International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Dynamic Password to Enforce Secure Authentication Using DNA.

J. Balaraju*[1], P. Ravinder Rao[2], V. Biksham[3] , P. V. R. D. Prasada Rao[4], Praveen Tumuluru[5]

Submitted: 16/08/2023 **Revised**: 07/10/2023 **Accepted**: 19/10/2023

Abstract: Every user accesses and stores data from the Hadoop cluster via a different node and a different node service, and they are all concerned with security and privacy to keep sensitive data safe from hackers. Password-based authentication is one of the many authentication solutions that are available to secure sensitive data. Both password security and authenticity have become crucial because of the quick growth of data storage, access, sharing, and moving. However, it is also necessary to guarantee the password's strength. Therefore, all cyber experts advise using complex password patterns. In this study, we create a special DNA algorithm that creates a special ID and a dynamic password based on the user's data. Through numerous experiments with single-node and multi-node Hadoop / SPARK clusters and varied users, we have evaluated our system. Our testing shows that the new authentication method we developed and the generated passwords are both completely reliable. We also put our created passwords to the test against the dictionary attack and the brute force attack, two methods for breaking passwords. By creating a user interface for a vast distribution system, we expanded our task. Our algorithm creates passwords that are easy to use and have the majority of the characteristics of being strong and uncrackable.

Keywords: *Hadoop Cluster, Authentication, DNA Cryptography, Security, Privacy.*

1. Introduction

Big Data [1] has recently been a hot topic for storage, analysis, and security. For IT services with a high level of efficiency and effectiveness, high end computer resources are needed. Because of the increased number of internet users and smartphone users as well as the widespread use of social networks and sensor networks, there is a growing need for data in the digital world. A collection of Semi-structured, unstructured data, and structured collections is referred to as big data. Data has grown from Exabyte's to Zettabytes and Yottabytes in size. Numerous difficulties are created by the quick growth rate of such vast amounts of data including the rapid creation of data, transmission speed, diversity of data, and security [2].

2. Hadoop Cluster

By leveraging commodity hardware, Hadoop Clusters (HC)[3] offers huge storage of data and analysis of large unstructured data in a distributed setting with parallel processing. Whether big or small, data organizations have stringent rules for allowing users access to the cluster's data. In the current digital era, that user community is rapidly expanding to encompass clients, partners, and joint venture companies. To secure the HC, authentication, and metadata security can work together [4].

3. Kerberos Authentication

Cloudera (CDH)[5], Hortonworks (HDP)[6], and MapR[7] are three Hadoop distributions that use Kerberos extensively for user and service authentication. Because Kerberos is a session-based authentication mechanism, additional steps were necessary to demonstrate one user for one session. The other drawback of Kerberos is that it completely relies on the KDC and employs timestamps to create secure communication [8]. Servers must therefore be synchronized within minutes; for example, they frequently need to synchronize their clocks. The main problem with Kerberos is that a malevolent user might implement authentication if the ticket was stolen or duplicated [9]. In any distributed systems authentication, Kerberos authentication[10] is more crucial for both user and service. If the Kerberos ticket is lost, a hostile client could compromise the authentication. High dependence on the Kerberos server (KDC) is one of Kerberos' limitations. To establish secure communication, it preserves timestamps. Password guessing attacks are not effectively thwarted by it, and Kerberos by itself cannot ensure that the password won't be transmitted across the network [11].

[1] *Department of CSE, Anurag University, Hyderabad – 500088,India*
ORCID ID : 0000-0002-6025-7524
[2] *Department of CSE,Anurag University, Hyderabad – 500088,India*
ORCID ID : 0000-0003-0354-3740.
[3] *Department of CSE, Anurag University, Hyderabad – 500088,India*
ORCID ID : 0000-0003-0532-594X.
[4]*Department of CSE , KLEF,Vijayawada – 522302, India*
ORCID ID : 0000-0001-7527-8013
[5]*Department of CSE , KLEF,Vijayawada – 522302, India*
ORCID ID : 0000-0001-9426-205X
** Corresponding Author Email: jb7443@gmail.com*

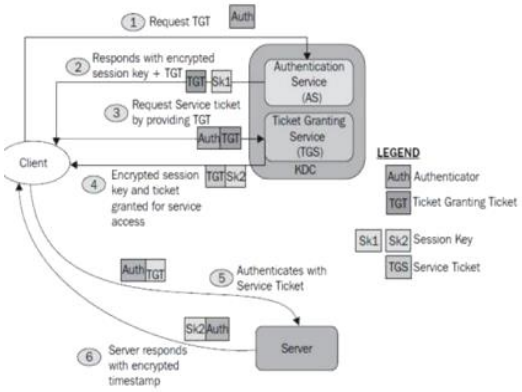

Fig 1: Kerberos Authentication Process**.**

4. Problem Definition

90% of customers are using unprotected open-source HC, according to recent polls. Kerberos and Apache Knox are the authentication security methods used in Hadoop data centers. Third-party security mechanisms cannot match the performance of Hadoop and Spark data centers. The Kerberos limitations are identified, the KDC's (Kerberos server) high reliance, To establish secure communication, it preserves timestamps. Password guessing attacks are not effectively thwarted by it, and Kerberos by itself cannot guarantee that the password won't be transferred over the network. The process of confirming and establishing a user's identity through authentication is necessary to develop a secure Hadoop cluster. Based on the user's identification, many authentication solutions have been introduced recently.

5. DNA.

The most alluring, appropriate, and intriguing field of study to achieve data security by leveraging a complex DNA structure are deoxyribonucleic acid (DNA)[12] based data hiding and encryption techniques. Compared to the DNA concealment mechanism, the conventional information security system requires a huge investment. Many scientists are interested in using the intricate structure of DNA as data security, which can also be used to secure sensitive metadata in HC. Utilizing DNA data concealing [13] techniques has several benefits, including the extreme difficulty of getting a key for an attacker and the complexity of comprehending the hiding process. Sensitive data must be secured in a distributed network, which DNA can do more effectively than other security measures [14].

6. Literature Survey.

B. Saraladevia et al [15] have analyzed HDFS security issues. The Hadoop Architecture's foundational layer, HDFS, has many information game plans and is more vulnerable to security issues. The three security and privacy measures below are based on HDFS. Any of the following ways, or a combination of these three approaches in HDFS, can advance big data security.

Kumar Abhishek et al[16] Integrated Hadoop Cloud Framework (IHCF), a new layer that supports numerous frameworks available in Hadoop ecosystems as cloud services, including Hive and Pig, was introduced. IHCF can be used by the average person to use Hadoop and will help to handle it more effectively and suitably. With the use of Hadoop ecosystems and other systems like IHCF, Pig, and Hive, massive amounts of data may be processed accurately and efficiently.

Michael Kenyeba et al [17] difficulties with authentication for Hadoop security on the cloud are explored. They mainly concentrated on security challenges related to Hadoop authentication for storing data on huge clusters of commodity hardware. They also talked about several Hadoop security and authentication concerns and potential solutions. Along with learning about computer security, network security, data security, and privacy, they are also learning about the significant advancement of Hadoop security for protecting data. Because the majority of the enterprise-critical data is kept in the cluster, access security becomes a top priority.

Yoon-Su Jeong et al [18] The authors covered both open-source and private platforms for developing big data applications. To process the security of Big Data, they are concentrating on various security methods in the Apache Hadoop framework. Hadoop's distributed file system, on which its modules are based, does not have a strong user authentication security system. The data node can use a block access token to authenticate HDFS clients under the proposed scheme. For higher communication, computation, and area efficiency than current HDFS systems, it leverages hash chain keys, in contrast to most HDFS authentication protocols that adopt public key exchange methodologies. By utilizing elliptic curve cryptography, this design strengthens the current symmetric key HDFS authentication. The produced authentication keys are made anonymous by the ECC technique used in this proposed approach, shielding them from breaches or unintentional exposures.

7. Implementation of DNA Algorithm.

The Importance of the proposed system is primarily diversely focused on user authentication with the combination of static and dynamic keys produced by DNA cryptography. Every user should be registered to enter HC, and only registered users can perform their activities in HC. The HC Interface takes the unique property of the user's mail ID and generates the first part of the unique key using the DNA hiding system and its corresponding mail ID. Once the static key is with the user, they can enter into HC as fixed as well as password for the

first time. The second part of the dynamic key is generated at the time the user logout, and it is caused by collecting the number of nodes accessed date, and time of the logout session. All the static and dynamic keys[19] are sent to the users' mail IDs at every logout time, and the user can log in with the static and dynamic keys from the second time onward [20].

Static Key Generation using mail_id Algorithm

Procedure

mail_to_Static_Key(email)

for *k=1 ….strlen(email)*

ascii_value[i]=email[i];

binary_format[i]=ascii_value[i];

DNA_format[i]=binary_format[i]

av[i]=decimal_value[i]

static_key[i]=decimal_value[i]

end for

return(Static_key)

end procedure

Static key Validation reverse – algorithm

Procedure

Static_Key_to_email(U_key)

for k in 1 …. strlen(Static_key)

decimal_value=static_key[i]

av[i]=decimal_value[i]

DNA_format[i]=av[i]

Binary_format[i]=DNA_format[i]

Ascii_value[i]=binary_format[i]

email[i]=ascii_value[i]

end for

return(email)

end procedure

8. Experimental Results

Since there are 6 distinct operations involved in the conventional Kerberos authentication process, it requires 6 units of time to complete. The following generalization applies to the computation time of the Kerberos authentication procedure: [21]

$$T(KA) = \sum_{i=1}^{U} [(6t) + x + y]$$

Where T(KA) stands for the total amount of time it takes to process user authentication requests (U).

- t= time stamp taken into account during authenticating.
- x=excess time for authentication optimal timestamp_t‘.
- y=time taken at AS

Table 1. Steps for Authentication in Kerberos.

Message No.	Message Exchange
1	$C \rightarrow AS: ID_C \| ID_{tgs} \| TS_1$
2	$AS \rightarrow C : E(K_c, [K_{c,tgs} \| ID_{tgs} \| TS_2 \| Lifetime_2 \| Ticket_{tgs}])$ $Ticket_{tgs} = E(K_{tgs}, [K_{c,v} \| IDC \| ADC \| IDv \| SN4 \| Lifetime4])$
3	$C \rightarrow TGS : ID_v \| Ticket_{tgs} \| Authenticator_c$ $Authenticator_c = E(K_{c,tgs}, [ID_C \| AD_C \| TS_3])$
4	$TGS \rightarrow C: E(K_{c,tgs}, [K_{c,v} \| ID_v \| TS_4 \| Ticket_v])$ $Ticket_v = E(K_v, [K_{c,v} \| ID_C \| AD_C \| ID_v \| TS_4 \| Lifetime_4])$
5	$C \rightarrow V : Ticket_v \| Authenticator_c$
6	$V \rightarrow C : E(K_{c,v}, [TS_5 + 1])$

Tickets, which are used to authenticate a client to the ticket distribution center, and an authenticator, which verifies the client's identity to the server of services—the ticket distribution center serves as an intermediary of various requests exchanged between the client and server of services to authenticate the client before access to the desired service [22].

Results Analysis-1

Table 2. Computations in Kerberos for 5 users.

	Kerberos	Computations
U1	**1**	**6**
U2	**2**	**12**
U3	**4**	**24**
U4	**1**	**6**
U5	**3**	**18**

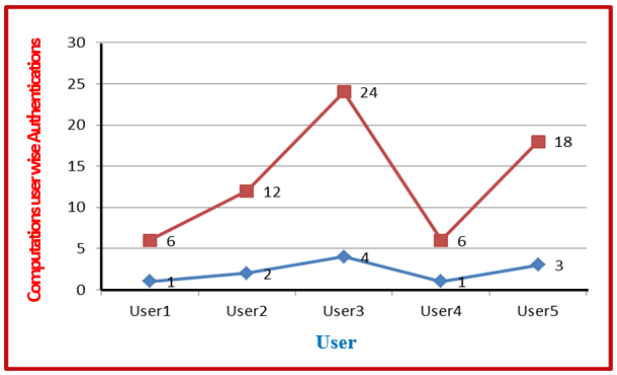

Fig 2. User computations analysis in Kerberos for five users.

Figure 2 from the above, provides a graphical breakdown of five users' Kerberos authentication. The y-axis displays the number of computations needed for authentications, and the x-axis lists usernames with the number of authentications. For instance, user2 successfully authenticated twice using 12 calculations.

Table 2. Five Users and Computations in DNA Algorithm.

	DNA_A	**Computations**
U1	1	5
U2	2	10
U3	4	20
U4	1	5
U5	3	15

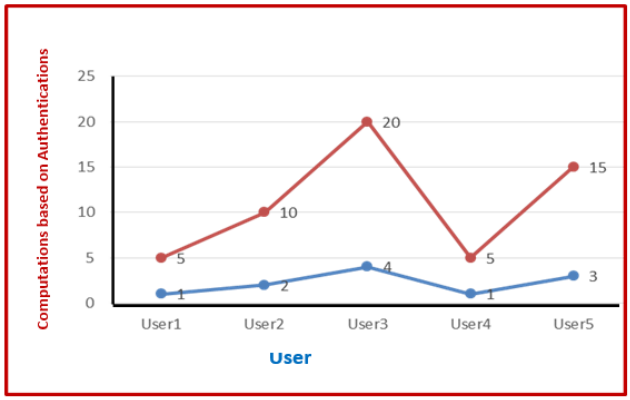

Fig 3. User computations analysis in DNA Algorithm for five users.

The aforementioned graph illustrates the authentication of 5 users using the created Secure Authentication Interface. The y-axis displays the number of computations needed for authentications, and the x-axis displays the username with the number of authentications. Let's say the user3 successfully authenticated four times using 20 calculations.Five separate processes, each of which can be completed in one unit of time, were taken into consideration when developing our Secure Authentication system. The developed secure authentication method's computation time is as follows.

$$T(_{DNA\_A}) = \sum_{i=1}^{U} (5)$$

Where T (DNA_A)=Total Computation of Authentication

(U) = User requests in the secure authentication method

Every time you leave the Hadoop cluster, a dynamic key is mailed to you. Therefore, the computation does not take into consideration the use of a timestamp at every step. Additionally, this technique does not make use of the Active Server (AS).

8.1. Results Analysis-2.

Table 4. Comparison of User Computations in Kerberos and DNA for 10 users

	Kerberos (Existing)	**Computations**	**Total Authentications / Day (DNA_A)**	**Total Computations**
User1	5	30	5	25
User2	**10**	**60**	**10**	**50**
User3	**8**	**48**	**8**	**40**
User4	6	36	6	30
User5	**2**	**12**	**2**	**10**
User6	3	18	3	15

User7	7	42	7	35
User8	6	36	6	30
User9	3	18	3	15
User10	**2**	**12**	**2**	**10**
Total	**52**	**312**	**52**	**260**

Fig 4. User computations analysis in DNA Algorithm for Ten users.

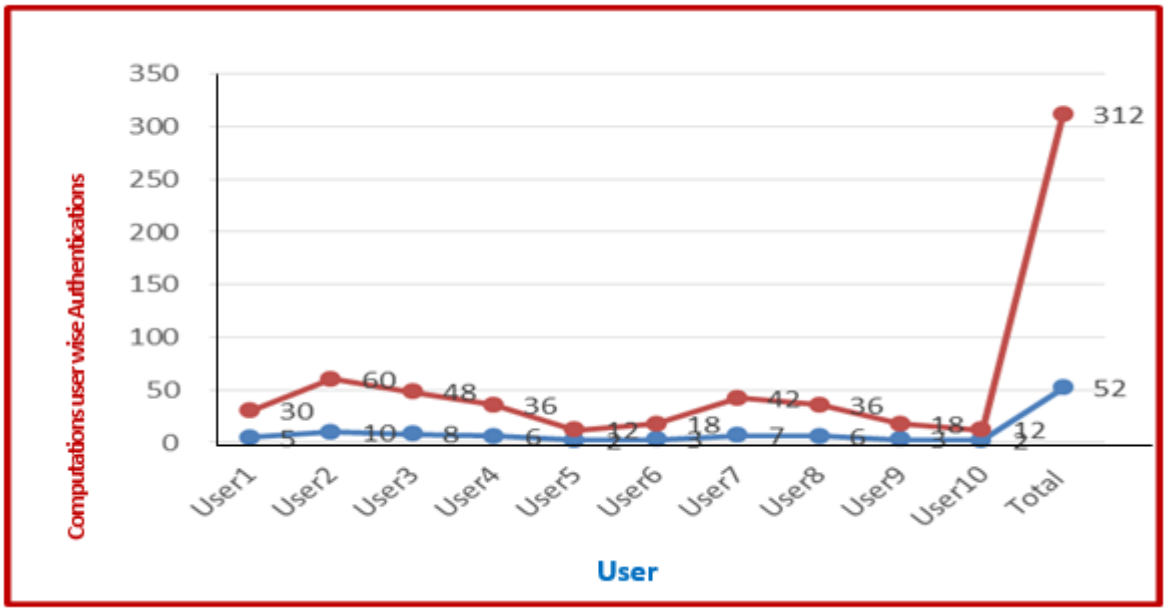

Figure 4 is an illustration of the authentication of 10 users using the current Kerberos authentication system. The y-axis displays the number of computations needed for authentications, and the x-axis displays the username with the number of authentications. All 10 users are authenticated 52 times while utilizing 312 computations, starting with user 2 who was authenticated 10 times using 60 computations.

Performance Analysis

Since there are 6 distinct operations involved in the conventional Kerberos authentication process, it requires 6 units of time to complete. The following generalization applies to the computation time of the Kerberos authentication process.

$$T(KA)=U[(6t)+x+y]$$

Five separate processes that can each be completed in one unit of time were taken into consideration when we designed DNA_A. The designed secure authentication mechanism has the following computation time.

$$T(DNA\_A)=U(5)$$

Every time you leave the Hadoop cluster, a dynamic key is mailed to you. Therefore, the computation does not take into consideration the use of time stamps at each step. Additionally, this technique does not make use of the Active Server (AS).

9. Conclusion and Future Scope.

The DNA_A algorithm developed is suitable for Hadoop/SPARK clusters and distributed systems, and it employs a novel authentication mechanism with a dynamic password mechanism for each time with limited computation. It is necessary required for develop a novel authentication system for both Hadoop / Spark clustersfor securing data. Finally, the DNA_A provides a good solution for increasing Hadoop and Spark cluster security and performance while decreasing elapsed time. The goal of this work in the future is to assign a dynamic hostname to each node in Hadoop and Spark clusters, which will confuse hackers and ultimately provide more security. This algorithm's computational burdens must be reduced before it can be applied to different domains.

References:

[1] Ishwarappa, J. Anuradha (2015), A Brief Introduction on Big Data 5Vs Characteristics and HadoopTechnology, Procedia Computer Science, Vol. 48, 319-324.

[2] P. Sirohi, A. Agarwal and S. Tyagi(2016), A comprehensive study on attacks on SSL/TLS protocol,2nd International Conference on Next Generation Computing Technologies (NGCT), 893-898.

[3] Can Uzunkaya, Tolga Ensari, Yusuf Kavurucu (2015), Hadoop Ecosystem and Its Analysis on Tweets, Procedia - Social and Behavioral Sciences,195,1890- 1897.

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

[5] Menon, R. (2014). Cloudera Administration Hand book Horton Works Data Platform HortonWorks Data Platform: New Book. (2015).

[6] Dunning, T., & Friedman, E. (2015). Real-World Hadoop.

[7] B. C. Neuman and T. Ts'o,(1994), "Kerberos: an authentication service for computer networks," IEEE Communications Magazine, 32, 9, 33-38.

[8] 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.

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

[10] Eman El-Emam, Magdy Koutb, Hamdy Kelash and Osama S. Faragallah, (2011),‐An Authentication Protocol Based on Kerberos 5‖,International Journal of Network Security, 12-2,147–158.

[11] Balaraju J., Prasada Rao. PVRD., ‐Dynamic Node Identification Management in Hadoop Cluster Using DNA‖. In Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies, vol 224. Springer,Singapore. https://doi.org/10.1007/978-981-16-1502-3_9.

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

[13] sMarwan, S., Shawish, A., Nagaty, K.: DNA-based cryptographic methods for data hiding in DNA media. In: ScienceDirect, pp. 110–118, Aug 2016

[14] Hossain, E.M.S., Alam, K.M.R., Biswas, M.R., Morimoto, Y.: A DNA cryptographic tech nique based on dynamic DNA sequence table. In: International Conference on Computer and Information Technology, IEEE 2016, Dhaka, Bangladesh, pp. 270–276, 18–20 Dec 2016.

[15] B. Saraladevia, N. Pazhanirajaa, P. Victer Paula, M.S. Saleem Bashab, P. Dhavachelvanc (2015),―Big Data and Hadoop-A Study in Security Perspective‖ ,Elsevier.

[16] Abhishek K., Kumar Verma, M., Shivam K., Kumar V., Mohan A.,(2017), Integrated Hadoop Cloud Framework (IHCF). Indian Journal of Science and Technology, 10,1-8.

[17] B. Agrawal, R. Hansen, C. Rong and T. Wiktorski,(2016),"SD-HDFS: Secure Deletion in Hadoop Distributed File System," , IEEE International Congress on Big Data (BigData Congress), 181-189.

[18] H. Ye, X. Cheng, M. Yuan, L. Xu, J. Gao and C. Cheng,(2016), "A survey of security and privacy in big data", 16th International Symposium on Communications and Information Technologies (ISCIT),268-272

[19] Hojabri, Mehdi.(2013), "Innovation in cloud computing: Implementation of Kerberos version5 in cloud computing in order to enhance the security issues." Information Communication and Embedded Systems (ICICES), International Conference on. IEEE.

[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

[21] T. A. Khaleel,(2020), "Review of Network Authentication Based on Kerberos Protocol," Journal of Basic Education college,16,1141- 1150.

[22] Kumari, Aruna, and Dharmender Singh Kushwaha. "Kerberos style authentication and authorization through CTES model for distributed systems." Computer Networks and Intelligent Computing. Springer, Berlin, Heidelberg, 2011. 457-462.

[23] Thomas, C., Wright, S., Hernandez, M., Flores, A., & García, M. Enhancing Student Engagement in Engineering Education with Machine Learning. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/123

[24] Rodriguez-Baca, L. S. ., Allagi, S. ., Larrea-Serquen,

R. ., Cruzado, C. F. ., Diaz, M. A. ., Garcia-Hernández, S. ., & Monteiro, J. D. . (2023). Experimental Study based on the Implementation of a Regulatory Framework for the Improvement of Cyber Resilience in SMEs. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 199–205. https://doi.org/10.17762/ijritcc.v11i3.6337

[25] Rohokale, M. S., Dhabliya, D., Sathish, T., Vijayan, V., & Senthilkumar, N. (2021). A novel two-step co-precipitation approach of CuS/NiMn2O4 heterostructured nanocatalyst for enhanced visible light driven photocatalytic activity via efficient photo-induced charge separation properties. Physica B: Condensed Matter, 610 doi:10.1016/j.physb.2021.412902