermediary computers while still maintaining stable connection and the best quality of security would use the application of

VI. Graph and Encryption in Onion Routing

Graph has many implementations, in onion routing it is used to visualize, and describe each connection between intermediary nodes, the client, and the destination. Each of the connection is represented as an edge in the graph. Although, each connection is pre-determined between each node, in actuality, if someone were to intercept the connection one wouldn't have a clue to where the nodes are connected to and what message it is trying to send through the connection.

The connection would be more like a cloud with no clue of how the nodes are connected. The reason for this is, while the route from the client to the destination is finite, the connection between the intermediary nodes is too many to count. This cloud of connection is beneficial to the user as it obscures the connection between each node, to make it more secure to attacks as it would be near impossible to determine the route of the message, let alone the connection between each intermediary node.

Although this can be beneficial, this routing method still needs to be efficient, as in it should be as quick as possible to send the message and get a respond from the destination from the point of view of the client while still maintaining the security of the message that is sent. The only downside to this method of routing is that if an attacker has control over the entry node and the exit node, it would be easy to correlate what message the client is trying to send. To generalize, if you were to try to access Google through onion routing, the message wouldn't make it obvious that you were trying to access google. If someone were to intercept the connection while the message is being sent, the message would be gibberish with no way of decrypting it, because each of the intermediary node only has one of the

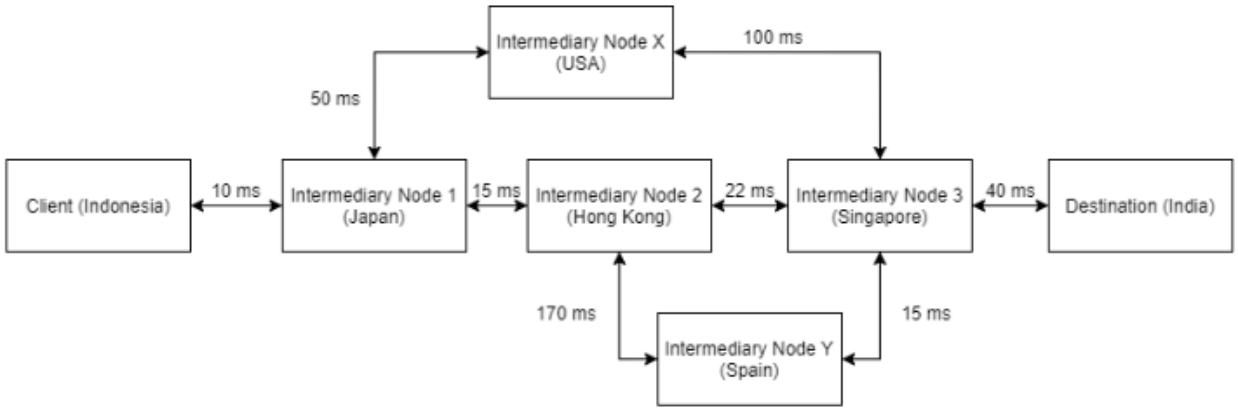

Figure 5. Example of finding the shortest route problem in onion routing

decrypting keys to decrypt the message, while only the client has all of the keys to decrypt the message, and there would be no clue to where the message is being sent to, reason is after the message is sent through an intermediary node, it is then decrypted by one other key, so the other node would not know what the message is trying to access, because after it was sent, it would be a different message from the one the node received.

Using graph, the algorithm needs to find the fastest method of sending the message, and has been said before, while maintaining the security of the message, which means it cannot just send the message to the destination as it would give no benefit to anyone as it would be just a common routing method. For example, if someone in Indonesia were to use onion routing method to send a message to India, the message sent from Indonesia might be routed to Japan, then to Hong Kong, then to Singapore, before reaching the destination of India. But the algorithm would choose the lowest possible latency route to send the message, to maintain the user's secure and great experience.

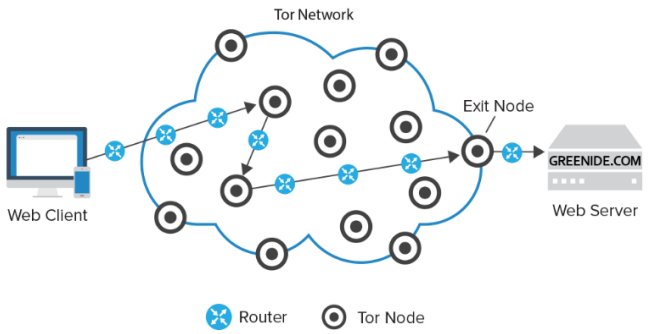

Figure 6. Tor Network represented as a clouded graph
Source: https://c2.staticflickr.com/6/5554/31209782171_2a27d3d3ca_o.png

Figure 6 shows how an actual implementation of an onion routing method would look like from the outside. From the outside it would look like a jumbled connection of intermediary nodes with the client and the destination besides it. In actuality, the intermediary nodes would be randomly connected.

The next problem would be choosing the best path from the user to determine the quickest or shortest path to take from the client to the destination, this is the shortest path problem of a weighted and directed graph. To simplify, the example given would find three intermediary nodes because it is the minimum amounts of nodes to connect to while still maintaining a great security.

For the example given in Figure 5, as mentioned before it is a simplified version of an actual onion routing network. In the example given, a client from Indonesia would like to send a message to India, yet the algorithm needs to find the shortest possible route while still using three intermediary nodes to maintain security of the encrypted message. In the example given above, the fastest possible route to send a message is to send it to Japan from Indonesia, then to Hong Kong, and Singapore, before reaching the destination of India. The weighted graph given above represents the latency between each of the intermediary nodes in milliseconds. From the example given, the connection from Indonesia to Japan would have a latency of 10 milliseconds, from Japan to Hong Kong would have 15 milliseconds, from Hong Kong to Singapore 22 milliseconds, and from Singapore to the destination is 40 milliseconds. The total would be the shortest possible path and meets the requirement of a minimum of three intermediary nodes before reaching the destination. In actuality the node before the destination or as most people call it, the exit node should be a secure node that would be hard to intercept to maintain security, so the exit node would be a trusted node whereas the rest of the intermediary node could be volunteers.

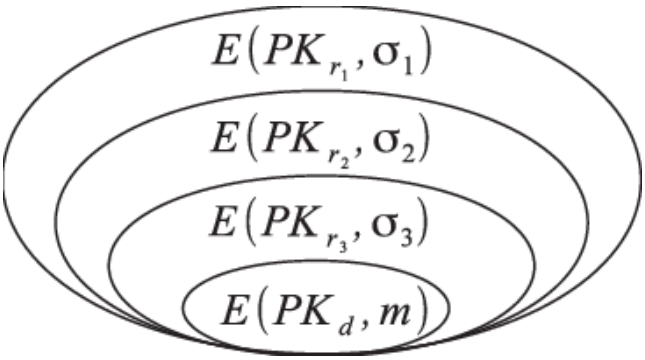

Figure 7. Example of an encrypted message in onion routing

Source: https://www.researchgate.net/profile/Faisal_Alanazi5/publication/315976819/figure/fig1/AS:678370828369921@1538747617336/An-example-of-message-encryption-in-onion-routing.png

The client from Indonesia would have all of the keys to unlock the message, the keys are key one, two, and three to simplify. The client would encrypt the message first with all of the three keys before sending it to other intermediary nodes. From the client to Intermediary Node 1, the message would be decrypted by Intermediary Node 1 using key 1 and then it would pass on the message to Intermediary Node 2, then it would decrypt the message using the key it has and pass it on again. This goes on until it reaches the destination of India, then the destination would respond and send the response to Intermediary Node 3, which would encrypt the message using the key it has, and pass it on to Intermediary Node 2, this step would go on until it reaches the client, which receives the message encrypted with all of the three keys. Then, the client would decrypt the response message using all of its keys and the result is a connection from the client's computer to the page of the site it is trying to access, in this example.

The directory server in onion routing would upload the list of intermediary nodes available to relay the message to from the client side. Then the client would choose nodes from the list uploaded by the directory server to relay the message to, the best possible and fastest route would be chosen to make the user's experience better as has shown above. The server then sends the three public keys to the client to use. While this is an oversimplification of how an actual onion routing works, the principle of onion routing is the same.

VII. Conclusion

Onion routing is a secure and anonymous method of communication over a network to relay a message from the client to the destination over multiple intermediary nodes. While it is secure, this method has its own vulnerabilities, for if an attacker intercepts the entry node and the exit node, guessing what the message sent is would be easy to do. Onion routing battles this vulnerability by using a secure exit node to make sure to make an interception made on its exit node would not be easy. The minimum number of intermediary nodes are three to maintain its security and user experience. Any more than three would not be necessary but could be done to make the connection even more secure. Graph and cryptography theory are used in onion routing as to find the best possible route to make the client's experience the best it can be while cryptography is used to secure the message sent by the client. Onion routing

VIII. Acknowledgment

The author thanks The One Almighty God, while this paper is far from perfection it would not be possible to be completed without the help from The One Almighty God. This paper would not have been complete without the help from friends, parents, and especially Mrs. Fariska Zakhralativa Ruskanda, lecturer of Discrete Mathematics IF2120 Class 03.

References

[1] https://www.onely.com/blog/how-much-content-not-indexed-google-2019/ accessed on December 8th, 2020, at 21:32

[2] https://traversals.com/blog/surface-web/#:~:text=Key%20Takeaways,data%20source%20for%20OSINT%20investigations. accessed on December 8th, 2020, at 21:35

[3] http://cryptowiki.net/index.php?title=Onion_encryption_and_its_application_for_Tor_anonymous_communication accessed on December 10th, 2020, at 19:22

[4] Munir, Rinaldi. Matematika Diskrit, Bandung: Informatika, 2010, edisi ketiga.

[5] Aniket Kate, Greg Zaverucha, Ian Goldberg. Pairing-Based Onion Routing. Waterloo, Canada.

[6] Paul F., David M., Michael G. Anonymous Connections and Onion Routing.

Pernyataan

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 7 Desember 2020

Naufal Alexander Suryasumirat
13519135