ISSN No:-2456-2165

Data and Cyber Security: The Study of Facebook Meta

Soniya[1]
LLB Student at Law Centre II,
Faculty of Law, University of Delhi
Address: RZ-344 A/2 street no. 18J, Palam Colony, New Delhi Delhi-110045, India

Abstract:- The world is on the verge of immense change in technology. It would be faultless to say that developed economies contain giant industries of technology such as Apple Inc., Google Inc., Intel Inc., Microsoft Inc., Facebook Meta Inc., etc. There are multiple factors due to which nascent technological industries approach developed nations one of the many factors is strong economic position. This renders, developing countries further behind, and developed countries grab the opportunity. Corporations that are incorporated or have their base locations in developed economies capture the trust of the people at their first instance. This study concluded the Encryption Methods, Data Security and Cyber Security followed and adopted by Facebook Meta. The methods and data contained in this study are for the year 2023. This study concluded that Artificial intelligence, Data Protection, and public and private key encryption are the major areas of interest for many researchers in the past 20 years. The study also contains an analysis in the area of threat and vulnerability of Cyber security. The paper attempts to explain JavaScript code, Multi-key Private Matching, Public key encryption in Easy Crypt, SALSA Cryptographic system, and Differential privacy used by Facebook Meta. The paper also includes a complete description of Millisampler and the role it plays in network intrusion in Facebook Meta and its impact on cyber security and data protection. Cyber security is fundamental for data protection managed on the computer database system. Therefore, Data security becomes particularly more important for the data stored on different platforms such as PCs, and various devices. A cyber-secured organization can efficiently and effectively manage its resources as well as human assets. Cyber security prevents any unauthorized usage, storage, alteration, deletion, encryption, decryption, access, addition, and manipulation of information or data of any user. Due to the expansion of the technological environment in every corner of the world, it would be significant for the technical giants to provide their users with an environment of security for their information. Cyber security not only prevents any data breach but also prevents data from cyber-attacks. This paper studies the modeling and analyzes the methods used by Facebook Meta.

Keywords:- *Facebook Meta, Cyber Security, Data Security, and Data Privacy.*

I. INTRODUCTION

Cybersecurity has been a popular subject in academics and literature. Cyber security is an application of systems and programs to protect the data of the users stored in a database. According to UNCTAD, the United States of America stands out at the top 1 in the ranking of 166 countries including developed and developing economies readiness to use Frontier Technologies. The research comprised the factors of ICT, Skills, R&D, Industry, and Finance. Frontier Technologies comprises of technologies such as intelligence, the Internet of Things, and green hydrogen[1]. A large number of works in this field of research have shown that due to the rise in technology cyber-attacks have been reduced. Hitherto the government and organization itself always focused on minimizing the threat and vulnerability of cyber security[2]. Numerous data protection and cyber security methods have been adopted and followed by technical giants. This study limits itself to the data protection and cyber security techniques followed by Facebook Meta 2017-2023. According to many authors and the study done by Meta itself, it has been found, a major potential threat is security[3]. The study focuses on how a technology giant in the world manages, maintains, and secures the records of its users.

- *Objective of Research*

There are numerous work in this field which is merely descriptive or prescriptive. In 2020 about 1.5 lakh research publications out of 26 lakh global publications in science, technology, and engineering domains published in peer-reviewed journals contributed by Indian students[4]. For a meticulous analysis, the study includes robust cyber security parameters followed by Facebook Meta to protect the data of its users. Every data used in the study is for the year 2020-23, so that it states the condition precisely. Considering this aspect, this paper focuses on explaining various methods and techniques followed by Facebook Meta to secure the data and information of the users.

- *My Contribution*

The present study involves the identification of various cyber threat resolution mechanisms adopted or followed by Facebook Meta to protect the data of its users. The study will indeed help many researchers and academicians to know as to how the technology giant Facebook Meta contends with various threat models in its security or privacy breach. The paper provides the analysis based on various research conducted by various professionals in the

field of technology which is being used by Facebook Meta (hereinafter referred to as The Inc.) for the year 2017-2023 which will help the researchers to identify various measures for cyber threats and can publish their quality research papers.

➢ *Background*

Facebook has emerged as one of the technology giants in the previous two decades. There are approximately 15 social media platforms having billions of active users at an international level[5]. Hitherto Facebook has maintained the highest among all social media platforms used having 2.989 Billion active users at present i.e., July 2023[See Figure 1 below].

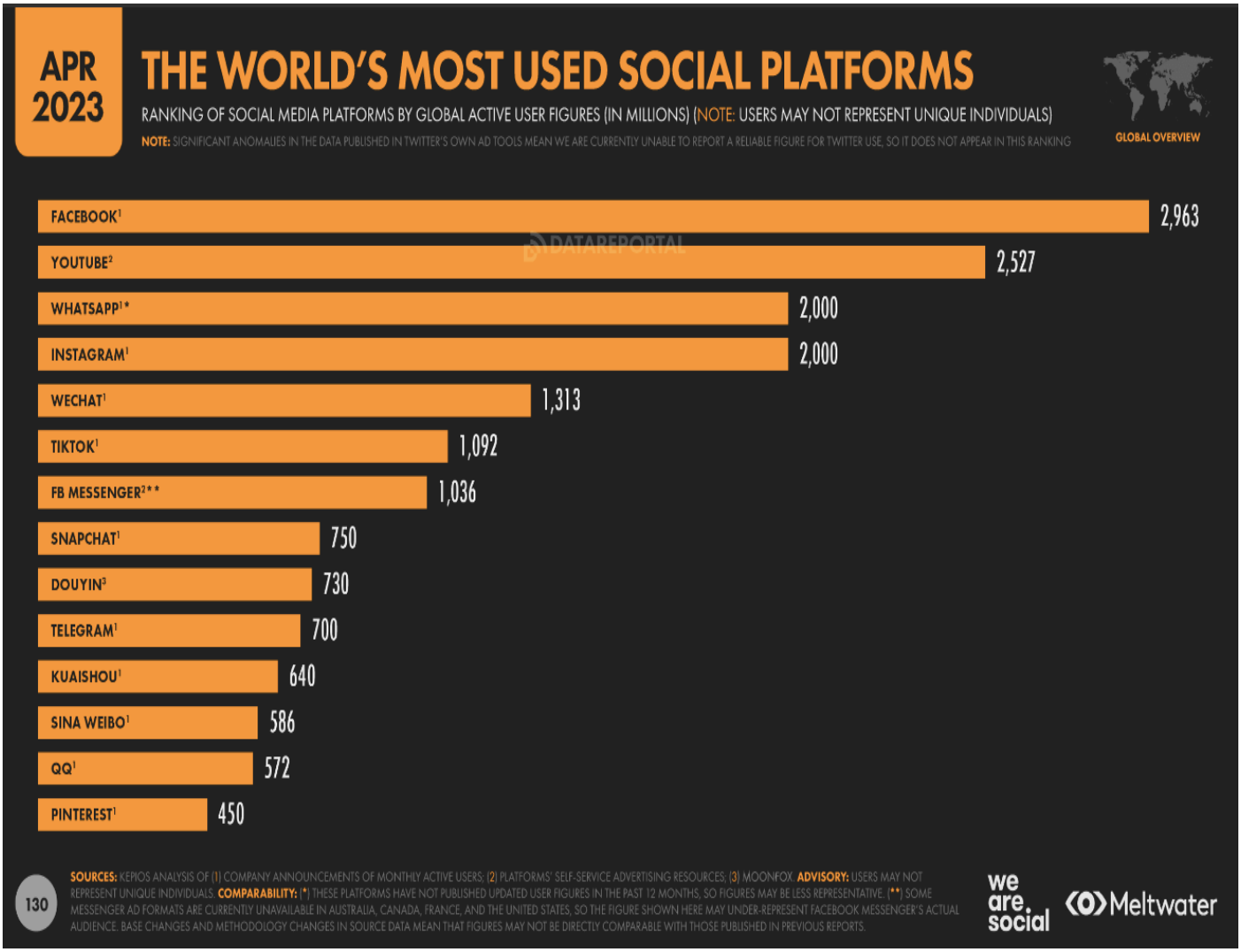

Fig 1 The World's Most used Social Platforms
(Source: Global Social Media Statistics)

There are many reasons for using social media platforms such as to get in touch with friends or family, spend spare time, entertainment, shopping, making new contacts, employment, etc. Countries like Libya, Mongolia, Philippines have the highest Facebook audience reach[6].[See Figure 2 below].

ISSN No:-2456-2165

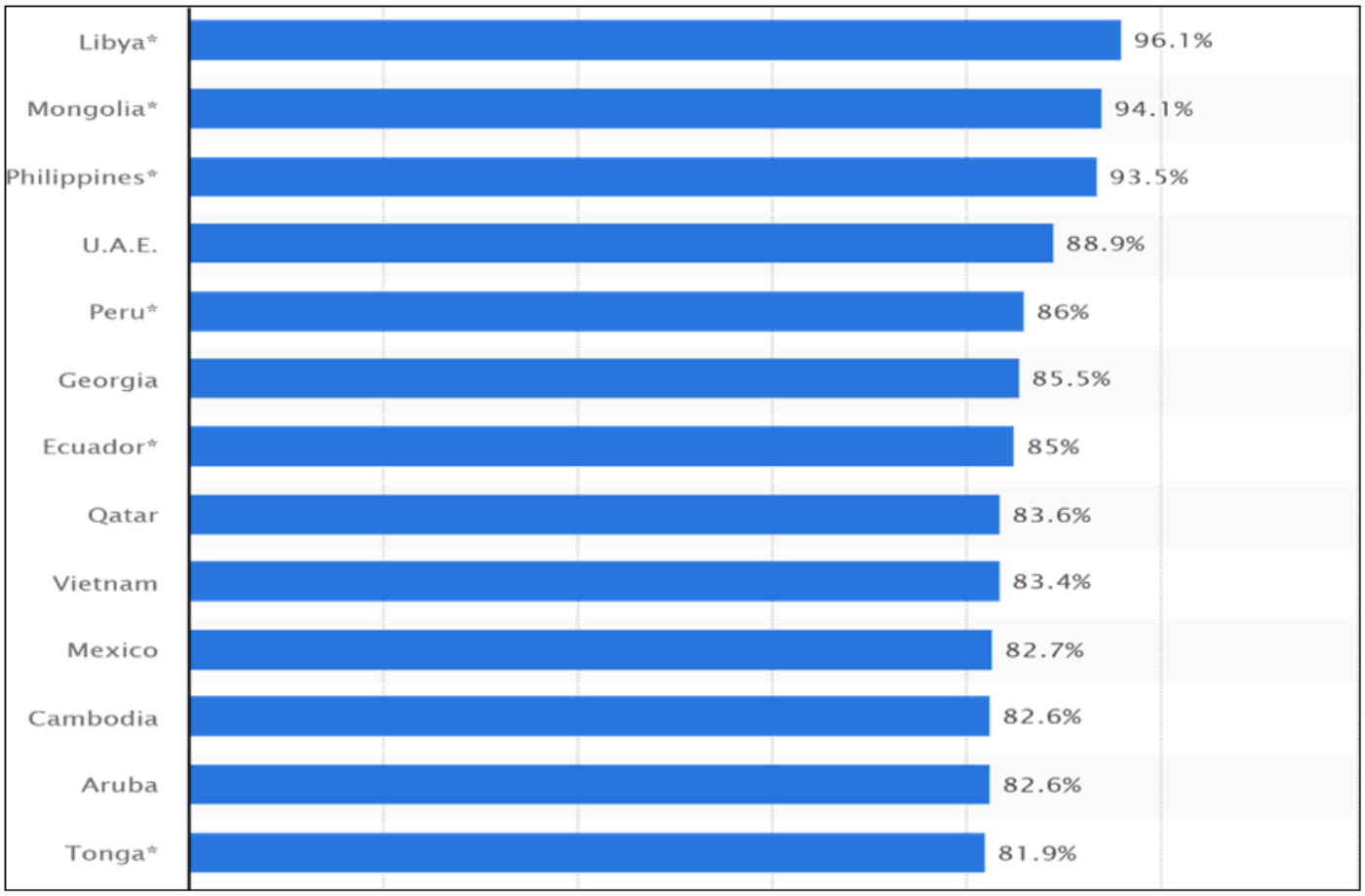

Fig 2 Countries with the Highest Facebook Audience Reach in January 2023
(Source: Social Media Statistics & Facts)

Facebook generated $116.8 billion in revenue in 2022. Approximately $65 billion came from the Facebook app[7]. Therefore, it is most significant, and utmost care is taken to protect the data or information of its users from data breaches. An organization is required to take efficient steps to give this effect. There are many threats such as insecure computer systems, ambiguity in the programming language used, vulnerabilities of private and public encryptions, traffic in the framework used to store data, etc[8]. All these threats have been resolved and precisely stated in the study based on different research conducted for this purpose.

II. METHODS FOLLOWED OR ADOPTED BY FACEBOOK META TO SECURE DATA OF ITS USERS

➢ *Java Script Code*

Many authors assume Cryptography perfectly retains the data secured unless their secret keys are made public. In formerly used methods by The Inc., it was assumed that hash functions and encrypted code deliveries were authentic and confidential. The sincere participants would patently desire that codes are transparent. However, fraudulent participants may want to present false evidence for the Javascript Code delivery that they have been delivered. To keep a check on this situation, the web server must be sincere. This renders the web server, application developer, and participants accountable and trustworthy to one another. There may be situations where fraudulent participants desire to retroact but the servers and data will be fully protected and secured since it is transparently maintained and publicly accessible.

To make this possible The Inc. uses transparent logs since, the transparency log is trusted, efficient, and available, and the same information is served to everyone. The data is maintained in the logs (the bulk of data), and all parties are satisfied that it is a public record and everyone has access to it. Therefore, these logs are transparent and anyone can see them or publicly accessible. A person can reach one web page to another via a hyperlink. Web pages are delivered via HTTP or HTTPS. In the latter case, “S” stands for secured and protected address. Thereby, ledgers are maintained to store the data for each URL and accountable. The users are required to declare that they trust The Inc. They can do this by *Asserting* to the hash algorithm (backed by a dash and encoded values) as intended or *Delegate* this responsibility to the third party or developer trust *Without Any Question (Blind Trust)* to the third party that is to say developer does not identify the code.

The initiator of the web browser is not recognized whereas the responder is recognized or identified since it has the public key and a certificate that links the public key to the main domain. The application developer builds content and makes it secure to the web server. The server transfers the code to the participant who can be anyone. However, the application developer and the website are associated with

the domain. Then, the participant requests a URL from the website[9].

➢ *Multi-Key Private Matching*

Data is stored in a remote server having confidentiality since no user has access to it however, there is still a possibility of leaking such data from the cloud storage[10]. The Inc. noted this case and uses multi-key datasets. Multi-key datasets include many to many relationships between the records. In other words, in a transaction of two parties, there is a possibility that the records match across the database such as email address and phone number. Major parties consist not only of a single identifier but several identifiers such as email addresses, phone numbers, and names in a single database whereas in another database there will be a combination of other identifiers and each identifier is important to the party. In previous private matching protocols, this was an onerous situation since it can match only a single type of identifier, for instance, a phone number. It was impracticable for such protocols to identify more than single identifiers where parties have more than one phone number or email address. To overcome this situation now, The Inc. uses Multi-key Private Matching for computing since it considers matching and not any other feature in the data. In multi-key private matching, both parties remain online which results in their secure joining. The Inc. came to this solution when they conducted research among Multi-key Private Matching for Compute, Delegated Private Matching for Compute (DPMC), and Compute, Delegated Private Matching for Compute with secure shuffling (DSPMC). On the contrary, this is not the case for DPMC and DSPMC since, in both these keys parties go offline and are required to encrypt their data. Moreover, both parties are required to carry the data and encrypted identifier which will be a cumbersome process. Therefore, Multi-key Private Matching is fundamental for computing in The Inc.

In Multi-key Private Matching, two parties generate a single set of keys for all their records without even revealing it to themselves. Both parties generate private keys to encrypt their associate data and once the party receives the data, they again encrypt it and that's how data is secured by a double encryption database. The Partner sends its doubly encrypted records to the Company after shuffling. The Company calculates a symmetric set difference, which would allow each party to get identifiers for records. Then, each party can generate an ID Spine by exchanging their encrypted records, undoing their shuffling, and attaching them to the records generated from the symmetric set difference. Therefore, this process restricts the number of possible matched records across the databases of the two parties[11].

➢ *Public Key Encryption in Easy Crypt*

It is well-versed that cryptographic construction has increased complexity in innovation and development. However, The Inc. found in some of the cases such cryptographic constructions were not accurate nor secured. Even, in some cases where such constructions are correct hitherto their application may be flawed. This significantly increases the application of computer-assisted cryptographic mechanisms partially. Therefore, In this case, cryptography is aided by the computer-aided mechanism for the application as well as the verification. These computer-assisted constructions are used to eliminate human labor used for the application and verification of cryptography. Such computer-devised systems patently increase the confidence of the users in cryptographic implementations in The Inc.

Moreover, such constructions have been engaged successfully at several platforms in constructions and verifications of the target. The Inc. uses Easy Crypt, Tamarin, and Jasmin for construction and verification proven as correct, effective, and efficient mannered for functionality. Easy Crypt espouses a code-based approach to provable security, modeling common security-related concepts, such as security notions and hardness assumptions, and well-defined possible programs. The tool required automated mathematical reasoning and other various built-in mechanisms. Therefore, The Inc. uses Saber's PKE scheme in Easy Crypt since, it envisages desired security (for security, this concerns the IND-CPA property), and correctness (for correctness, this concerns the δ-correctness). Furthermore, The Inc. taking the steps for construction and extension of Easy Crypt libraries (which can be found in the code or standard library in the Easy Crypt) having generic definitions and properties of multiple concepts that are used in the formal verification of Saber's PKE scheme along with the extension of formal libraries. Since these libraries are abstract and reusable can be used in the analysis of other cryptographic systems that use similar mathematical structures[12].

This ultimately serves the purpose of construction, verification, and implementation for making an informed decision simultaneously, increasing the confidentiality as to security among the users of The Inc.

➢ *SALSA Cryptographic System*

The Inc. at present uses a public-key cryptosystem which is at risk of cyber threats. Public-key cryptographic systems are unable to solve many problems such as distinct algorithms in the group, with their security. Thereby, the data of the users suffer from vulnerabilities. The study has been done between SALSA, lattice-based cryptography, and Learning With Errors (LWE)-based cryptosystems to know which protocol will be most suited to cryptosystem.

SALSA is a machine-learning attack on LWE-based cryptography. SALSA can fully recover secrets for small-to-mid-size LWE-based secrets. In lattice-based cryptography, the shortest vector problem and its indefinite variants are the hard mathematical problems that are not going to be addressed. The dots below form a lattice having vectors b1 and b2 however shortest vector is a green vector i.e., v. [See Figure 3 below]

ISSN No:-2456-2165

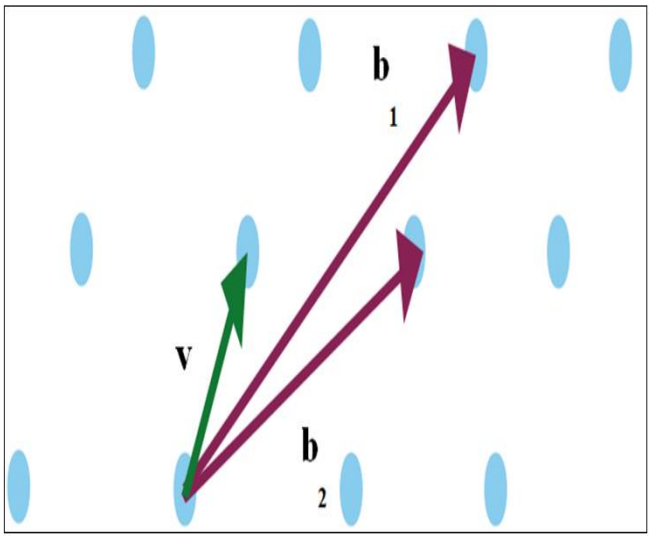

Fig 3 Showing Lattice-based Cryptography
(Source: SALSA: Attacking Lattice Cryptography with Transformers Research paper)

Application of an LWE-based cryptosystem is hard since it was hard to learn the secret vector, unlike machine learning where patterns can be learned. SALSA is a technique for secret-recovery Attacks on LWE by the chain of models. Meaning thereby SALSA attacks problems not addressed by LWE-based cryptosystems. SALSA contains a transformer model, a secret recovery algorithm, and a secret verification procedure. However, other parameters of full-strength homomorphic encryption such as secret density are within SALSA's reach[13].

Therefore, it was proposed that SALSA can break LWE problems. It was also shown that SALSA enables the recovery of secrets which was not the case in other cryptosystems, thereby it prevents any data breach. However, it was only a proposal for potential threats likely to arise in the future.

- *Differential Privacy*

All the technology giants employ means to safeguard the security of the data of their users. This can be done by preserving machine learning models that can be trained over the distributed data. There may be situations where such models can also leak the information of the users therefore, different types of models to accomplish the best results in preserving privacy and security are used by technological giants. The Inc. uses tree-based approaches i.e., Gradient Boosted Decision Tree (GBDT) models like XGBoost, LightGBM, and Cat-Boost, which are based on cryptographic mechanisms such as Homomorphic Encryption (HE) and Secure Multi-Party Computation (MPC). However, such mechanisms may not necessarily be an assurance of data privacy. GBDT methods are used for many reasons such as speed, ease of use, performance, etc.

Differential Privacy can be defined as the adoption of such mechanisms which can eliminate the risk of privacy breaches. Through differential privacy, it is possible to achieve very high utility while maintaining strong levels of privacy. In Differential privacy, the data inhabit the user device, and a small portion of information about the model update is accumulated from the clients. This minimizes the risk of breach of privacy. GBDT algorithm is the best balance of efficient privacy accounting and utility. GBDT algorithm has been experimented with based on five components by The Inc. team, these are Split Method, Weight Update, Split Candidate, Feature Interaction, and Batch Update. The dominant approach has been found in our iterative Hessian (IH) method for split candidates. However, other models are also privacy friendly which were proposed to be adopted in the future. Private GBDT framework methods can be adopted by incurring additional privacy costs[14].

- *Millisampler*

For efficient data security, traffic management is fundamental for every organization. This requires continuous monitoring. Traffic management must be done at short traffic dynamics i.e., millisecond time scale. The Inc. uses these two tools for this purpose: Millisampler and Syncmillisampler.

The existing data traffic center focuses on the utilization of links as well as the duration. The Inc. noted various contentions that occurred when various bursts arrived simultaneously shared at the framework. Buffering is generated when various bursts line up together at different chains and arrive simultaneously at the framework. This causes buffers in every chain which results in loss, latency in performance, etc. Millisampler is a lightweight host-centric approach, it corresponds to other traffic patterns. It collects all the data from different servers into a framework which in turn, synchronizes the data to ensure that the packets refined by the framework simultaneously appeared at another host. It specifies traffic over all the servers within the data framework at the same time. Millisampler helps in understanding burstiness, and workloads, identifying difficult traffic patterns, characterizing traffic at both fine time scales and large network scales, and troubleshooting the interactions between application behavior and the network. Through Millisampler it is easy to collect the data, store, and serve data at hosts, unlike switches which do not cooperate uniformly. The Inc. found out that there was a wide amplitude of buffering even during business hours in a single region. In a joint analysis, it was found that higher burstiness is not the reason for losses incurred. However, losses are most likely to occur with contended bursts that are a few milliseconds in duration.

Modern data center networks comprise distinct services and communication patterns. These patterns include capacity planning, fabric design, and tuning of transport parameters, all to provision an efficient network that provides low loss and latency to services as well as the dynamics of communication at millisecond-scale intervals. However, there is bursty traffic at these timescales. Buffer policies, control designs, and other server algorithms depend upon various factors including characteristics of traffic bursts[15].

III. FUNDING

A. The Present Manuscript has no Funding Source to Declare.

➢ *Availability of Data Material*

- https://unctad.org/tir2023
- https://ncses.nsf.gov/pubs/nsb20221/
- https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/#:~:text=According%20to%20the%20Datareportal%20April,within%20the%20last%2012%20months
- https://www.statista.com/statistics/278435/percentage-of-selected-countries-internet-users-on-facebook/
- https://www.businessofapps.com/data/facebook-statistics/

➢ *Ethics Approval and Consent to Participate*
Not applicable

➢ *Consent for Publication*
Not applicable

➢ *Declaration of Competing Interest*
The Author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1]. Conference on Trade and Development, U. (2023). *Technology and Innovation Report*. Geneva: United Nation Publication.
[2]. Tushar P. Parikh, D. A. (2017, June 6). *Raijmr*. Retrieved from https://www.raijmr.com: https://www.raijmr.com/ijrmeet/wp-content/uploads/2017/12/IJRMEET_2017_vol05_issue_06_01.pdf
[3]. Rosen, G. (2023). *Meta's Q1 2023 Security Reports: Protecting People and Businesses.* USA: Facebook Meta.
[4]. Board, N. S. (2022). *Science & Engineering Indicators.* USA: National Science Board.
[5]. Chaffey, D. (2023). *Global Social Media Statistics.*
[6]. (2023). *Social media - Statistics & Facts.* Stacy Jo Dixon.
[7]. Iqbal, M. (2023). *Facebook Revenue and Usage Statistics.*
[8]. Yunchuan Sun, J. Z. (2014). Data Security and Privacy in Cloud Computing. *International Journal of Distributed Sensor Networks*.
[9]. Ilkan Esiyok, P. B.-G. (2023, February 27). *Facebook*. Retrieved from https://research.facebook.com/: https://research.facebook.com/publications/accountable-javascript-code-delivery/
[10]. Wang, R. (2017). *ScienceDirect*. Retrieved from www.sciencedirect.com: https://pdf.sciencedirectassets.com/278653/1-s2.0-S1877705817X00052/1-s2.0-S1877705817302862/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEAaCXVzLWVhc3QtMSJIMEYCIQD56qHVnW24WkMRxwwHX4Ps3ghyAcFPhTi2CdW7n6gZlgIhALklnLvUVZ0yxDkRrTxLf%2B%2BfnHxXAW3xXIWekq%2
[11]. Prasad Buddhavarapu, B. C. (2023, January 30). *Facebook*. Retrieved from https://research.facebook.com/: https://research.facebook.com/blog/2023/1/delegated-multi-key-private-matching-for-compute-improving-match-rates-and-enabling-adoption/
[12]. Andreas Hülsing, M. M.-Y. (2022, August 12). *Facebook*. Retrieved from https://research.facebook.com: https://research.facebook.com/publications/formal-verification-of-sabers-public-key-encryption-scheme-in-easycrypt/
[13]. Emily Wenger, M. C. (2022, November 27). *Facebook*. Retrieved from https://research.facebook.com: https://research.facebook.com/publications/salsa-attacking-lattice-cryptography-with-transformers/
[14]. Samuel Maddock, G. C. (2022, November 9). *Facebook*. Retrieved from https://research.facebook.com: https://research.facebook.com/publications/federated-boosted-decision-trees-with-differential-privacy/
[15]. Ehab Ghabashneh, Y. Z. (2022, October 25). *Facebook*. Retrieved from https://research.facebook.com: https://research.facebook.com/publications/a-microscopic-view-of-bursts-buffer-contention-and-loss-in-data-centers/

➢ *Submission Declaration*
The work described has not been published previously or that it is not under consideration for publication elsewhere, and its publication is approved by the author and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

➢ *Author Bio*
Ms. Soniya is pursuing an LLB from **Law Centre II, Faculty of Law, University of Delhi** as well as pursuing company secretaries at the professional programme. She has done her graduation from the University of Delhi itself B.com (Hons.). Her areas of interest include corporate laws, Cyber laws, competition laws, and securities laws. She has written various articles such as Gist of Securities laws, Debt funding, and Equity funding which have been published on various reputed platforms and are available to read online. She has been honored with Legal Writer's Award by Legal Services India for outstanding legal research work in her article Sprouting of Cryptocurrency.