International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

A Unified Approach to Defend the Timing Analysis Attacks on Web Traffic

S. Pandiaraj, Suraj Kaushik, Yuvraj Singh Chouhan, Nagendra Sharma, Shreyansh Singh

Abstract*: Security of information transmitted over open systems has pulled in much consideration over ongoing years. Studies demonstrate that attacks dependent on traffic investigation empower attackers to separate helpful data about interchanges between people, regardless of whether the data is transmitted through an encrypted channel. Encryption is one of the most essential requirements in the security mechanism. It provides a safe way to transfer funds, do online purchase, etc. using online banking methods. It is a way that provides the public and organizations to communicate safely and avoid eavesdropping. It should astonish nobody, that encrypted communications are the main prey of DDoS and timing attacks. Such administrations empower access to an abundance of the individual, secret, and monetary information. Personality hoodlums and digital crooks can have a field day in the event that they prevail with regards to breaking web communications encryption. This paper introduces web traffic timing attacks on encrypted web traffic. In this method, the attackers use only the time value mentioned on the TCP packet and use it to perform the attack. Therefore, this type of attack cannot be merely solved by just padding. The reason this attack is difficult to defend is that the attacker does not require information about the sending and receiving ends due to which its effective against traffic streams. In the proposed system, a new lower overhead tunnel is suggested whose traffic cannot be analyzed. The fundamental thought utilized is to guarantee that, given a watched bundle follow, a wide range of groupings of web gets could sensibly create this follow. Clients are hence given solid deniability that a particular website page was gotten. This concept of imperceptibility is not new, however, the approach and methodology used to acknowledge it is new and much better.

***Index Terms*: Encryption, Padding, Timing-attack, Tunnel.**

I. INTRODUCTION

Today's internet world has a large number of online systems, that can be attacked by hackers from all around the world. These systems need to be protected from all sought of attacks as they can compromise the user data and cause great harm to the industry business by stealing its secret information. Daily we hear of large number of attacks being performed and different kind of security breaches happening in organizations and people's online applications. Web traffic time analysis attack is also a similar type of breach that should be taken care of. Presently the essential system for shielding system traffic from snooping is by utilization of encryption. Encryption is the strategy by which plaintext or some other kind of information is changed over from a clear structure to an encoded form that must be decoded by another element on the off chance that they approach an unscrambling key. Encryption is a standout amongst the most vital strategies for giving information security, particularly for the sender to receiver insurance of information transmitted crosswise over systems. Industries and people likewise normally use encryption to ensure delicate secret information present on their devices. This is viable at disguising the substance of individual packets, however, it is additionally realized that the packet streams can even now uncover data about client action for example by means of the packet sizes, timing, sender/receiver IP address and so on. Considering attacks and protections based around this kind of bundle stream data is the focal point of this theory. Our point is to pick up an improved comprehension of assaults that prompts the structure of better guards. Website fingerprinting (WF) is a traffic investigation attack with the potential capacity to break the protection that secured encrypted network tends to provide. WF permits the intruder to recognize site pages in an encoded association by examining designs in system traffic. This permits a nearby and detached organize foe, for example, a client's Internet specialist co-op or somebody sniffing the client's remote association, to recognize the sites that the client has visited and the information they have shared in spite of user using encrypted system. Web servers have a wide range of clients; few may connect with vindictive exercises to corrupt or totally square system administrations, for example, flooding assaults. Thus, bunches of asset and data transfer capacity on sites may be squandered. There are a few fundamental ways to deal with the barrier that has been considered to date. First is to jumble the timing of data parcels by sending bundles at a consistent rate, utilizing buffering and inclusion of fake packets. This method was quite successful in upsetting assaults but had a large overhead due to which it cannot be used for daily security purpose. A second, application-layer, methodology is to randomly change the pipelining of the requests shaping the get a page and furthermore to infuse fake requests. Notwithstanding, while at the same time bringing about a low overhead this is generally incapable against current assaults. A third profession depends on moulding traffic with the goal that the communicated follow is comparable, in some suitable sense.

Manuscript published on 30 April 2019.
* Correspondence Author (s)
S. Pandiaraj*, Computer Science & Engineering, SRM Institute of Science & Technology, Assistant Professor, Ramapuram, Chennai 600089, India.
Suraj Kaushik, Computer Science & Engineering, SRM Institute of Science & Technology, Student, Ramapuram, Chennai 600089, India.
Yuvraj Singh Chouhan, Computer Science & Engineering, SRM Institute of Science & Technology, Student, Ramapuram, Chennai 600089, India.
Nagendra Sharma, Computer Science & Engineering, SRM Institute of Science & Technology, Student, Ramapuram, Chennai 600089, India.
Shreyansh Singh, Computer Science & Engineering, SRM Institute of Science & Technology, Student, Ramapuram, Chennai 600089, India.

This paper presents another class of lower overhead tunnel that is impervious to traffic investigation. The fundamental thought utilized is to guarantee that, given a watched bundle follow, various groupings of web brings could sensibly create this follow. Clients are in this manner given solid contradiction that a particular website page was gotten. This lack of definition thought isn't new, yet the methodology used to acknowledge it is advanced. Essential highlights are that (I) our methodology is lightweight and does not, for instance, require pre-grouping of website pages and (ii) bundles from various web gets are accumulated subsequently enabling us to sharply lessen the number of fake parcels sent amid occupied occasions when numerous fetches are at the same time dynamic while keeping up precise security properties. It was noticed that the fake bundle overhead is normally less on delicately stacked connections and the division of fake bundles tumbles to zero as the traffic load increments, for example, the passage is limit accomplishing. Vitally, we manufacture an exploratory model of the tunnel and do a broad execution assessment that shows its viability under a scope of system conditions and genuine site page gets.

The significance of utilizing an employed execution is double. First thing is that the buffering related with traffic moulding influences the conduct of the web get, for example, postponement of demand has thump on consequences for the planning of the reaction which may influence consequent requests, and such communications are not caught by the standard re-enactment approach of repeating a recorded bundle grouping. Furthermore, functional issues, for example, treatment of DNS requests and communications between traffic forming and TCP congestion control, that have gotten less consideration in the traffic analysis writing, are featured.

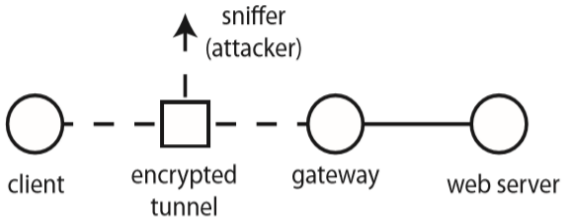

Fig. 1: The Figure represents the basic working of webpage client and server.

II. PRELIMINARIES

A. Life structures of a Web Page

In this module, we will discuss how the data packet timing is having this much of information, which an intruder can misuse. Let's begin with how the data packet is encrypted? The traffic is encoded then the data packet is sent. The starting address and the end address and the packet format and ports are hidden and the data packets are padded. The person who is trying to capture the information, first of all, tries to capture the encoded traffic with the help of which it is possible for the intruder to find the data packet timings and the direction flow of the packet. As shown in Fig. 2 the plot shows as about the timestamp trace of different webpages. In X-axis, represent the Packet number and the Y-axis represents Timestamps. It is observed from the fig. 2 that different value at X-axis and Y-axis are showing different timestamp traces correspondingly.

Let's get in more detail how there is different webpages having different data packets timestamp as shown in Fig. 2, it will explain in detail how the webpage process. When a client side webpage open the first thing which takes place is a TCP/IP connection is established with the Server Side with the help of an URL and assigns an HTTP POST or HTTP GET Request. The server side talks with the HTTP GET or POST requests only. The User request from the Client side using the HTTP GET/ HTTP POST to the server. The Server side responds to the client side using the same.

The Client can also request to other servers also and get back the responses from the server. In this case, the client is required to create a TCP/IP connection with the new server to get a response from the server. Here the page loading is done using asynchronous loading using dynamic ways. The complex request and responses here are handled by AJAX. After the TCP/IP connection is established with the server, it remains connected until the webpage is fetched completely and it remains connected so that it can be further used if needed. When a number of process are taking place then objects in a webpage create time trace that is called as Signature.

B. Risk Model

When we consider privacy, the thing that comes into mind is how the internet and the client side are connected. By which means they are connected. There are generally two types of major risk that we consider:

- The mobile hotspot and Internet Service providers can capture the data packets sent to throw the wired Connection/ DSL by connecting to the end of the network.
- When a device is connected by Wireless link then the Wi-Fi capture data packets sent across the wireless ends.

Here the main thing needs to focus on is the data packets. They could be traced. The data packets is of two parts: 1) Packet Header and the other one is 2) Data/ Content. The data packet consists of the timing, which the packet is sent with is parts.

Let consider the data packet are padded for the same size that the user has already basis defense instead of sending traffic using a suitable tunnel and the data is encrypted this is shown in Fig. 1 The main part for an intruder is to find the data packet timing for data packet passing.

Since the major content of traffic on the internet is related to the literature. Therefore, here we focus on the web Traffic for finding web traffic attacks. Here our main purpose is not to know what the client sitting at the end is surfing, but here our main motto is to find and stop the intruder from deducing which webpages are surfed by the client.

As we have discussed before the data packets timing, they are an abundant source of information for the intruders, and it is more than enough to help an intruder to capture the data of web page being surfed by the client with a high probability ratio. The data packet padding is not capable of hiding all the features like the total packet size, time to load the page and data burst.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

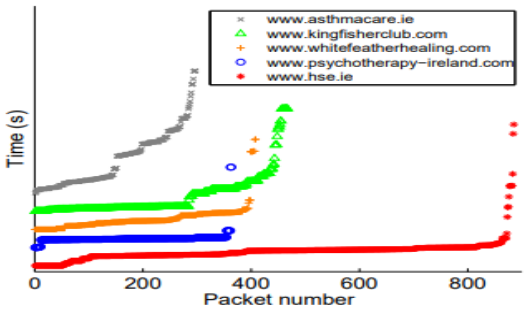

Fig. 2: The Figure represents the time trace of different websites.

Even though the data packets are, sent from the client devices to the internet in many ways such as through web application, Email, and web and it is to be taken care that the intruder cannot be able to guess the number. The website Time trace always shows a different kind of pattern. Fig. 2 will show the time trace of some websites. Time trace is a kind of attack, which is done by the intruder with exact same configuration. Here the microseconds are considered. The traces are created in such a way that they do not overlap each other. They are shown vertically and we can compare time trace of different websites.

Here, when we look at the fig. 2 then we can see the flow of data packets over the time is quite difficult to guess by the intruder. So here, we need a security software that is applicable on, when a single web page is fetched or many web pages are fetched at the same time. Here we are using security software because to reduce the effect of the overhead attacks the intruder might do that. Virus's kind of attacks are not checked.

III. RELATED WORK

Here when we consider the attack then timestamp is considered the main factor in our paper. As we look at the internet, the main content we will find is literature. For every webpage, a signature is generated for the size of bytes received by the webpage on its different port, and the signature depends upon the webpage objects and the number of objects for each webpage. As we can consider that, we have some datasets that can provide some hints using which we can guess the attack. Therefore, there are some hints to be considered to stop attacks a) The size of the data packet seen. b) The port number that is used by the user. c) The moving direction of the data packet (which check the data packet is coming or going from one point to other). d) Data packet, data is captured from it with the starting and ending of each webpage.

To protect the system from timestamp attacks here we will be using a technique that is known as Tunneling. Tunneling is a rule that helps in secure communication and the data packets are sent from source to destination in a secure way in a network or different networks also. In our project we observe a) The size of the data packet seen. b) The port number that is used by the user. c) The moving direction of the data packet (which check the data packet is coming or going from one point to other). d) Data packet, data is captured from it with the starting and ending of each webpage. The intruder here uses the timestamp for attacking. First, the data packet is having a header, data, and timestamp. When we want to send the data from the source, then the data is provided at the beginning of the network layer, it takes time for the data to reach the application layer, and then only the data is sent. Here the timestamp is used when the data is sent from the application layer to the destination and back to the source. The RTT (Round trip timing) is considered here. Here we need to remove the time taken for data to move from the lower layer to the application layer, we do not consider that time in RTT. So, what does intruder do with the timestamp? The intruder is a very smart person, first of all, they use the same configurations for everything whatever the user is using like CPU with same processing speed, same OS, same software, each and everything will be same. Now the attacks take place like I have two fields for login: Username and Password. When the intruder tries to guess and he enters the Username and checks the timestamp if the time taken for the query is more then, the timestamp attack is done successfully but if the time taken is very less then it means his/her guess is wrong. The time here is in microseconds that cannot be measured by human beings, the time measurement is done by computers. Using timestamp attack is first done for the username if he finds that the attack is correct and then he/she proceed to do the brute force attack on the password to find it. Here with the help of time difference timestamp attack is done.

To protect from timestamp attacks we use tunneling. In tunneling, we are putting a size limit and the constant time for the data packets to be transferred. Now we the intruder tries to do the timestamp attack and he guesses for the username, if the username guessed by him is correct then he will get a constant time and if he/she entered a wrong username then also he/she will the same time. Here when the guess is correct then no dummy packets are added to the tunnel but when the guess is wrong the time taken by the query is less so, dummy packets are automatically added to make the time constant. Now the intruder will not be able to get the timestamp difference and every time he/she tries, they will get the same time duration. That will make the connection more secure from attack.

Here we will also be using machine-learning algorithms for predicting about the attacks based on their timestamps timings. We are going to collect the data such as timing for the user correct login and the timing for the user incorrect login and attacks. Using that data, we will create a plot, with the help of which we can predict the ratio of attacks happened and we can also use tunneling, after that we can plot, and then we can compare the different plots.

IV. ALGORITHMS USED

In our project, we employ two widely used machine learning (ML) algorithms: k-nearest neighbour (k-NN) and support vector machines (SVM).

In the following sections, we provide brief explanation of this two techniques.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

A. k-Nearest Neighbour (k-NN)

The k-NN algorithm can be used for both classification and regression. In website fingerprinting (WF), we use k-NN as a classifier. As k-NN is a supervised ML algorithm, we feed the data along with the associated label to train the classifier. In our case, the instances of the websites are the data, and the name of the websites are the labels.

We train the k-NN classifier with training examples (x, y) where x is the feature and y is the target label or class. The objective is to learn the function g : X → Y to find the relationship between x and y so that g(x) can predict the correct label of an unseen x. For every data points, the algorithm measures a similarity index between the k nearest neighbours and the data point x. A popular similarity matrix among researchers is Euclidean distance.

B. Support Vector Machines (SVM)

Support vector machines (SVM) is one of the popular traditional machine learning techniques for solving classification problems. The basic idea is to plot the data points (features) in a n-dimensional space, where n is the number of features. weight or value of each feature is the coordinate value in the space. The Classification is done by separating the features by finding hyper-planes that divide data points in each class. We want our features to be as farther as possible from the hyper-planes. Hyperplanes can easily be useful to classify the linear features. Classification becomes challenging when the data points are non-linear. For example, network traffic data used in WF attacks have non-linear features. To address this, we apply a technique called kernel trick. The kernel trick transforms low-dimensional data into high-dimensional data. In the higher dimension, the data points become linearly separable

V. ARCHITECTURE AND MODULES

A. AJAX Request and Response

When the user start the webpage then first a connection is established with the server with the help of URL using the TCP connection. The client request using HTTP GET or POST to the server. The client can request from server and get data like images, audio, video docs, scripts, etc. The clients can requests to other servers also to get some other data from them but it requires to form a connection and start using TCP/IP with the new server. The connection will remain connected so that the user can request data and get responses from the server. To get the asynchronous dynamic data here we are using AJAX. AJAX is used for a process of sending complex request to the server and receiving the complex responses from the server to the client.

B. Attacks

As we look at the internet, the main content we will find is literature. As we can consider that, we have some datasets that can provide some hints using which we can guess the attack. Therefore, there are some hints to be considered to stop attacks a) The size of the data packet seen. b) The port number that is used by the user. c) The moving direction of the data packet (which check the data packet is coming or going from one point to other). d) Data packet, data is captured from it with the starting and ending of each webpage.

C. Defenses

To protect from timestamp attacks we use tunneling. In tunneling, we are putting a size limit and the constant time for the data packets to be transferred. Now we the intruder tries to do the timestamp attack and he guesses for the username, if the username guessed by him is correct then he will get a constant time and if he/she entered a wrong username then also he/she will the same time. Here when the guess is correct then no dummy packets are added to the tunnel but when the guess is wrong the time taken by the query is less so, dummy packets are automatically added to make the time constant. Now the intruder will not be able to get the timestamp difference and every time he/she tries, they will get the same time duration. That will make the connection more secure from attack.

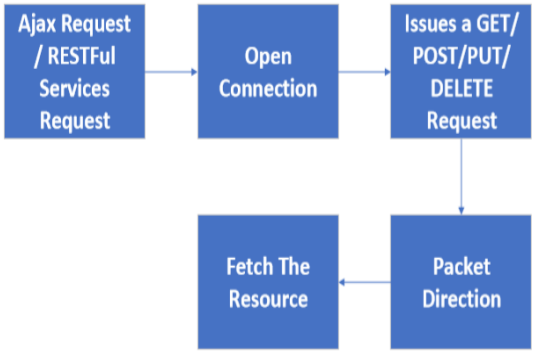

Fig. 3: The Figure represents the architecture

VI. DUMMY PACKETS USES

The dummy packets are very useable thing that we will be using to secure our system from intruders. We are be using dummy packets dynamically. The dummy packets are managed by the dummy manager depending upon the overhead load. They are helpful in securing the privacy of the user. We use dummy packets when we use the tunneling approach. When the traffic is low then the number of dummy packets inserted is more because of the fixed size is to be maintained for the tunneling effect and when the

traffic load is quite high then the dummy packets can be zero to as much as required to fill the fixed size. This help is reducing the time trace attacks.

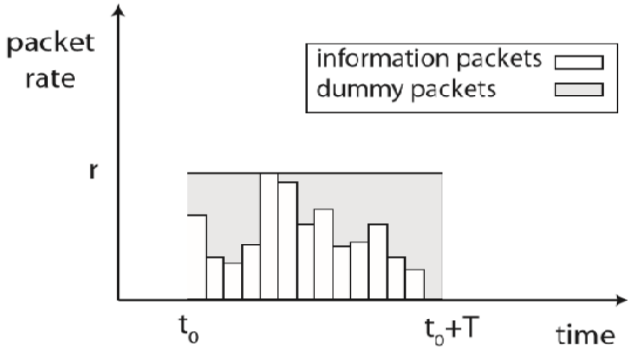

Fig. 4: The Figure represents a graph between dummy packets and information packets

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

VII. ACHIEVING PRIVACY

In the segment the traffic model is to verify the indistinguishability that implies that when a packet is traced then through that traced path different combinations of web fetches.

It can be inconsiderably accomplished by sending packets frequently and utilizing buffering in addition to inclusion of duplicate packets as required, the overhead included is intemperate. The test is to strike a harmony among lack of definition and the related traffic model overheads of buffering delay and duplicate packets.

There are mainly two conclusions. First is the traffic modelling which can conform on the trace-based method. The other method is the optimization of traffic profile which can be done by using the information of web fetch packet traces. But by this method it finalises to traffic profiles which consumes less number of duplicate packets, but it can be hard to trade with web pages which are unknown. Second is that, instead of using traffic modelling on each and every single web fetch the other method is to find the average of traffic sharing and then using traffic modelling on it. The process of modelling is processed on every web fetches and the overhead on every fetch is constant and the total tunnel overhead of scales with different fetches. There is no use of duplicate packets when the traffic is low then the number of dummy packets inserted is more because of the fixed size is to be maintained for the tunneling effect and when the modelling is processed on the average traffic and because of the modelling technique used there is a rise in the traffic due to that the overboard becomes null.

There are three ways in the privacy can be achieved and those three ways are going to be discussed in detail in this section only. The three ways are-

A. Reduction In Duplicate Packets Overhead Obtained By Modelling Of Average Traffic

B. Adaptive Trace-Based Approach

C. Indistinguishability

A. Reduction In Duplicate Packets Overhead Obtained By Modelling Of Average Traffic

The process of buffering is applied on the packets by time analysis attacker at the tunnel ingress till the number of packets are enough to remove the trace that the overhead of the duplicate packets be avoided.

B. Adaptive Trace-Based Approach

Information packets are buffered when too many of packets are available and the buffering is done till the transmission property arises. In this adaptive trace-based approach when the time take by a web fetch is more than the duration of an individual trace than addition traces are activated. In this approach problem arises with the packets generated by web fetch is that it consists of more than one packet or precisely saying maybe hundreds or thousands of packets and because of the number of packets these can be distinguishable easily because it is a large amount of data. So, to solve this problem the data or the packets are divided into groups and then transmitted.

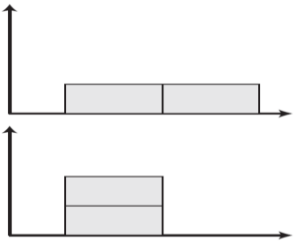

Fig. 5: The Figure represents a graph of dummy and information packets

C. Indistinguishability

The time analysis attacker is visiting the pages which have been visited recently adjacently surfing the internet also. So the indistinguishability of the information of the web fetches and the packets which are generated by because of the privacy embedded in the indistinguishability.

So, in indistinguishability there is a different approach for the privacy. The web fetches of a individual web page is not processed instead of that every web page fetch is used. We require fetches of every web page because if a large web page fetch is there it would require some time to be generated and for small pages it might require less time but they are distinguishable because of the time taken so because of that web fetches of small web pages are generated because these fetches cannot be distinguished because they are atomic in nature and are indistinguishable from one another.

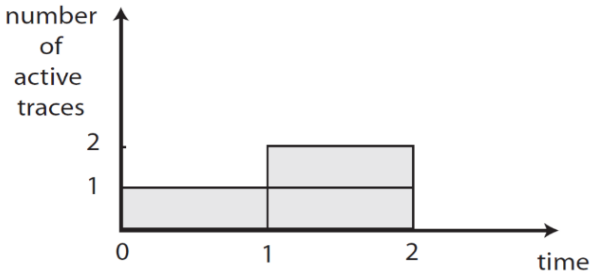

Fig. 6: The Figure represents a sequence of traces.

VIII. CONCLUSION

The main purpose of our research is to find useful timing features that can be used in Website Fingerprinting attacks. In the proposed system a new lower overhead tunnel is suggested whose traffic cannot be analysed. The fundamental thought utilized is to guarantee that, given a watched bundle follow, a wide range of groupings of web fetches could sensibly create this follow. Clients are hence given solid deniability that a particular website page was gotten. The architecture is made in such a way that the users traffic demands are satisfied by the communication of traces while it also reduces the delay and throughput overhead due to fake or dummy packets.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

REFERENCES

1. G. D. Bissias, M. Liberatore, D. Jensen, and B. N. Levine. Privacy Vulnerabilities in Encrypted HTTP Streams. In G. Danezis and D. Martin, editors, Privacy Enhancing Technologies, volume 3856 of Lecture Notes in Computer Science, pages 1–11. Springer Berlin Heidelberg, 2006.
2. X. Cai, R. Nithyanand, and R. Johnson. CSBuFLO: A Congestion Sensitive Website Fingerprinting Defense. In Proceedings of the 13th Workshop on Privacy in the Electronic Society, WPES '14, pages 121–130, New York, NY, USA, 2014. ACM.
3. K. P. Dyer, S. E. Coull, T. Ristenpart, and T. Shrimpton. Peek-a-Boo, I Still See You: Why Efficient Traffic Analysis Countermeasures Fail. In Securityand Privacy (SP), 2012 IEEE Symposium on, pages 332–346, May 2012.
4. S. Feghhi. Seculink Tunnel Project: Codes and Measurements, 2016. available at: www.scss.tcd.ie/disciplines/stable/seculink.
5. S. Feghhi and D. Leith. A Web Traffic Analysis Attack Using Only Timing Information. IEEE Transactions on Information Forensics and Security, 11(8):1747–1759, April 2016.
6. T. Flach, N. Dukkipati, A. Terzis, B. Raghavan, N. Cardwell, Y. Cheng, A. Jain, S. Hao, E. KatzBassett, and R. Govindan. Reducing Web Latency: The Virtue of Gentle Aggression. ACM SIGCOMM Computer Communication Review, 43(4):159–170, 2013.
7. X. Cai, R. Nithyanand, T. Wang, R. Johnson, and I. Goldberg. A Systematic Approach to Developing and Evaluating Website Fingerprinting Defenses. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14, pages 227–238, New York, NY, USA, 2014. ACM.
8. X. Cai, X. Ch. Zhang, B. Joshi, and R. Johnson. Touching from a Distance: Website Fingerprinting Attacks and Defenses. In Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS '12, pages 605–616, New York, NY, USA, 2012. ACM.

AUTHORS PROFILE

S. Pandiaraj received his Bachelor of Engineering in Computer Science and Engineering from Madras University in 2001. After graduating in his bachelor's degree, He pursued in Master of Engineering at Sathyabama University, Chennai in the year 2010. He is Assistant Professor at SRM Institute of Science and Technology, Chennai, Tamil-nadu. His Research interest is in AI, Network Security, Computer Graphics, Compiler Design, DBMS, Web Technology.

Suraj Kaushik is currently pursuing his Bachelor of Engineering and Technology in Computer Science and Engineering at SRM Institute of Science and Technology in 2019. He is passionate about the DBMS, SQL, Machine Learning, Cloud Computing and Networking. He has developed the project based on cloud computing and Data Analysis.

Yuvraj Singh Chouhan is currently pursuing his Bachelor of Engineering and Technology in Computer Science and Engineering at SRM Institute of Science and Technology in 2019. He is passionate about the Devops, Machine Learning, Cloud Computing. He has developed the project based on cloud computing and Data Analysis.

Nagendra Sharma is currently pursuing his Bachelor of Engineering and Technology in Computer Science and Engineering at SRM Institute of Science and Technology in 2019. He is passionate about the DBMS, SQL, Machine Learning, Cloud Computing. He has developed the project based on cloud computing and Data Analysis.

Shreyansh Singh is currently pursuing his Bachelor of Engineering and Technology in Computer Science and Engineering at SRM Institute of Science and Technology in 2019. He is passionate about the DBMS, SQL, Machine Learning, Data Analytics. He has developed the project based on cloud computing and Data Analysis.

.