Big Data Security Using RSA Algorithms in A VPN Domain

Big Data Security Using RSA Algorithms in A VPN Domain

1st Abel Yeboah-Ofori
School of Computing and Engineering
University of West London
United Kingdom
Abel.yeboah-ofori@uwl.ac.uk

1st Aishat Ganiyu
School of Computing and Engineering
University of West London
United Kingdom
21491949@student.uwl.ac.uk

Abstract*— Big Data security using encryption algorithms has become imperative due to the increased reliance on the volume, velocity, veracity, and value of data that organizations require to manage business processes, information sharing, and vulnerabilities that can be exploited. VPN tunneling ensures the confidentiality of data transmission over the network and remains secure from unauthorized parties. Big Data security within a VPN environment using RSA encryption to secure the data traversing between the established VPN tunnel. However, recent attacks on Big Data such as Man-in-the-middle, Evil twin attacks, DNS cache poisoning, phishing, injection, and DoS attacks, among others, have impacted greatly on organizations, leading to financial losses, data breaches, reputational damage, litigation issues, and trust. This paper explores how Big Data Security uses the RSA Algorithms in the VPN environment to establish secure tunneling and enhance security. The contribution of this paper is threefold. The foremost objective is to explore existing literature and state of the art to identify and analyze the prevalent Big Data challenges, threats, risks, and vulnerabilities that can compromise Big Data. In addition, we would compare encryption algorithms, such as AES, RSA, and DES, to determine secure features and relevance during data transmission in a VPN environment. Furthermore, we implement a VPN tunnel and encrypt the end-to-end network infrastructure for configuration and to secure the data traversing the network. Finally, we recommend security mechanisms to improve Big Data security in a VPN environment. The results highlight issues of improper data storage, inadequate authentication, and insufficient data protection mechanisms; it also discusses examples of Big Data security challenges and how RSA encryption could improve security on the VPN.

Keywords—Big Data Security, Encryptions, VPN, RSA, Cyber Security

I. Introduction

Big Data security has become crucial for businesses and organizations due to Big Data's 10 V's, which comprise visualization, velocity, variety, variability, volume, vulnerability, validity, volatility, veracity, and value [1] in information storage and retrieval. Companies such as Google, Amazon, Walmart, Sainsbury, and NHS, among others, use Big Data to acquire new information and improve existing processes. The popularity of Big Data systems has led to various attacks, such as DNS Spoofing, evil-twin attacks, Man-in-the-middle (MITM), phishing, denial-of-service (DoS), and injection attacks [2][3].

Regarding Big Data, [4] described it as networked, linked, and traceable data via the Internet and accompanying mobile technologies. The advent of Big Data not only presents the opportunity to collect more data but also raises privacy and security concerns [5], making it challenging to analyze such data with typical statistical analysis software. The authenticity and authorization of Big Data have grown difficult, owing to threat actors leveraging weaknesses in VPN setups, which results in various attacks. VPN tunnel provides a service that hides real traffic and can be used to establish a secure channel for communication. It establishes an SSL-protected channel between the client and the server; the established VPN tunnel is where all internet activities occur [6]. Fig. 1 depicts VPN tunnel configurations and how cyberattacks could exploit vulnerabilities and deploy attacks such as DoS, MITM, and DNS attacks.

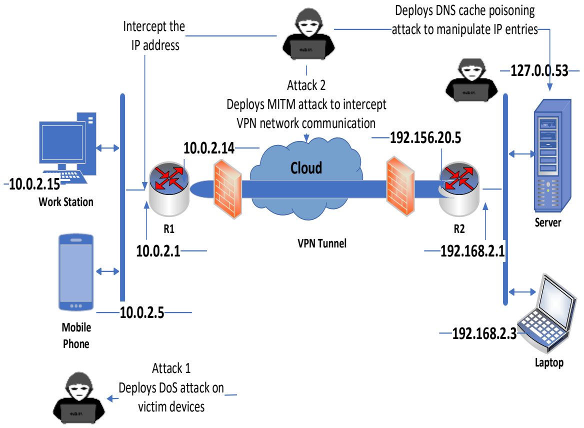


Fig. 1. VPN Virtual Client and Server Set-Up

Recent attacks on Big Data have impacted organizations, leading to financial losses, DoS, false narratives, and predictions. The issues of improper data storage, inadequate authentication, and insufficient data protection mechanisms are examples of Big Data security challenges. Thus, a VPN tunnel is necessary to secure information that traverses the network. The advantage of Big Data security using encryption algorithms is that its approaches can help combat cyber threats and ensure incident management [7]. The VPN SSL network guarantees the confidentiality of data transmission and ensures they remain secure from unauthorized parties [8].

Big Data challenges and the main characteristics of its vulnerabilities have increased recently in volume, velocity, value, veracity, and variety, often referred to as the 3Vs or 5Vs [9] to 10Vs as in Fig. 2 [1]. The volume considers the magnitude of the data, as Big Data sizes are captured and reported in terabytes and petabytes. Big data usage is a technological advancement that permits the use of structured, semi-structured, and unstructured data. Hence, Variety in Big Data refers to the structural heterogeneity in a dataset [9]; it involves the difference in sources and data types. Velocity considers how fast the data is produced and its processing time. However, the value is considered to be the usefulness of the information [10].

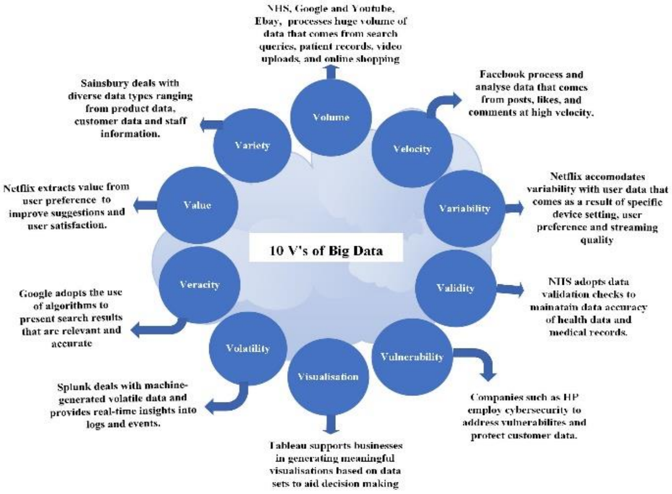

NHS, Google and Youtube, Ebay, processes huge volume of data that comes from search queries, patient records, video uploads, and online shopping
Sainsbury deals with diverse data types ranging from product data, customer data and staff information.
Facebook process and analyse data that comes from posts, likes, and comments at high velocity.
Volume
Variety
Velocity
Netflix extracts value from user preference to improve suggestions and user satisfaction.
Value
Variability
Netflix accomodates variability with user data that comes as a result of specific device setting, user preference and streaming quality
10 V's of Big Data
Veracity
Validity
Google adopts the use of algorithms to present search results that are relevant and accurate
NHS adopts data validation checks to mainatain data accuracy of health data and medical records.
Volatility
Visualisation
Vulnerability
Splunk deals with machine-generated volatile data and provides real-time insights into logs and events.
Companies such as HP employ cybersecurity to address vulnerabilites and protect customer data.
Tableau supports businesses in generating meaningful visualisations based on data sets to aid decision making

Fig. 2. Big Data 10 V's and Challenges

VPN tunnel provides access control that permits users to access a private network between two entities over the Internet. VPN tunneling involves encrypting the communication channel between systems in the VPN environment using TLS/SSL encryption protocols, among others. However, recent attacks on Big Data, such as MITM, Evil twin attacks, DNS cache poisoning, phishing, injection, and DoS attacks, and others [2][3], have impacted organizations greatly, leading to financial losses, data breaches, reputational damage, litigation, and trust issues.

This paper explores how Big Data Security uses RSA Algorithms in the VPN environment to establish secure tunneling and enhance security. The contribution of this paper is threefold. The foremost objective is to explore existing literature and state of the art to identify and analyze the prevalent Big Data challenges, threats, risks, and vulnerabilities that can compromise data. Further, we compare encryption algorithms such as AES, RSA, and DES to determine secure features and relevance during data transmission in a VPN environment. Furthermore, we implement a VPN tunnel and encrypt the end-to-end network infrastructure for configuration to secure the data traversing the network. Finally, we recommend security mechanisms to improve Big Data security in a VPN environment. The results highlight issues of improper data storage, inadequate authentication, and insufficient data protection mechanisms. In addition, we will also discuss examples of Big Data security challenges and how RSA encryption could improve security on the VPN.

II. Related works

This section discusses the state-of-the-art Big Data and related works, as well as the existing encryption approaches such as AES, RSA, and DES algorithms, which are used during VPN tunneling, to identify existing challenges and limitations with Big Data security.

Regarding cryptography algorithms, Patil et al. [11] carried out a comprehensive evaluation of cryptographic algorithms, including DES, 3DES, AES, RSA, and Blowfish, to determine the best-suited algorithm for a given function due to the variability with the cryptographic algorithms. The authors described the RSA as an asymmetric encryption algorithm involving two different keys; the public key is used for encryption, and the secret key is used for decryption. Additionally, AES and DES algorithms use symmetric encryption algorithms on a single private key for encryption and decryption. However, the work did not consider VPN secure communication channels using encryption.

Yu [12] analyzed Big Data threats in computer networks and recommended firewalls and DMZ for security strategies. However, the author did not consider it from the perspective of VPN security. Tian and Jiang [13] explored the application of Big Data in information security by highlighting data leaks and using an algorithm to calculate data flow in a network. Goel. et al. [14] reviewed Big Data privacy and security challenges by highlighting issues of the Vs in managing data, infrastructure security, Hadoop security, privacy, and integrity. However, the work did not address encryption issues in VPNs. Qureshi [15] conducted a comparative study on recent trends to secure Big Data and its security challenges. They proposed security management tools such as Hadoop security. However, the work did not consider privacy issues during data transition and using VPN. Hussain et al. [16] analyzed the application of Big Data based on network security and Intelligence by reviewing the various technology applications. However, the work did not consider VPN security issues. [17] highlighted challenges and issues of Big Data, including data capture, storage, searching, sharing, analysis, and visualization in huge sizes (terabytes and petabytes) and various data types (structured, semi-structured, and unstructured data) [9]. Big Data security challenges are vast. [7] posit that one of the disadvantages of Big Data is the need to utilize the latest technologies that are not mature enough for a comprehensive test against cyberattacks. Consequently, attackers are using sophisticated tools to deploy various attacks, such as MITM and distributed denial of service (DDoS), to infiltrate, exfiltrate, manipulate data, and penetrate the network due to the popularity and value of Big Data. These attack patterns come in various forms and are mostly conducted by generating a request for the software service. They are most often accomplished by utilizing a network's weakness [18].

Regarding authentication and encryption algorithms, [19] investigated the performance of the authentication and encryption algorithms in IPsec-based Mobile VPN (MVPNs) and its impact on the quality of MVPN design implementation by running C++ code and compiling them with Microsoft Visual Studio. The results show that hash-based message authentication code (MD5) - HMAC(MD5) outperformed AES encryption. However, the test was not extended to RSA and Blowfish algorithms. Tillah et al. [8] explored using VPN SSL to secure data transmission between the access control module and the database server to prevent MITM and data theft attacks by implementing IPsec and SSL protocol on OpenVPN for configuration. The results show that the use of two encryption algorithms secures the values in layers, making it difficult to exploit the encrypted variables. We discuss key concepts including the encryption algorithm RSA and TLS/SSL.

A. *RSA (Rivest, Shamir, Adleman)*

The function of the RSA is to ensure integrity, confidentiality, authentication, and nonrepudiation of data [20]. RSA deals with public and private keys for encrypting and decrypting. The process involves Key generation, where the public and private keys are created. Encryption, where plain text is converted to cipher text. Decryption is the reverse of encryption as it involves reversing the cipher text to the original text. RSA should be utilized with a high number of bits to maintain its security. Recent developments in RSA have heightened the threats as adversaries constantly try to crack the encryption, so the highest number of bits possible is recommended to ensure that the RSA implementation is not tampered with. We chose the RSA encryption algorithm since RSA provides better key exchange, digital signatures, and PKI relevant for Big Data transition in VPN.

B. *TLS / SSL Protocols*

TLS and SSL are security protocols that provide an encrypted link between a browser and a web server, ensuring that the transmitted data between the two connections remains secure and private [21]. TLS establishes confidentiality by blocking all access to the actual content of the transmitted message between the two ends of the medium. TLS verifies integrity through the detection of tampered data within the channel. TLS authenticates at least one end of the channel to avoid implementing a protected channel with an attacker [22]. On the other hand, SSL VPN provides end-to-end encryption (E2EE) between the VPN server and the VPN client, making it difficult for anyone to sniff the packets or intercept the data [21].

Overall, the existing literature considered various encryption algorithms for network access controls and VPNs. However, the works did not consider the role of Big Data security in the VPN environment. Thus, it gives us the rationale to implement RSA encryption within VPN tunneling to secure Big Data.

III. Approach

This section presents an overview of the approach used for the paper, with a primary focus on assessing the security implementation using the RSA algorithm on a VPN tunnel and analyzing the attacks on Big Data. Our approach considers using the RSA encryption algorithm on the VPN to understand the attack pattern in order to address cybersecurity issues on the VPN tunneling. The paper uses a qualitative approach by implementing attack methods to identify vulnerabilities through the VPN tunneling implementation threat landscape.

Our approach aims to identify the threats and attacks that can be deployed to compromise Big Data security. The initial step involves setting up four virtual machines (VMs) using Virtual Box for the experimental technique. We installed an Ubuntu Linux environment to perform various activities and test the effectiveness of our approach by running the RSA algorithm in a C programming language and using the OpenSSL library to monitor the time it takes to generate the prime numbers and formulate the private and public key pairs. Fig. 3 shows how we set up a VPN tunnel in a network work environment by creating three clients and a server and configuring them to establish a secure connection using an RSA encryption mechanism. Further, we set up a secure VPN tunnel between the server and the clients. Then, we established a TCP connection between the server and a client using the RSA encryption mechanism to encrypt messages that pass through the VPN. Finally, we discuss Big Data security using RSA encryption and VPN. As illustrated in Fig. 1, VPN tunneling attacks, some of which are DoS, MITM, and DNS spoofing attacks, can be deployed on the network. DoS is a cyber-attack committed on a specific network or server to disrupt normal operations [18]. MITM attack occurs through the interception of traffic traversing along two users [23]. It occurs when a device that contains essential data is compromised. DNS Spoofing/ DNS Cache Poisoning is the term used when an adversary manipulates the IP entries in the DNS server to enable the client to bypass the authentic website and reroute to the threat actor's fraudulent one.

This paper demonstrates the DNS Cache poisoning attack in operation and the measures that can be taken to prevent such attacks in VPN tunneling.

IV. Implementation

We conducted the following tasks for the implementation: setting up the VPN tunnel between the server and the client, establishing communication between the server and the internal and external clients, and using a Firewall to block specific ports from the machines.

A. *Set Up Four Virtual Machines (VMs)*

Using Virtual Box software, we set up four virtual machines (VMs) and connected them to the same local area network (LAN). Among the four VMs, three were configured as clients, namely PubClient-U, PubClient-S, and PrtClient-V. The fourth virtual machine (VM) was configured as the VPN server.

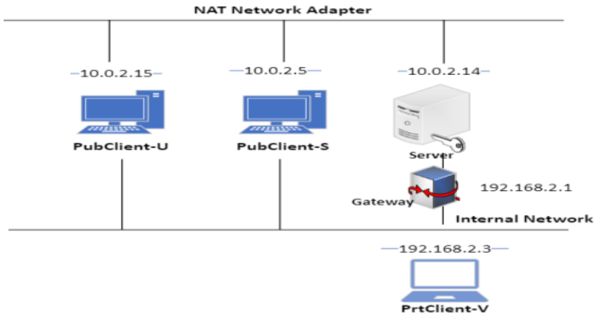
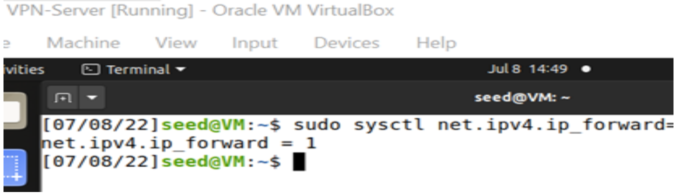


Fig. 3. Virtual machine network configuration

To enable the server to function as a gateway rather than a host like the clients, we initiated a command sudo sysctl net.ipv4.ip_forward=1. [24] that enabled IP forwarding, as demonstrated in Fig. 4.


VPN-Server [Running] - Oracle VM VirtualBox
e Machine View Input Devices Help
ivities Terminal
Jul 8 14:49
seed@VM: ~

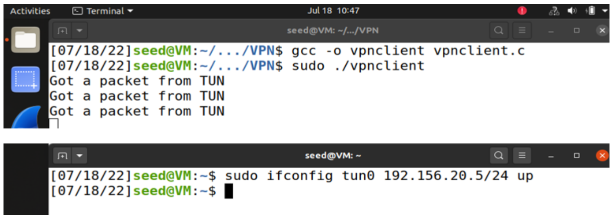
```
[07/08/22]seed@VM:~$ sudo sysctl net.ipv4.ip_forward=
net.ipv4.ip_forward = 1
[07/08/22]seed@VM:~$ █
```

Fig. 4. Enabling IP forwarding for the Server to function as a Gateway.

B. *Code Generation for the Encryption*

When establishing a VPN tunnel, the machines must be able to communicate with each other. We set up this tunnel by utilizing the tun/tap interface, an enabling technology for the TLS/SSL VPN [22]. Tun/tap interfaces are virtual network interfaces used to create virtual instances of physical connections; their primary function is to pass data from one host to another [25]. The TUN is an essential entity within the tun/tap interface. TUN is an acronym for the word tunnel. It is an alternative to a physical port whose role is to send/receive packages through a cable [26].

Subsequently, we created a server and a client program where we initiated the TUN device and server configuration. We connected the machines using sockets and established a secure VPN tunnel. The server and the client program were made executable using their respective commands -sudo *./vpnclient* and sudo *./vpnserver*. This resulted in creating the tun0 tunnel interface, which was activated by assigning an IP address to tun0, as illustrated in Fig. 5, leading to the establishment of the socket connection between the two devices.



```
[07/18/22]seed@VM:~/.../VPN$ gcc -o vpnclient vpnclient.c
[07/18/22]seed@VM:~/.../VPN$ sudo ./vpnclient
Got a packet from TUN
Got a packet from TUN
Got a packet from TUN
```

```
[07/18/22]seed@VM:~$ sudo ifconfig tun0 192.156.20.5/24 up
[07/18/22]seed@VM:~$ 
```

Fig. 5. The client establishes a connection with the TUN interface

C. *Test the UDP connection with the firewall.*

After successfully setting up the socket connection and initiating the TUN processes, we created a firewall rule to accept all connections from the specified port number. This was done by inputting the following command *sudo ufw allow 55555/udp* to the terminal to update the firewall table for the VPN server and client VM. After configuring the firewall rule, the port becomes visible and open for connection, making it easy for the client and server to communicate.

This UDP socket connection is quick and easy to set up, but it has a vulnerability that makes it susceptible to threat actors. We migrated from the UDP socket connection to protect the data that traverses between the client and the server. We utilized the TLS and TCP approach to establish a secure connection.

D. *Establishing a secure connection with TLS and RSA:*

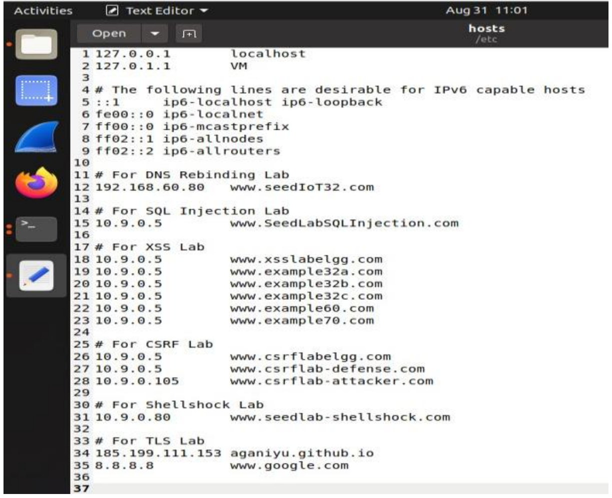
Before integrating the protocol into our program, the first task to be completed is creating or using an existing certificate authority (CA). The CA is an essential entity whose function is to establish the integrity of a server. Due to the non-existence of an existing CA, we created a new CA for this research using the OpenSSL library. The next task involved generating the RSA public and private key pair for the root CA and the intermediate CA; the intermediate CA will then issue an x509 certificate, which will be deployed to our web server to establish an HTTPS connection. The final step involved generating a certificate signing request (CSR) by requesting a certificate and passing the key pair as a parameter. Upon receiving the public key certificate signed and issued by the root CA, the certificate is then deployed to a web server. The firewall table was then updated, and a rule was set up to allow connections to port 443. Once the CA was established and the keys were generated with RSA, we executed a TLS socket program and tested the connection, which worked successfully. The only drawback is the manual creation of the CA, which led to numerous warnings from the browser that the root CA was not authorized due to the manual insertion of the root CA into the browser-trusted certificate for experimental purposes and to test our connection.

V. Results and Discussions

The paper has discussed several challenges and issues that impact Big Data security, including attacks deployed in the VPN environment to exploit vulnerabilities. Du [22] states that public key cryptography is the foundation for secure communication. Thus, discussing the various encryption algorithms, such as RSA, AES and DES as well as their implementation issues during secure tunneling in this paper was essential. [17] states that confidentiality and integrity must be achieved to secure a tunnel; using the TLS, a transport layer protocol. Hence, we utilized TLS which is built on top of TCP, a more secure protocol for the network.

In our solution, to evaluate the weaknesses of the existing encryption algorithm and the VPN configuration, we used a Big Data application -Google, to test our solution, where we discovered two principal vulnerabilities: DNS spoofing and MITM. We identified the vulnerabilities by acting as a threat actor and manipulated the DNS entries, as seen in Fig. 6. We set Google's IP address to 8.8.8.8 as the IP addresses are dynamic, and change from time to time, as highlighted in Fig. 7.

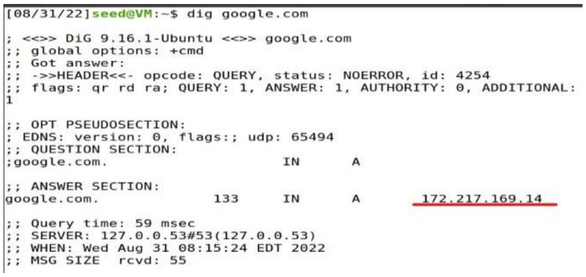
Fig. 6 shows how we manipulated the DNS configuration file by manually inserting the IP address and the hostname into the /etc/hosts file, and made the changes by logging in as a root user.



```
127.0.0.1       localhost
127.0.1.1       VM

# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

# For DNS Rebinding Lab
192.168.60.80   www.seedIoT32.com

# For SQL Injection Lab
10.9.0.5        www.SeedLabSQLInjection.com

# For XSS Lab
10.9.0.5        www.xsslabelgg.com
10.9.0.5        www.example32a.com
10.9.0.5        www.example32b.com
10.9.0.5        www.example32c.com
10.9.0.5        www.example60.com
10.9.0.5        www.example70.com

# For CSRF Lab
10.9.0.5        www.csrflabelgg.com
10.9.0.5        www.csrflab-defense.com
10.9.0.105      www.csrflab-attacker.com

# For Shellshock Lab
10.9.0.80       www.seedlab-shellshock.com

# For TLS Lab
185.199.111.153 aganiyu.github.io
8.8.8.8         www.google.com
```

Fig. 6. Manipulated DNS entries, inserting a new IP address

The dig command tool in Linux can be used to retrieve information about a domain. Fig. 7 shows how we use the *dig google.com* command to check Google's IP address to realize its dynamic state when we simulate the attack process of deploying DNS spoofing. Showing the IP address as *172.217.169.14.*



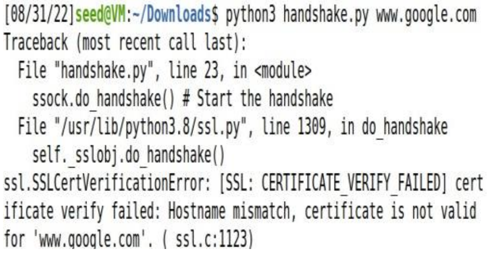
```
[08/31/22]seed@VM:~$ dig google.com

; <<>> DiG 9.16.1-Ubuntu <<>> google.com
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 4254
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 
1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 65494
;; QUESTION SECTION:
;google.com.                    IN      A

;; ANSWER SECTION:
google.com.             133     IN      A       172.217.169.14

;; Query time: 59 msec
;; SERVER: 127.0.0.53#53(127.0.0.53)
;; WHEN: Wed Aug 31 08:15:24 EDT 2022
;; MSG SIZE  rcvd: 55
```

Fig. 7. Google's dynamic IP address -1

To complete the experiment, we imported the CA as we had done for the TLS within the VPN tunnel. Then, we modified the /etc/host file and inserted the static IP address. We executed the command to gain access to Google's website but were redirected. Hence, we demonstrate how an attacker can initiate a DNS spoofing attack or a DNS cache poisoning attack, be it an insider attacker or any threat actor.

For testing purposes, we manually edited the /etc/hosts file, which could lead to an MITM attack as the potential adversary may manipulate the user's DNS cache by switching the IP address of a legitimate server to that of the adversary to commit an MITM attack.

To address and prevent this vulnerability, we included the hostname check in the client program as recommended by [22] to ensure that the address on the certificate matches the client address; this was implemented by updating the program and including the hostname check. Thus, Fig 8 shows re-execution of the program denies the manipulated entry and demands a valid certificate for the hostname.



```
[08/31/22]seed@VM:~/Downloads$ python3 handshake.py www.google.com
Traceback (most recent call last):
  File "handshake.py", line 23, in <module>
    ssock.do_handshake() # Start the handshake
  File "/usr/lib/python3.8/ssl.py", line 1309, in do_handshake
    self._sslobj.do_handshake()
ssl.SSLCertVerificationError: [SSL: CERTIFICATE_VERIFY_FAILED] cert
ificate verify failed: Hostname mismatch, certificate is not valid
for 'www.google.com'. ( ssl.c:1123)
```

Fig. 8. Image illustrating connection failure when DNS cache poisoning.

For testing purposes, we manually modified our clients' and servers' local DNS files to simulate the attack. The DNS contains a list of IP addresses and domain names that allow the system to locate and communicate with each other. The threat of DNS cache poisoning or DNS spoofing is a challenge within the VPN tunnel that weakens Big Data security. DNS spoofing attacks occur when an adversary intends to deceive the user by providing a fraudulent IP address to victims and tricking them into communicating with an unintended machine [22].

Our research findings clearly demonstrate that our measures have successfully reduced the impact of the vulnerability on our system. We will now discuss the proposed solution.

A. *Proposed Solutions*

Implementing RSA algorithms in the virtual private network environment to improve data security and privacy has become essential due to the constant challenges and changes in the 10Vs of Big Data, including the vagueness, venue, vocabulary, and semantics of Big Data, among others. These challenges regarding the size of Big Data, inaccuracies, heterogeneity, dynamics, speed of generating Big Data, and invalidities have brought with them vulnerabilities that are being exploited by cybercriminals. The use of RSA encryption in securing Big Data during data transition is critical considering the structured, semi-structured, and unstructured data in the heterogeneity of the dataset [6] [18]. Thus, using RSA encryption in VPN tunneling has become necessary for securing information that flows across a network to ensure confidentiality to authorized users.

B. *Big Data Security and Privacy*

Factors impacting Big Data security and privacy include key management, data encryption, quantum cryptography, data anonymization, access control and authentications, firewall configuration mechanisms, and secure VPN tunnels. Table 1 addresses some security recommendations to improve VPN and Big Data security.

Our results demonstrated that implementing RSA encryption is effective in securing both a VPN tunnel and a Big Data system to improve security.

VI. Conclusion

The paper has explored Big Data security challenges, vulnerabilities, and attacks that are being deployed to exploit the VPN network. We have discussed the state of the art and encryption algorithms that are being used in the VPN environment. We have discussed some of the existing threats and attacks that can compromise Big Data security while utilizing protocols and RSA encryption mechanisms to establish secure communication. We implemented the VPN tunnel, which was used to determine the vulnerabilities when Big Data is traversing the network. We have recommended a practical approach to securing Big Data using the RSA encryption algorithm and control mechanisms to secure communication, as discussed in Table 1. The results have highlighted how the issues of improper data storage, inadequate authentication, insufficient data protection mechanisms, and challenges of Big Data could impact network systems and how RSA encryption could improve the security of VPN tunnels. Table 2 discusses the various encryption algorithms and their secure features.

Future work will consider secure VPN tunneling in Quantum computing to improve Big Data security.

TABLE I. VPN and Big Data Security Recommendations

Big Data Security	**Attack**	**Control Mechanisms**
Access Control	Session Hijacking attack on network connections to obtain IP for DoS attack	Apply Secure VPN and RSA encryption algorithm to conceal IPs and DNS address leaks.
Authentication	Attack on user password credentials	Implements Multifactor Authentication, digital certificates, and a strong password policy
Data Encryption	Exploit encryption keys during key exchange.	Apply strong encryption algorithms such as RSA and tokenization to protect privacy and sensitive data.
VPN Tunnel Security	Exploit SMTP, HTTP, and FTP protocols between networks to intercept communication.	Apply appropriate OpenVPN and IPSec protocols to encrypt data in transit. Use key management to provide confidentiality and integrity.
Data Anonymization	Data De-anonymization attacks by Intercepting networks and manipulating big data to generate false narratives.	Apply masking and anonymization techniques to protect data from inferences and dictionary attacks.

Socket End Points	Attack the socket programming and the API functions that connect processes and applications for the sender and receive data to inject false data.	Apply sanitization to validate input. Use TLS for website encryption and integrity checks. Use IDS/IPS to monitor traffic and for session management.
Firewalls	Exploit firewall misconfigurations on webservers	Employ expertise to implement appropriate firewall configurations such as DFI, application-level gateway firewall, NGFW

TABLE 2. Secure Features and Relevance of Encryption Algorithms

Encryption Algorithm	**RSA (Rivest-Shamir-Adleman)**	**AES (Advanced Encryption Standard)**	**DES (Data Encryption Standard)**	**ECC (Elliptic Curve Cryptography)**
Secure Features	Asymmetric encryption: Utilizes public and private key pairs for secure communication and digital signatures.	Symmetric encryption: Requires secure key exchange as it shares key for encryption and decryption with 128, 192, and 256-bit key lengths.	Symmetric encryption: Shares key for encryption and decryption and 56-bit key length.	Asymmetric encryption: Elliptic curves in mathematics are used for key generation and to establish secure connections.
Relevance	Strong security, efficiency, and fast encryption and decryption, RSA provides better key exchange, digital signatures, and public key cryptography that is relevant for Big Data transition in VPN.	AES provides strong security, efficiency, and fast encryption and decryption but does not perform well in Big Data transition in VPN.	Deprecated, Small key size, vulnerable to cyber-attacks in a network environment.	ECC has the potential to provide efficient and strong security algorithms for the future.

REFERENCES

[1] H. Soomro, “Mastering the 10 vs of big data,” datasciencedojo, 31 January 2023. [Online]. Available: https://datasciencedojo.com/blog/10-vs-of-big-data/. [Accessed 24 December 2023].

[2] A. Yeboah-Ofori, I. Darvishi, A.S. Opeyemi, "Enhancement of Big Data Security in Cloud Computing Using RSA Algorithm," 2023 10th International Conference on Future Internet of Things and Cloud (FICloud). 2023, pp. 1-7, doi: 10.1109/FiCloud58648.2023.00053

[3] A. Yeboah-Ofori and A. Hawsh, "Evil Twin Attacks on Smart Home IoT Devices for Visually Impaired Users," 2023 IEEE International Smart Cities Conference (ISC2), Bucharest, Romania, 2023, pp. 1-7, doi: 10.1109/ISC257844.2023.10293225.

[4] K. Mills, Big Data for Qualitative Research. Oxon: Routledge, 2019.

[5] Dincer and E. Zeydan, "Big data security: Requirements, challenges and preservation of private data inside mobile operators," 2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Istanbul, Turkey, 2017, pp. 1-6, doi: 10.1109/BlackSeaCom.2017.8277711.

[6] M. Zain ul Abideen, S. Saleem, M. Ejaz. (2019). “VPN Traffic Detection in SSL-Protected Channel” Security and Communication Networks [Online]. Vol. 2019, Article ID 7924690, 17 pages. Available: https://www.hindawi.com/journals/scn/2019/7924690/

[7] E. Damiani, C.A. Ardagna, F. Zavatarelli, E. Rekleitis, L. Marinos (2016) “Big Data Threat Landscape and Good Practice Guide”. [online] Academia.edu. Available at: https://www.academia.edu/22838790/Big_Data_Threat_Landscape_and_Good_Practice_Guide

[8] A. M. Tillah, D. Ogi, M. Febriyanto and D. A. Farhatin, "Access Control System based on Secret Sharing Scheme with Secure Web Database and SHA-3 Password Authentication," 2021 6th International Workshop on Big Data and Information Security (IWBIS), Depok, Indonesia, 2021, pp. 145-152, doi: 10.1109/IWBIS53353.2021.9631847.

[9] A. Gandomi, M. Haider (2015) “Beyond the hype: Big Data concepts, methods, and analytics,” International Journal of Information Management, vol. 35 no. 2, pp. 137-144, Available: https://www.sciencedirect.com/science/article/pii/S0268401214001066

[10] L. Cui, F. R. Yu and Q. Yan, "When big data meets software-defined networking: SDN for big data and big data for SDN," in IEEE Network, vol. 30, no. 1, pp. 58-65, January-February 2016, doi: 10.1109/MNET.2016.7389832.

[11] P. Patil, P. Narayankar, D.G. Narayan, S.M. Meena. (2016) “A Comprehensive Evaluation of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish”, Procedia Computer Science, [Online] Vol. 78, pp. 617-624. Dec 2023 Available: https://www.sciencedirect.com/science/article/pii/S1877050916001101

[12] X. Yu, "Analysis of the Security Strategy of Computer Network Data under the Background of Big Data," 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2021, pp. 13-16, doi: 10.1109/ICAIBD51990.2021.9459026.

[13] Q. Tian and H. Jiang, "Application of Big Data Technology in Information Security," 2022 World Automation Congress (WAC), San Antonio, TX, USA, 2022, pp. 133-137, doi: 10.23919/WAC55640.2022.9934346.

[14] P. Goel, R. Patel, D. Garg and A. Ganatra, "A Review on Big Data: Privacy and Security Challenges," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 705-709, doi: 10.1109/ICSPC51351.2021.9451749

[15] K. Qureshi, "A Comparative Study on Recent Trends to Secure Big Data," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-3, doi: 10.1109/ICECCME57830.2023.10252572.

[16] K. Hussain, S. Sah, B. Seth, N. Fatima Rizvi and B. V. Febiyola Justin, "Analysis Application of Big Data-based Analysis of Network Security and Intelligence," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 1481-1485, doi: 10.1109/ICAIS56108.2023

[17] C.Y. Zhang, C.L.P. Chen, (2014) “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” Information Sciences, 275, pp. 314-347, Available: https://www-sciencedirect-com

[18] M.J. Awan, U. Farooq, H.M.A. Babar, A. Yasin, H. Nobanee, M. Hussain, O. Hakeem, A.M. Zain, (2021) “Real-Time DDoS Attack Detection System Using Big Data Approach”. Sustainability, Available: https://www.mdpi.com/2071-1050/13/19/10743/htm

[19] A. V. Uskov, "Information Security of IPsec-based Mobile VPN: Authentication and Encryption Algorithms Performance," 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, UK, 2012, pp. 1042-1048, doi: 10.1109/TrustCom.2012.187.

[20] M.P. Babitha, K. R. R. Babu. (2016) "Secure cloud storage using AES encryption," International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 859-864. Available at: https://ieeexplore.ieee.org/document/7877709.

[21] R. Walsh, 2018 “What is a SSL VPN? A Quick Guide to SSL & TLS” Available: https://proprivacy.com/vpn/guides/ssl-vpn-quick-guide-ssl-tls

[22] W. Du (2017) Computer Security A Hands-on Approach CreateSpace

[23] A. Yeboah-Ofori and A. Hawsh, "Effects of Cyberattacks on Virtual Reality and Augmented Reality Technologies for People with Disabilities," 2023 IEEE International Smart Cities Conference (ISC2), Bucharest, Romania, 2023, pp. 312-319, doi: 10.1109/ISC257844.2023.10293659. Available: http://www.dfd.gov.uk/R4D/Output/188391/Default.aspx

[24] Seed Labs (2014) “Virtual Private Network”. Available: https://web.ecs.syr.edu/~wedu/seed/Labs/VPN/VPN.pdf

[25] R. Walsh, 2021 “What is TUN/TAP and Why do VPNs Use Them?” Available: https://proprivacy.com/vpn/guides/tun-tap

[26] H. Yuksel and Ö. Altunay, "Host-to-host TCP/IP connection over serial ports using visible light communication," Physical Communication, 2020