Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

Utilizing the SRP Protocol for Authentication and Authorization in Massive Data Sets

Surendra Shukla[1], Bhasker Pant[2], Rajesh Upadhyay[3]

[1]Department of Computer Science & Engineering,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[2]Department of Computer Science & Engineering,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[3]School of Management, Graphic Era Hill University, Dehradun, Uttarakhand India, 248002

ABSTRACT

Big Data is the term for the accumulation of massive volumes of data made possible by technological developments in the digital age. If we can quickly examine this massive amount of data, we can learn a lot. However, as Big Data may have catastrophic consequences if it is intercepted by unauthorised parties, its security has become a crucial consideration. In this work, we survey the protocols currently in use for managing authentication and access control. We also suggest a redesigned protocol that unifies authentication and authorisation to speed up the whole process.

Keywords: Secure Remote Password Protocol, Access Control, Authentication

INTRODUCTION

The term "Big Data" was coined by Google1 and refers to "very massive data sets that may be processed computationally to discover patterns, trends, and relationships, particularly pertaining to human behaviour and interactions." Big Data, thus, is a storehouse of massive amounts of information in a variety of formats, including structured, unstructured, and semi-structured data. Consistent analysis of the massive amounts of data produced by ubiquitous technologies like smartphones, e-commerce platforms, social networking sites, wireless sensor networks, and global positioning systems might provide useful insights. The size of Big Data is really enormous. In the present day, it is measured in zettabytes, which is more than a trillion gigabytes (1,099,511,627,776 GB, to be precise, or 1021 bytes), and this amount is predicted to grow exponentially in the future [1]. Big data analytics is very crucial in ways that are hard to fathom. When collected, processed, and analysed correctly, Big Data provides businesses with invaluable insights that can be used to boost productivity, revenue, cost-effectiveness, service quality, and product quality. For instance, because to the widespread availability of smartphones and other GPS devices, businesses may reach out to customers as they walk by their establishment of choice. This creates opportunities for several enterprises to reach a wider audience and generate new streams of income. However, broadcasters have begun tailoring commercials to specific viewers based on information about their demographics and

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

viewing habits.
There are six core aspects of Big Data that set it apart from conventional or relational databases . These traits are characterized by:

Data created by machines, networks, and systems (such as social media) is massive, thus the term "big data." Analyzing this volume of information is no simple feat.

• Diversity: information from a wide range of sources, both organised and unstructured, constitute the diversity feature. Images, audio files, movies, pictures, etc. are now commonplace forms of data storage, alongside the more traditional database files and Excel spreadsheets. The storage and analysis of this kind of unstructured data presents a number of challenges [2].

• Velocity: The tremendous pace at which the data is created, also known as "velocity." This is a big and persistent stream. Another problem for big data systems is the analysis of real-time data at such a high rate of speed [3].

• Authenticity: The data has to be cleaned up since it includes a lot of biases, noise, and irregularities. The "dirty data" that might arise from a lack of oversight in the collection of large amounts of data is resisted by the "veracity" of Big data [4].

• Volatility: Validity is the threshold beyond which data collection is pointless, and so analysis is terminated [5].

Data accuracy and reliability are two issues that are addressed by the concept of "validity." There should be no room for error while evaluating massive amounts of data. In Section II of the study, we provide an in-depth analysis of the SRP Authentication protocol and suggest our revised implementation, which adds an authorization step. Following this, we defend the SRP protocol in part III by detailing the myriad of security threats it effectively sidesteps. Section IV wraps up our research and discusses what lies beyond the scope of this article.

WHY SECURITY IS SO CRUCIAL FOR BIG DATA

Providing security for Big Data is difficult because of its volume and complexity. Due to the nature of Big Data, security is a crucial factor to address. If data is intercepted by a competing company, that company may utilize the information to its advantage, perhaps leading to the development of policies designed to undermine the target organization. There are three major concerns about Big Data security:

Safety of Information
The definition of data confidentiality is the prevention of disclosure of information without the owner's consent. It may also mean stolen information that has been manipulated [7]. In other words, the goal of data confidentiality is to set up systems of authorised user access. The nature of Big Data necessitates the incorporation of data from many different sources, each of which may have its own access policy. A complex job, merging several different access control rules into a single policy. In addition, the velocity trait may be mitigated by automating Big Data authorization [8].

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

Safety of Information
The integrity of data guarantees that it stays the same over the whole big data lifecycle. It's difficult with huge data to use data provenance and data correlation approaches.

Confidentiality of Personal Information
The purpose of data privacy is to prevent unauthorised individuals from gaining access to personally identifiable information contained inside a database. Big data research is being conducted to construct and analyse profiles of us, and if the obtained information is not safeguarded, it may be used for evil reasons [9]. This has made the topic of Data Privacy more important with the rise of social media. In this study, we aim to address the problems of data privacy and data veracity in the context of big data [10]. In this paper, we introduce the Secure Remote Password (SRP) Protocol, a robust authentication method, and present a modified version that incorporates attribute based access control inside the authentication stage to speed up the process, essential when dealing with Big Data.

SECURE REMOTE PASSWORD PROTOCOL

Identification by mutually accepted means, such as the verification of credentials, is called authentication. With Big Data, authentication becomes more difficult since data from many sources must all prove their identity to the same server. When it comes to authentication, the Secure Remote Password (SRP) protocol stands head and shoulders above the others [10]. The SRP protocol's elegance comes from its ease of use. Since the password is never sent over the network, zero-knowledge security is achieved. Mathematical equations are constantly used to calculate and transmit the value. Therefore, the customer never shares their password with anybody else. We are inspired by the SRP protocol and propose a variant of it that simultaneously implements attribute based access control (ABAC). The first draught of the SRP read as follows:

The client must first register with the Big Data server in order to proceed with Authentication. Because the client is a regular human being, all they need to remember is their login and a weak password. The first step in the authentication procedure is for the user to provide his ID or username to the server.

PROTOCOL FOR LOW-VOLTAGE SIGNAL RELAY PROCESSING IN VEHICLES

Attribute based authorisation for Big Data servers is suggested via the LV-SRP protocol, a variation of the SRP standard. Authorization is handled independently of authentication in conventional protocols. When implementing selective limits on access to resources, authorization, or access control, is required since not every user should be granted the same level of privileges. Rights and permissions may include the ability to read, write, modify, delete, create, etc. The term "attribute based access control" describes a mode of authorization in which privileges are conferred on users based on rules that take into account a combination of characteristics about those individuals. Time, date, user IP address, and corporate restrictions such as minimum withdrawal amounts, maximum allowable transactions, and so on are all examples of characteristics that may change on the fly. The LV-SRP protocol retains the standard registration procedure. The user calculates the verifier (using $v = gx$) and sends it, together with the login and salt value, to the server through an insecure

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

network. These values are kept in a table on the server.

In order to function, the LV-SRP relies on "level" values, which are singular identifiers for a certain rights hierarchy. Access at Level-1 will be Read-only, at Level-2 Read-Write, and so on. The client will transmit a user name and password to the server for authentication. Following this, the server does a lookup for each specified attribute and, based on the attribute values, assigns a Level value. One example of this is a banking application, which we will use to demonstrate our argument. First, let's say there are three constraint attributes: maximum withdrawal amount, maximum number of transactions, and time in milliseconds. The bank establishes a policy, and an if-then ladder is constructed appropriately.

Assignment of L = 1; ## only read if (Amount >= 50,000 || Trans >= 50 || Time >= 18:00).

If (Amount > = 50,000 and Trans is less than fifty and Time is less than 18:00), then assign L = 2; ## Read, Write. The L value for a specific user is determined by the server and is tied to the current value of the property in question at that instant in time. The updated SRP protocol calculates the LV value using a Hash add function and then utilises that value to partition the table. LV = A (L, v) The function A () is defined as the addition of L to V before the most significant bit. As a result, the LV values for a given access control right, say L = 1, will be constrained to a certain interval. If L=1 and v=24532, then LV=124532 is the derived value. When L = 1, all LV values will be between 10000 and 19999. (inclusive). Therefore, the determined access privileges are within the range of the estimated LV.

After this, the server will provide the client the salt value and the L value. Based on the client's password, we get back the value 'x'. At the client's end, the verifier is produced again, and then the L value is attached to it, yielding the LV. From then on, wherever V would have been used, LV will be substituted.

Utilization of Excessive Force
The attacker use a trial-and-error strategy to guess the user's password or other sensitive data. In order to get into the system, an attacker will try every conceivable combination of passwords, and in the worst-case scenario, will have to scan the whole search space. In contrast, the verifier in LV-SRP is an exponential number, and it is this value that is calculated and sent across the insecure channel. Therefore, not only is the assault laborious, but it also has high financial costs. Aiming to Use Words as Weapons

An example of a Brute Force assault, a dictionary attack simply tries every possible string in the dictionary. The dictionary is a collection of the most promising strings.The expense of this attack, however, would outweigh its benefits if the user were required to provide a password hashed using an exponential function. Thus, LV-SRP is still safe from such an assault.

Where g is the publicly known basis and x is any positive integer from zero to infinity, we get V= gx....(i). Discrete logarithms are the inverse of discrete exponentials, and discrete exponentials are used to compute V. Finding the discrete logarithm is a computationally challenging job, especially

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

for high values of n. (512 bits). Incidences That Happen Again

When a third party intercepts a command during transmission and then retransmits it at a later time, this is called a Replay attack. A hacker may get access to a protected computer or run instructions that are ordinarily unintelligible if they intercept the right signals. Deciphering a command before putting it to use is usually unnecessary. As a result, "Replay attacks are often easy to conduct and need little or no expertise.". There are two reasons why such assaults are futile against SRP. To begin, the LV-SRP protocol does not need the channel to be used for password transmission. Both the verifier and salt are sent along in the transmission. If an eavesdropper were to gain the verifier, the inverse logarithm of such a large exponential number would be very difficult to calculate. Suppose it is an integer with less than 128 bits. The inverse logarithm is easily determined under these conditions. It is impossible to learn the password or retransmit this x even if the inverse is discovered and the value of x is retrieved. A one-way hash is used to calculate the value 'x,' making it almost difficult to recover. As a result, LV-SRP prevents replay assaults.

Malicious Intermediary Attack
An eavesdropper attempting a Man-In-The-Middle attack will pose as either the client or the server in an effort to alter the sent message or get access to sensitive data. This kind of attack is precisely described as "Computer security breach in which a malevolent user intercepts and perhaps modifies data passing via a network." Since LV-SRP uses a one-way hash function, even if an eavesdropper obtains either the verifier or the salt, they will be unable to successfully impersonate either system. If he attempts to use this "password" to authenticate himself, he will be denied access even if it looks like any other password.

An eavesdropper monitoring a non-secure network would only be able to decipher the letters A, B, and u. Since the LV-SRP Protocol may be reduced to the Diffie-Hellman Protocol , this still won't provide him the session key.

The eavesdropper gains no useful information since the client and server both arrive at the identical numbers. The snooper can listen in.

restricted limited to numbers like A, B, and u. The session key S needs the random variables a (0 a n) on the client side and b (0 b n) on the server side, neither of which are provided by these values. Both the client and the server have access to the predetermined random values, but these values are never sent across the network. A Denning-Sacco Assault The eavesdropper in a Denning-Sacco attack obtains the session key, K, which is calculated by hashing the inputs A and B. If he is unsuccessful, he may try to impersonate the system or use a brute-force assault to crack the password. As was just discussed, an eavesdropper will not be able to glean any useful information just by acquiring M [1], M [2], or K. Therefore, the SRP is secure against such assaults.

This protocol, thus, protects against many assaults even if the channel is not totally secure. As has been shown, the system is able to fend against a variety of threats, safeguarding both individual components and the system as a whole. Obtaining the previous session key is likewise useless since the random variables are rotated at the beginning of each authentication cycle. As a result, no

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

previously-gathered data could ever be used to facilitate a breach in the present or the future.

CONCLUSION

Most systems need two different steps—one for authentication and one for authorization—which adds unnecessary complexity and overhead. In this study, we attempt to merge these two distinct security mechanisms into a unified whole, so as to increase the system's overall efficacy. SRTP is now the most used authentication protocol. Through incorporating the work done in a static role-based model with the SRP Protocol, we present a paradigm to enable dynamic attribute-based access control. We have also provided elaboration on the assaults that are defeated by the LV-SRP protocol and provided mathematical proof of this defeat.

Because of its adaptability, this model may be altered to meet the needs of a variety of organisations with varying specifications for model attributes. Because of the breadth and depth of possible applications, it attempts to provide a sophisticated and robust system for controlling access. It also makes an effort to speed up the combined process in contrast to the conventional approaches.

REFERENCES

1. Algaradi, T. S., & Rama, B. (2018, September). Big data security: a progress study of current user authentication schemes. In 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 68-75). IEEE.
2. Nguyen, M. C., & Won, H. S. (2017, February). Gateway-based access interface management in big data platform. In 2017 19th International Conference on Advanced Communication Technology (ICACT) (pp. 447-450). IEEE.
3. Ji, S., Gui, Z., Zhou, T., Yan, H., & Shen, J. (2018). An efficient and certificateless conditional privacy-preserving authentication scheme for wireless body area networks big data services. IEEE Access, 6, 69603-69611.
4. Ouda, A. (2016, March). A framework for next generation user authentication. In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-4). IEEE.
5. Nguyen, M. C., & Won, H. S. (2016, December). A Case Study on Web-based Analytic Workflow in Big Data Platform. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 427-430). IEEE.
6. Reddy, Y. (2018, May). Big data security in cloud environment. In 2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing,(HPSC) and IEEE International Conference on Intelligent Data and Security (IDS) (pp. 100-106). IEEE.
7. Singh, K. K., Dimri, P., & Rohatgi, S. (2016, November). Cloud testing and authentication model in financial market Big Data analytics. In 2016 International Conference System Modeling & Advancement in Research Trends (SMART) (pp. 242-245). IEEE.
8. Lee, K. Y., & Sim, J. Y. (2019, September). Cloud removal of satellite images using convolutional neural network with reliable cloudy image synthesis model. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 3581-3585). IEEE.
9. Ibrahim, A., & Ouda, A. (2016, October). Innovative data authentication model. In 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 1-7). IEEE.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/99

10. Chintakindi, R., & Mitra, A. (2020, February). Execution of real-time wide area monitoring system with big data functions and practices. In 2020 IEEE 9th Power India International Conference (PIICON) (pp. 1-6). IEEE.