Form 2: Literature Documents

**1. Team No:** 11

**2. Project Title**: Investing and finding a DNA Cryptography layer for securing data in spark cluster.

**3. Problem Statement**

Spark is a very versatile, adaptable platform for storage of data and focusing data analytics, data driven applications, however it was not begun in light of security or authorization for data. Security and controlling data are most significant fragments of any distributed platform that wants to break into the endeavor standard.

**Existing method(s) disadvantages**

The primary existing problem in user authentication within spark is the reliance on third-party security measures, such as Kerberos, which introduces computational overhead. The fragmented nature of security capabilities, developed independently, complicates user authentication and access control.

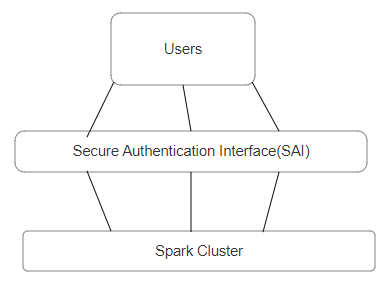
Users face challenges in securing their data, particularly in the absence of a unified security mechanism. The need for a single security instance for user authentication becomes apparent to address these existing authentication issues in spark.

**4. Problem Illustration:**

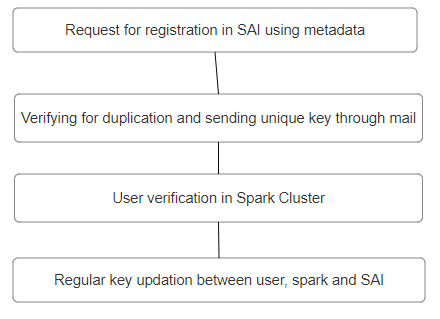
Consider a scenario where multiple users attempt to access and interact with data stored in a Spark Cluster. In the current setup, the user authentication process relies on third-party security protocols, such as Kerberos.

1. Fragmented Authentication Mechanisms:
   * The illustration depicts different users (User A, User B, and User C) attempting to authenticate through various security mechanisms.
   * User A encounters authentication through Kerberos, while User B goes through a different process.
   * This fragmentation in authentication mechanisms complicates the overall security landscape, making it challenging to maintain a cohesive and standardized approach.
2. Computational Overhead:
   * Each user's authentication involves multiple computational steps and interactions with external authentication services.
   * This process introduces computational overhead, as illustrated by the intricate network of calculations and verifications.

**5 .Concept Tree**

****

**Fig 1-Proposed Secure Authentication Interface.**

****

**Fig 2 –Steps between User, Spark cluster and SAI**

**6. Comparison of Existing Strategies for Problem solve**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.No** | **Strategies** | **Advantages** | **Disadvantages** |
| 1 | Kerberos Authentication | • **Strong security**: Kerberos is a well-established and widely used authentication protocol.  • **Centralized authentication**: It uses a centralized Key Distribution Center (KDC) for secure authentication. | • **Complexity**: Setting up and maintaining a Kerberos infrastructure can be complex.  • **Overhead**: Introduces computational overhead due to the need for ticket requests and validations. |
| 2 | Custom Token-based Authentication | •**Flexibility**: Custom token-based authentication allows for a flexible and tailored solution.  • **Stateless**: Tokens can be designed to be stateless, reducing the need for server-side storage. | •**Development effort**: Implementing a custom solution requires additional development effort.  • **Security risks**: Custom solutions may have security risks if not designed and implemented correctly. |
| 3 | Multi-factor Authentication (MFA) | • **Enhanced security**: Adds an extra layer of security through multiple authentication factors.  •**Compliance**: Meets security compliance requirements in certain environments. | • **User experience**: MFA may introduce additional steps for users, impacting user experience.  •**Implementation complexity**: Implementing and managing MFA can be complex. |
| 4 | OAuth 2.0 Integration | **• Industry standard**: OAuth 2.0 is a widely adopted industry standard for authentication.  • **Third-party authentication**: Allows for integration with third-party identity providers. | •**Complexity**: Implementing OAuth 2.0 can be complex and may require careful configuration.  • **External dependencies**: Relies on external OAuth providers, which may introduce dependencies. |

**7. Comparison of Existing Method from selected Strategies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.No** | **Author** | **Strategies** | **Advantages** | **Disadvantages** |
| 1 | Nan Zhang, Et al. | Centralized Key Distribution Center (KDC) | •Kerberos enables Single Sign-On(SSO)  • Authenticate once and access multiple services  • Receive a ticket granting ticket (TGT) for various services. | • Spark's reliance on Kerberos infrastructure  • It dependent on the availability and reliability of the KDC  • Introducing a potential single point of failure. |
| 2 | Yong-Tae Kim Et al. | Stateless Authentication | • Compatibility with RESTful Architectures as tokens can be included in HTTP headers. | • Limited Integration with External Identity  •may face challenges when integrating with external identity providers |
| 3 | Paul J. E. Velthuis | Adds an extra layer of security by requiring users to provide multiple forms of identification. | • MFA reduces the risk of unauthorized access, even if one factor (e.g., a password) is compromised. | •Potential User Friction:  MFA may introduce friction for users, requiring additional steps in the authentication process, which could impact user experience. |
| 4 | Tamjid Al Rahat ,Et al. | Integrate seamlessly with third-party identity providers. | • Reduced Credential Exposure: OAuth 2.0 uses access tokens rather than exposing user credentials | • Limited Control Over Authentication Process:  Integrating OAuth 2.0 may limit the level of control Spark has over the authentication process compared to custom or in-house solutions. |

**8. References**

[1] Nan Zhang, Xiaoyu Wu, Cheng Yang, Yinghua Shen and Yingye Cheng,"A lightweight authentication and authorization solution based on Kerberos," 2016 IEEE Advanced Information Management, Communicates,Electronic and Automation Control Conference (IMCEC), Xi'an, 2016, pp.742-746,DOI:10.1109/IMCEC.2016. 7867308.

[2] Yoon-Su Jeong & Yong-Tae Kim “A token-based authentication security scheme for Hadoop distributed file system using elliptic curve cryptography” August 2015Journal of Computer Virology and Hacking Techniques 11(3) DOI:10.1007/s11416-014-0236-5

[3] Paul j. e. Velthuis, "New authentication mechanism using certificates for big data analytic tools", kth royal institute of technology school of information and communication https://kth.divaportal.org/smash/get/diva2:1149007/FULLTEXT01.pdf

[4] Tamjid Al Rahat,Yu Feng, Yuan Tian, " Vulnerability Detection in OAuth Service Provider Implementations",ACM SIGSAC Conference on Computer and Communications Security (CCS ’22) November 2022 DOI:10.1145/3548606.3559381

**Signature Team Members Signature Supervisor**

**1.**

**2.**

**3.**