International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Inspired Symbol-based Authentication against Dictionary attacks Using ML Algorithm

[1]J.Balaraju, [2]C.Dastagiraiah, [3]K.Jyothi, [4]Voore Subrahmanyam
[1]Department of Computer Science & Engineering
Anurag University, Hyderabad, India
balarajucse@anurag.edu.in
[2]Department of Computer Science & Engineering
Anurag University, Hyderabad, India
dastagiraiah.cse@anurag.edu.in
[3]Department of Computer Science & Engineering
Anurag University, Hyderabad, India
jyothi.kurremula@gmail.com.
[4]Department of Computer Science & Engineering
Anurag University, Hyderabad, India
voore.subrahmanyam206@gmail.com

Abstract— Verifying an individual's identity before granting them access to a linked device, an internet service, or another resource is known as user authentication. The significance of authentication lies in its ability to safeguard data, apps, and networks for companies by limiting access to their protected resources to only authorized individuals or processes. This paper used the popular Big data technology Apache Spark for storing, and processing large data and proposed a novel authentication framework. A viable replacement for conventional alphanumeric passwords, bio-metric and facial authentications with dynamic symbol selection as an authentication. This authentication is tested in the Apache spark cluster which is the most distributed system. In these methods, SHA512 cryptography is used in several ways and comparison is done using existing authentication and machine learning algorithms. The straightforward authentication scheme and applied Apache Spark distributed system with 10 nodes produced the best results.

Keywords- Authentication, Distributed System, SHA512, Cryptography, Spark

I. INTRODUCTION.

Big Data [1] security in Apache spark [2] cluster security is the weakest link and is frequently regarded as a human factor. The development of secure systems, security operations, and human-computer interaction are the three main areas where these fac- tors are crucial. The authentication issue is the main concern here. In the majority of situations involving computer, and Big data security, user authentication is a key component. Authentication is the process of verifying that the person requesting a resource is who they claim to be. Most authentication solutions these days incorporate the use of a username and password. A password is a type of confidential authentication information that is used to manage data and resource access. It is kept secret from those who aren't authorized access, and those who want access are asked if they know the password before being either granted or refused access. The password's disadvantage is that it needs to be kept secret and that the user needs to remember it. Every authentication method has its own set of guidelines and limitations, such as the need for both alphanumeric and special characters in the password and a minimum length for the password. These passwords are primarily text-based. Passwords have been used since the beginning of time when warriors defending a spot would trade passwords and only let those who knew them inside[3].

A. Apache Spark

Using Hadoop's essential components, Apache Spark is an open-source framework for processing large amounts of data. It employs a multi-staged in-memory processing method that produces 100 times quicker processing than map-reduce processing when compared to other large data processing tools like Hadoop and Storm. Any Hadoop data source may be accessed and processed by Spark, which can operate on top of Hadoop clusters. Some fundamental features of task scheduling, memory management, fault recovery, and communication with storage systems are included in the core of Apache Spark. [4].

B. Cryptography

Cryptography [5] involves the use of mathematical algorithms and protocols to encrypt and decrypt data to prevent unauthorized access and ensure data integrity.

Cryptography is an essential technology used in symbol-based authentication (SBA) to ensure the confidentiality and integrity of the authentication process. In symbol-based authentication, cryptography is used in several ways. This is typically achieved using cryptographic hash functions, which convert the password or image template into a fixed-length hash value that can be stored or transmitted securely without revealing the original data. Cryptography is used to protect the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

communication between the client and server during the authentication process [6].

C. SHA512 Algorithm.

A hashing algorithm called SHA-512 operates on data that is provided to it. Blockchains, digital certificates, and internet security are all areas where hashing algorithms are applied. Hashing algorithms are so important to digital security and cryptography. It is a member of the SHA-2 family of hashing algorithms, which also includes SHA-256 and is used to hash the Bitcoin blockchain. Hence, SHA-512 completes its task in phases[7].

II. PROBLEM DEFINITION.

Nowadays most of the data reside in popular open-source Hadoop and Apache Spark distributed systems and it does not have their own securing methods and depends on third-party security protocols. Apache Spark uses Hadoop cores for storing and processing big data. Hadoop does not have its own security mechanism and it completely depends on third-party security mechanisms like Kerberos and Apache Knox. The existing authentication methods are time-consuming and an amount of password reuse and other authentication methods, not only personal accounts but also corporate accounts and, consequently, corporate data are in danger due to unauthorized accessing of data.

III. LITERATURE REVIEW.

Kiran, T. Srinivasa Ravi, et al [8] have put up a symbol-based entrance plan that defends against shoulder surfing attacks by utilizing the convex hull technique. In order to click inside the convex hull that these pass items make, the user must be able to identify pass objects. A huge collection of items can be used if the user wishes to make the password difficult to guess, however this will make the photos appear extremely crowded and the objects nearly indistinguishable from one another. If fewer items are used, the password space may be smaller, resulting in a big convex hull.

Balaraju and Prasada Rao et al [9] discussed that storage, security, and protection for Big data administrations are exceptionally thought-provoking. Rather than using any other big data technology, the Hadoop framework can offer the solution to the aforesaid problem. Over the past ten years, Hadoop's security features have rapidly increased with each new iteration. In this paper's second half, Hadoop and Spark continue to face Issues with authentication, metadata security, and data leaking.

Gao, Haichang, Wei Jia, Fei Ye, and Licheng Ma. et al [10] has created a graphical password method in which the user clicks on roughly specified areas on a given image to access pre-determined destinations. By enabling the user to click on different items in the right order to verify their legitimacy, Passlogix expanded on this concept.

Balaraju and Prasad Rao et al [11] have spoken about the security measures for Hadoop and Spark authentication. They employ significant computational overhead third-party security providers like Kerberos and Apache Knox to secure data. Despite computational expense, this security method does not succeed in establishing a trustworthy environment. In addition, they only authenticate users—not their processes. The authors also discussed how to secure Big Data in a Hadoop/SPARK c cluster using Secure Authentication Interface, a single, unique security method.

Meng, Yuxin. et al [12] have created a method that uses a mouse to draw a digital signature for authentication. The method consists of two steps: the first is registration, and the second is verification. The stage user signs his name using a mouse during registration, and the system then extracts the signature area. It accepts the user signature as input, goes through the normalization procedure, and then extracts the signature's parameters at the verification step.

Balaraju and Prasada Rao et al [13] suggested New Algorithm Authentication Based on Access (BABA) as a security instance connected with a Hadoop instance to protect metadata security from Hadoop failures and protect data in the Hadoop cluster from attackers. With computing power increasing data security and offering a solid solution for HC, this mechanism offers a safe hybrid cloud (HC) without the need for additional security measures, saving operating costs.

IV. MATERIALS AND METHODS.

By separating master and slave services, the Spark Multi-node Cluster can handle up to 16000 nodes, and a large number of users may operate within the cluster. With the aid of master services, every user may join the cluster with identical rights, and this service does not require user authentication..

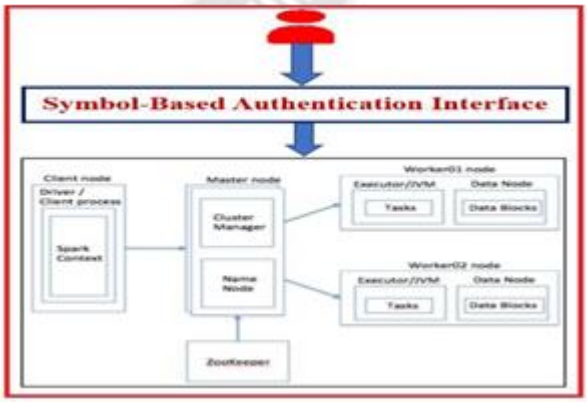

Fig 1: Proposed Symbol-based Authentication for SPARK Cluster.

The multi-node Hadoop/Spark cluster setup was established using the below configurations: Two servers with Xeon processors, 32 GB RAM, and 256 × 4 GB SAS controller hard

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

disk capacity as masters; Core i3 processor, 4 GB RAM, and 512 MB storage with 117 numbers; Core i5, 8 GB RAM, and 1 GB HDD with 126 numbers as slaves. The cluster has 245 nodes total, all of which are running CentOS Linux, an open-source operating system, and are connected to a fast gigabit switch. This setup offers a conducive setting for upcoming performance analysis and related endeavors. The graphical interface is designed as a symbol-based authentication mechanism that uses images, graphics, or visual cues instead of traditional alphanumeric passwords to authenticate a user's identity. Instead of typing in a password, the user selects a series of images, clicks on specific points, or traces a pattern using a mouse or touchscreen. Symbol-based authentication is considered to be more secure than traditional alphanumeric passwords because images and patterns are more difficult to guess or crack using brute-force methods. Additionally, it is believed to be more memorable than traditional passwords, making it easier for users to recall their authentication information.

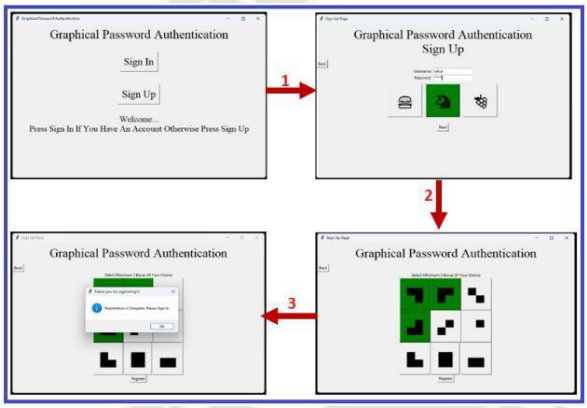

Fig 2: Process of User registration using Symbols.

There are several types of symbol-based authentication methods, including im-age-based, gesture-based, and 3D virtual environment-based authentication. Some examples of image-based authentication include Pass faces, which involves selecting faces from a group of images, and DAS, which involves selecting a series of digits from a grid. Finally, 3D virtual environment-based authentication is a newer and more experimental method that involves selecting and interacting with objects in a virtual environment to authenticate a user's identity. Overall, symbol-based authentication offers a promising alternative to traditional alphanumeric passwords and has the potential to provide stronger security and a better user experience[16].

A. Recognition- Based Technique

With recognition-based tactics, a user is made to prove their identity by asking them to recognize one or more photos they chose during the registration process. For in-stance, in recognition systems, the system needs to remember which photographs are from a user's portfolio in order to show them. This information must be saved such that the system may access it in its original form (perhaps protected by reversible encryption), and anybody with access to the stored data may also have access to it such as a shoulder surfing assault or a phishing attempt.

B. HASHLIB Module

This module's main goal is to encrypt a string using a hash function in a way that makes it nearly impossible to decrypt. The length of an encrypted string makes it nearly hard to recover the original string most of the time. A common interface for several secure hash and message digest algorithms is created by this module. Included are the FIPS secure hash algorithms (specified in Internet RFC 1321) and the MD5 technique from RSA. The FIPS secure hash algorithms include SHA1, SHA224, SHA256, SHA384, and SHA512. It is possible to use "message digest" and "secure hash" interchangeably. Message digests were a term used for earlier algorithms. Secure hash is the word used nowadays[19].

Fig 3: Symbol-based Authentication Login process.

V. RESULTS AND DISCUSSION.

In the existing system, to provide security for the Apache spark cluster Kerberos and Apache Knox authentication mechanisms are used. The developed authentication method is applied for the Apache Spark cluster which stores large sets and processing tools by using reduced computation and provides more security than existing Authentication methods. The comparative study is done with symbol-based authentication and other authentication like Kerberos, Secure Authentication interface and DNA Authentication [20].

Authentication Methods	Computations per Authentication
SAI	5
DNA	4
Kerberos	6
SBA (Proposed)	1

Table 1: Single User Authentication and its Computations.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

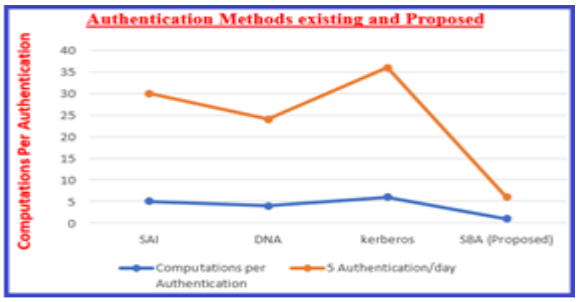

Figure 4: Single User Authentication and its Computations.

Authentication Methods	Computations per Authentication	5 Authentication /day
SAI	5	25
DNA	4	20
Kerberos	6	30
SBA (Proposed)	1	5

Table 2: Single User Authentication, its Computations, and 5 times authentication/day

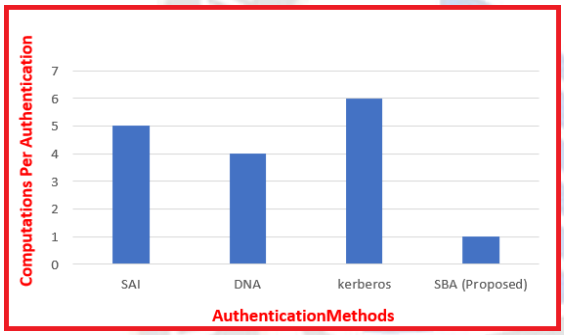

Figure 5.: Single User Authentication, its Computations, and 5 times authentication/day.

VI. CONCLUSION

The proposed a framework for a fast and efficient authentication system for securing Big data in the Apache Spark cluster. It is an efficient alternative to existing authentication methods. The system's successful completion of its goal was to build a symbol-based authentication system applied to popular big data technologies of Apache spark cluster for restricting unauthorized users to access data and this can use in any mobile or web-based application. In this authentication SHA512 cryptography is used in several ways. They are appealing since words are often forgotten more slowly than images. The straightforward authentication scheme and applied Apache Spark distributed system with 10 nodes produced the best results.

VII. FUTURE SCOPE.

The future work of this is applying for Hadoop and spark multinode cluster with the size of 1000+ nodes distributed system because the above clusters are supporting commodity hardware and increasing nodes supports day by day. This mechanism is also applying for data accessing as future work by providing a unique id for each user and storing the user activities.

REFERENCES.

[1]. Balaraju, J., Prasada Rao, PVRD., –Recent advances in big data storage and security schemas of HDFS: a surve, Special Issue (Emerging Trends in Engineering Technology) ,6,132-138.

[2]. Chellappan S., Ganesan D. (2018) Introduction to Apache Spark and SparkCore. In: Practical Apache Spark.Apress, Berkeley,CA. https://doi.org/10.1007/978-1-4842- 3652-9_3.

[3]. M. M. Shetty and D. H. Manjaiah,(2016), "Data security in Hadoop distributed file system", International Conference on Emerging Technological Trends (ICETT),1-5.

[4]. Salloum, S.; Dautov, R.; Chen, X.; Peng, P.X.; Huang, J.Z. Big data analytics on Apache Spark. Int. J. Data Sci. Anal. **2016**, 1, 145–164.

[5]. Amalraj, A.J. and Jose, J.R., 2016. A survey paper on cryptography techniques. International Journal of Computer Science and mobile computing, 5(8), pp.55-59.

[6]. Revenkar, P. S., Anisa Anjum, and W. Z. Gandhare. "Secure iris authentication using visual cryptography." arXiv preprint arXiv:1004.1748 (2010).

[7]. Gueron, Shay, Simon Johnson, and Jesse Walker. "SHA-512/256." In 2011 Eighth Interna- tional Conference on Information Technology: New Generations, pp. 354-358. IEEE, 2011.

[8]. Kiran, T. Srinivasa Ravi, et al. "A symbol-based graphical schema resistant to peeping at- tack." International Journal of Computer Science Issues (IJCSI) 10.5 (2013): 229.

[9]. Balaraju J., Prasada Rao. PVRD, –Innovative Secure Authentication Interface for Ha- doop Cluster Using DNA Cryptography: A Practical Study‖. In (eds) Soft Computing and Signal Processing. ICSCSP 2019. Advances in IntelligentSystems and Computing, vol 1118. Springer, Singapore.https://doi.org/10.1007/978-981-15-2475-2_3

[10]. Gao, Haichang, Wei Jia, Fei Ye, and Licheng Ma. "A survey on the use of graphical pass- words in security." J. Softw. 8, no. 7 (2013): 1678-1698.

[11]. Balaraju, J., Prasada Rao. PVRD, "Investigation and Finding A DNA Cryptography Layer for Securing Data in Hadoop Cluster." Int. J. Advance Soft Compu. Appl 12.3 (2020).

[12]. Meng, Yuxin. "Designing click-draw based graphical password scheme for better authenti- cation." In 2012 IEEE Seventh International Conference on Networking, Architecture, and Storage, pp. 39-48. IEEE, 2012.

[13]. Balaraju, J., Prasada Rao. PVRD, –Designing authentication for Hadoop cluster using DNA algorithm‖. Int. J. Recent. Technol. Eng. (IJRTE) ,8(3), 2019. ISSN: 2277-3878. https://doi.org/ 10.35940/ijrte.C5895.0983.

[14]. Li, Yue. "On Enhancing Security of Password-Based Authentication." (2019).

[15]. Jiya, Gloria Kaka, Ishaq Oyebisi Oyefolahan, and Joseph O. Ojeniyi. "Recognition based graphical password algorithms: A survey." (2021).

[16]. Saranya Ramanan1, Bindhu J S "a survey on different symbol-based authentication schemes" IJIRCEVol. 2, Issue 12, December 2014. [6] Greg E. Blonder (1996). U.S. Patent No.5559961.

[17]. Towseef Akram, Vakeel Ahmad, Israrul Haq, & Monica Nazir. (2017). Symbol-based Authentication.

[18]. Awais, A., Muhammad, A., M., K. H., & Talib, R. (2016). Secure Symbol-based Tech- niques against Shoulder Surfing and Camera-based Attacks.

[19]. Anwar, Muhammad Rehan, Desy Apriani, and Irsa Rizkita Adianita. "Hash Algorithm In Verification Of Certificate Data Integrity And Security." Aptisi Transactions on Techno- preneurship (ATT) 3.2 (2021): 181-188.

[20]. Balaraju, J.,Prasada Rao. PVRD, –A Novel Node Management in Hadoop Cluster using DNA‖, International Journal of Information Technology ProjectManagement.,12(4), June 2021, ISSN: 1938-0232.