Journal of Mines, Metals and Fuels, 71(5): 695-697; 2023. DOI: 10.18311/jmmf/2023/34172

Print ISSN : 0022-2755

Journal of Mines, Metals and Fuels

Contents available at: www.informaticsjournals.com/index.php/jmmf

User Profiling Behaviour and Activity of Persistent Surveillance of User Activity in Complex Networks

Koustav Dey* **and Pratik Bera**

Swami Vivekananda University Kolkata, India. E-mail: koustav1998dey@gmail.com / pratikbera.c@gmail.com

Abstract

The objective of this paper is to study the behaviour of multiple user activity on network traffic.

The traditional approach of using the source IP address as a tracking identifier does not work when used to identify mobile users. Insider misuse has become a significant issue for organizations. Traditional information security has focused on threats from outside rather than employees. A wide range of concepts has been undertaken to develop approaches to detecting the insider.

Recent efforts to address this problem by exclusively not relying on web browsing behaviour, collecting data from routers, and predicting human psychology (when a router receives a packet, the router checks its routing table to determine if the table lists the destination address in the header. If the table does not contain the destination address, the router forwards the packet to another router that is listed in its routing table).

This paper builds the concept of tracking behaviour-based identification of multiple user activity on a router. Here we find out the traffic of HTTP or HTTPS Protocol and the Results are promising. find out as the result we are capable of tracking human psychology.

Keywords: *Router activity, Surveillance, Machine Learning, Networks.*

1.0 Introduction

The Internet is the greatest invention of science. It helps us in getting every piece of information with a click. We can share information, and get connected with people around the world through the use of the internet.

In today's world, everyone uses the internet and searches for data as required. But the uses of the internet totally depend on the user. Users can search for any kind of useful or offensive content. It is little to find which people use the internet for offensive content and identify the user. It's also hard to figure out why the user searching for offensive content.

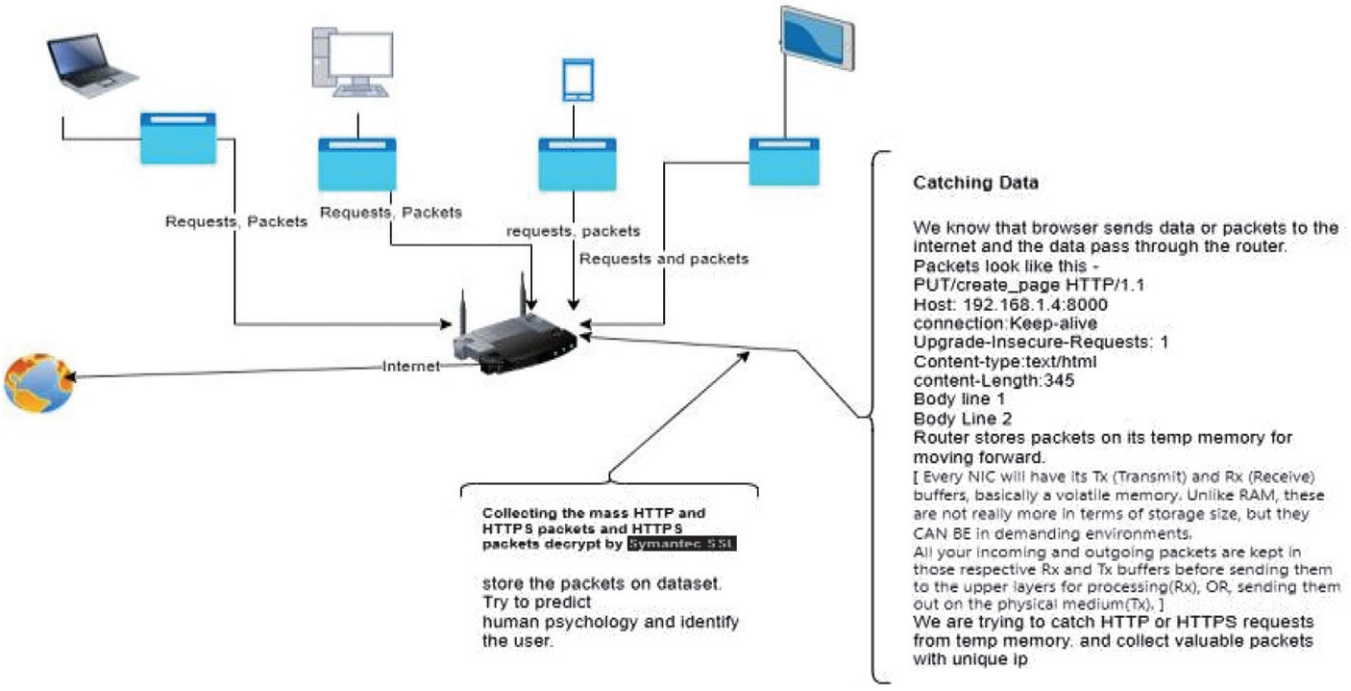
But in these days many tools are already developed to find and identify the user.

In the past days, max work was system browser activity-dependent but now we are introducing an environment that is network-dependent. So, now we don't need permission from any individual system. Our system is taking data in-network level so we can track user browser activity who is under system installed network. Previously[1,2,3], much research work was done on browser activity but our concept does not only depend on the user system it depends on a network system. Previous work only collected data from browser cookies, traffic, server, etc. However, we are acting like a router and gathering senders' packets from the core network model.

We will attempt to forecast human psychology using the packets fed into an ML algorithm (similar to SVM, RF, and kNN), and through the packet, information must identify the user.

On the other hand, we noticed a maximum number of tools are used for monitoring browser activity[2] for security

*Author for correspondence



Model: User Profilling behaviour and activity of persistent surveillance of user actyivity in complex networks

purpose[4]. But it is not a complete system it's a long-term process. Our idea establishes an easier concept for identifying user activity on complex networks. Every institution or company uses its own networks. Additionally, all systems are linked to the same networks. As a result, a particular router passes through each request for a system packet. Requests are forwarded by the router to servers or the internet. We can try to locate packet requests in the router's temporary memory, collect them, and utilize the requests to determine who the user is. Then we'll be able to predict user psychology and user thinking with AI algorithms.

2.0 Methodological Aspects

Modelling Profiling behaviour-based surveillance

Our behaviour-based tracking technique exclusively relies on the packets sent by the router, thus we are basing certain notions on the positive outcomes of earlier research investigations on the complex network[6].

In this model, the timing and request order is not taken into account.

We presume that any service provider using behaviour-based tracking on a complicated network can capture user interactions with a number of target hosts. Now we are try to understand the process to track behaviour based activity and predict human physiology.

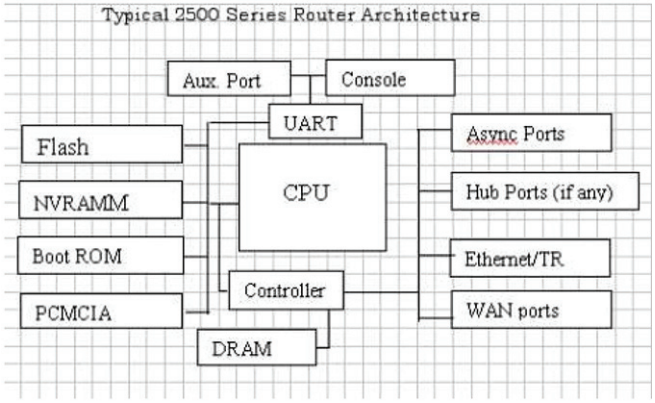
We assume that each user is represented by a dynamic IP address. Furthermore, we are aware that routers can support 4 different types of memory. The active tasks are kept in RAM. NVRAM is used for the startup setting. The POST and bootstrap software information is stored in ROM. and Flash Memory, which houses the Router IOS operating system.

That, after a predetermined period of time, is changed, moreover, As far as we are aware, a browser transmits data or packets to the internet, which then transit through a router. Packets look like this -

PUT/create_page HTTP/1.1 Host: 192.168.1.4:8000
connection:Keep-alive
XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
Upgrade-Insecure-Requests: 1
Content-type:text/html
content-Length:345
Body line 1
Body Line 2

Router stores packets on its temp memory called ROM for moving forward. collecting each packet through wiretrapping. [Wiretapping is the act of recording communications between parties, often without their consent] and collects the packet. Most of the browser packets are encrypt by https protocol [Hypertext Transfer Protocol Secure] or the packets are pass through by http protocol]. Here the encrypted packet decrypt by ssl stript[5] and we collect the raw data. now the dataset is ready.

After that, the system will make an individual report for each mac address identified person and focus on each keyword of the packet, depending on the words system will predict the



type of word by taking the help of the ML algorithm and predict the thought process of the person. For generating the thought process as a final report our system will depend on the previously collected data which was already generated in the previous certain time period.

A router uses a Linux-based OS (Operating System) like OpenWrt, DD-WRT, MikroTik, etc and there is also a processor and ram for data processing. That is like Linux Operating System installed PC. Our system will make a copy of the data and MAC address that is passing through this router. The most copied data will be in HTTP or in HTTPS protocol format. The fact that HTTP is a client-server protocol that enables internet users to ask web servers for web pages. It is a commonly used application-level protocol on the Internet. A browser sends an HTTP Request message to the web server whenever a user wants to visit a web page. The web servers typically use TCP port 80.

Today, HTTP/1.1 is the version of HTTP that is most often used. There is a more recent version, HTTP/2, which is supported by the majority of browsers. On the other side, HTTPS, which stands for Hypertext Transfer Protocol Secure, is an encrypted version of HTTP. By utilizing encryption, this protocol facilitates secure communication between a client, such as a web browser, and a server, such as a web server. HTTPS encrypts data using the Transport Layer Security (TLS) protocol or Secure Sockets Layer (SSL), which is one of its predecessors. The well-known TCP port 443 is used by HTTPS. Browsers will use this port for sending HTTPS requests if the port is not mentioned in the URL. So it needs to decrypt all this data. Our system will decrypt the data by using a tool like an SSL strip or SSL downgrade tool. SSL stripping is a technique by which a website is downgraded from HTTPS to HTTP. After that, our system will store this meaningful data, MAC address, IP address, date and time in the database. After some time gap, our system will sort the data and make a new dataset depending on the MAC address and store all queries searched by a particular MAC-addressed user. then our system will make focus on searched words for individual MAC addresses. Our system will set a flag for how much time the user searched one kind of word like offensive, mathematics-related, food-related, or technology related. Then it will try to make a prediction of which kind of word the user is searching, for by using an ML algorithm and dataset. Then our system will try to find if there is any link between different query which is searched by the same user. Then our system will give a tagline for defining the user depending on their searched data. This user is interested in maths, this user is interested in cyber crime or this user is a foodie. After a while, our system will remove all unnecessary data.

3.0 Conclusion

The behaviour-based tracking scheme studied in this paper is capable of monitoring the user. Now the concept is playing an important role in the corporate sector or institutions for security purposes. The future scope and the extension of this paper is possible on large scale.

According to the survey, many people look for undesirable things on search engines, which has a bad impact on global civilization. That's why router monitoring and user perception identification may be viable alternative to protect society from this menace. In another hand, we can monitor the list of productive, responsible, and safe people for an Institute or Society. If we use it on a large scale in the future then it can help National Defense System by identifying Terrorist mindset persons and can spot their location.

Our system is easily extendable and can be improved further for new and improved upgrades. Our system has a big scope to do a predicted list of people depending on their browsing activity and provide protection of the company trade secrets and for the better work output of employees, analysis of members under a single environment (School, College, Digital Library, Society), and can upgrade it to the national level if needed in future. Though our project is mature enough still there is still scope for betterment as it's always an open door.

4.0 References

1. Forouzan, Behrouz A. (2007): Data communications and networking. Huga Media.
2. Taglienti, Claudio, and James Cannady. (2016): "The user attribution problem and the challenge of persistent surveillance of user activity in complex networks." *Journal of Computer Security* 24, no. 2, 235-288.
3. Ms. Sakshi Sanklecha, Mr. Darshit Deotale, Ms. Jyoti Yadav, Ms. Dipti Mishra, Prof V. P. Yadav, (2022): User Activity Monitoring System/SPYWARE, *International Journal for Research in Applied Science & Engineering Technology* (IJRASET) ISSN: 2321-9653; Volume 10 Issue III Mar 2022, PP 1382-1389.
4. Karamanos, Emmanouil. (2010): "Investigation of home router security.".
5. Stallings, William. (2006): "Cryptography and network security principles and practices 4th edition."
6. Fan, Liyue, Luca Bonomi, Li Xiong, and Vaidy Sunderam. (2014): "Monitoring web browsing behaviour with differential privacy." In Proceedings of the 23rd international conference on World wide web, pp. 177-188.