

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum - 590014, Karnataka



**CRYPTOGRAPHY
AAT REPORT**

On
ChaCha20

Submitted by

Sarthaka Mitra GB (1BM23CS305)
Shaikh Uzair Ahmed (1BM23CS307)
Shashank U (1BM23CS314)
Subramanya J (1BM23CS343)

Under the Guidance of

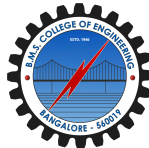
Dr. Nandhini Vineeth
Associate Professor
Department of CSE
B. M. S. College Of Engineering

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING



B. M. S. College Of Engineering
(Autonomous Institution under VTU)

BENGALURU - 560019

February 2025 to June 2025

B. M. S. College Of Engineering,
Bull Temple Road, Bengaluru 560019
(Affiliated To Visvesvaraya Technological University, Belgaum)
Department of Computer Science and Engineering



CERTIFICATE

This is to certify that the project work entitled **ChaCha20**, carried out by **Sarthaka Mitra GB (1BM23CS305)**, **Shaikh Uzair Ahmed (1BM23CS307)**, **Shashank U (1BM23CS314)**, **Subramanya J (1BM23CS343)**, who are bonafide students of **B. M. S. College of Engineering**, is in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belgaum during the year 2024-2025. The AAT has been approved as it satisfies the academic requirements of the course CRYPTOGRAPHY (23CS4ESCRP) work prescribed for the said degree.

Signature of the Guide

Dr. Nandhini Vineeth

Associate Professor

Department of CSE

B. M. S. College Of Engineering

Signature of the HOD

Dr. Kavita Sooda

Professor and Head

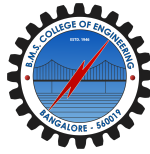
Department of CSE

B. M. S. College Of Engineering

Name of the Examiner

Signature with Date

**B. M. S. College Of Engineering,
Department of Computer Science and Engineering**



Declaration

We, Sarthaka Mitra GB (1BM23CS305), Shaikh Uzair Ahmed (1BM23CS307), Shashank U (1BM23CS314), Subramanya J (1BM23CS343), students of 4th Semester, B.E, Department of Computer Science and Engineering, BMS College of Engineering, Bengaluru, hereby declare that this AAT entitled **ChaCha20** has been carried out by us under the guidance of **Dr. Nandhini Vineeth**, Associate Professor, Department of CSE, B. M. S. College Of Engineering, Bangalore, during the academic semester **April 2025 - August 2025**.

We also declare that, to the best of our knowledge and belief, the development reported here is not a part of any other report by any other students.

Signatures

Sarthaka Mitra GB (1BM23CS305)

Shaikh Uzair Ahmed (1BM23CS307)

Shashank U (1BM23CS314)

Subramanya J (1BM23CS343)

Contents

1	Introduction	4
2	Methodology	7
2.1	Key and Nonce Generation	7
2.2	Initialization	8
2.3	Block Processing	8
2.4	Message Authentication	10
2.5	Decryption Process	10
3	Results and Discussion	11
4	Conclusion and Future Work	12

List of Figures

List of Tables

Chapter 1

Introduction

ChaCha20 is a stream cipher designed by Daniel J. Bernstein in 2008 as a variant of the Salsa20 cipher [1]. The primary motivation behind the development of ChaCha20 was to create a cryptographic algorithm that is both secure and efficient, particularly in software implementations. Bernstein aimed to address some of the limitations and vulnerabilities found in existing ciphers, particularly in the context of high-performance applications.

The development of ChaCha20 was primarily motivated by the need for a cryptographic algorithm that could achieve high security while maintaining efficiency, particularly in software implementations. Prior to ChaCha20, many cryptographic algorithms, such as RC4 and even Salsa20, were often optimized for hardware execution. This hardware-centric design approach leveraged the parallel processing capabilities of dedicated cryptographic hardware, which could perform operations at high speeds. However, as the landscape of computing evolved, there was a significant shift towards software-based implementations, especially in consumer devices, mobile applications, and cloud services.

Software implementations of cryptographic algorithms faced several challenges. Many existing algorithms were not designed with the constraints of software execution in mind, leading to inefficiencies such as high computational overhead and increased memory usage. For instance, algorithms that relied heavily on complex mathematical operations or extensive key schedules could result in slower performance on general-purpose processors. Additionally, software implementations were more susceptible to side-channel attacks, such as timing attacks, where an attacker could infer information based on the time taken to execute cryptographic operations.

ChaCha20 was specifically designed to address these shortcomings. It employs a simple and efficient structure that minimizes the number of operations required to generate the keystream, making it suitable for software execution. The algorithm's design allows it to perform well even on devices with limited processing power, such as smartphones and IoT devices. Furthermore, ChaCha20's key schedule is designed to be resistant to related-key attacks, enhancing its security profile.

By focusing on both security and efficiency, ChaCha20 has become a widely adopted algorithm in modern cryptographic protocols, including TLS and SSH, ensuring robust encryption in a software-driven environment. Its ability to deliver high performance without compromising security has made it a preferred choice for developers seeking to implement secure communication in a variety of applications.

Before the introduction of ChaCha20, several stream ciphers were widely used, including RC4 and Salsa20. RC4, designed by Ron Rivest in 1987, was popular due to its simplicity and speed. However, it was later found to have significant vulnera-

bilities, such as biases in the keystream that could be exploited in certain contexts, leading to potential security breaches. RC4's vulnerabilities were primarily due to inherent biases in its keystream, resulting in a non-uniform distribution of output bytes. This meant that certain bytes were more likely to appear than others, creating predictable patterns. Attackers could exploit these biases, especially in scenarios where the same key was reused, such as in HTTPS sessions. For instance, in the "Fluhrer, Mantin, and Shamir" attack, attackers could recover portions of plaintext by analyzing the keystream. This predictability undermined the confidentiality of encrypted data, leading to significant security breaches and prompting the cryptographic community to recommend against the use of RC4 in favor of more secure alternatives.

Salsa20, developed by Bernstein himself, was an improvement over RC4, offering better security and performance. However, it still faced scrutiny regarding its resistance to certain cryptanalytic attacks. Specifically, concerns arose about its key schedule and the potential for related-key attacks. Salsa20, designed by Bernstein as an enhancement over RC4, provided improved security and performance through its innovative design and efficient operations. However, it was not without vulnerabilities. Cryptanalysts raised concerns about its key schedule, which could potentially allow attackers to exploit related-key attacks, where knowledge of one key could lead to the compromise of another. Additionally, while Salsa20 demonstrated resilience against many known attacks, the possibility of future cryptanalytic advancements necessitated ongoing scrutiny and evaluation of its security in various contexts.

ChaCha20 was created to enhance the security and performance of Salsa20 while addressing these weaknesses. It features a more robust key schedule and improved diffusion properties, making it resistant to known attacks and suitable for high-performance applications, particularly in software environments where efficiency is crucial.

ChaCha20 operates on 64-byte blocks and utilizes a 256-bit key along with a 12-byte nonce (number used once). The algorithm employs a series of operations, including addition, bitwise XOR, and rotation, to produce a keystream that is combined with plaintext to generate ciphertext. The design of ChaCha20 emphasizes simplicity and speed, making it suitable for a wide range of platforms, from low-power devices to high-performance servers.

The security of ChaCha20 has been extensively analyzed, and it has been adopted in various cryptographic protocols, including TLS (Transport Layer Security) and SSH (Secure Shell). Its resilience against known cryptanalytic attacks and its performance characteristics have made it a preferred choice for modern encryption needs.

2. Why This is Relevant Today In an era where data breaches and cyberattacks are increasingly common, the need for robust encryption solutions has never been more critical. ChaCha20's relevance today stems from several factors:

Increased Data Sensitivity: With the proliferation of digital data, including personal, financial, and health information, there is a heightened need to protect sensitive data from unauthorized access. **Performance Requirements:** As applications become more demanding, the need for fast and efficient encryption algorithms is paramount. ChaCha20's design allows it to perform well on a variety of hardware, including mobile devices and embedded systems, where computational resources may be limited. **Regulatory Compliance:** Organizations are often required to comply with data protection regulations, such as GDPR (General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act). Implementing strong encryption methods like ChaCha20 helps organizations meet these legal re-

quirements and protect user privacy. 3. What Weaknesses Are We Tackling? The implementation of the ChaCha20 file encryption and decryption utility specifically addresses several weaknesses associated with traditional encryption methods:

Nonce Reuse: Many encryption algorithms are vulnerable to attacks if the same nonce is reused for multiple encryption operations. ChaCha20 mitigates this risk by generating a unique nonce for each encryption session, ensuring that even identical plaintexts yield different ciphertexts. Key Management: The utility emphasizes secure key management by deriving encryption keys from user-provided passwords, which can be further enhanced with key stretching techniques. This approach helps protect against brute-force attacks on weak passwords. Data Integrity: The utility incorporates the Poly1305 message authentication code (MAC) to ensure the integrity and authenticity of the encrypted data. This addresses the risk of data tampering and provides assurance that the data has not been altered during storage or transmission. By leveraging the strengths of ChaCha20 and addressing these critical weaknesses, the utility provides a secure and efficient solution for file encryption and decryption in today's digital landscape.

Chapter 2

Methodology

Steps in the Complete ChaCha20 Implementation :

2.1 Key and Nonce Generation

1. Key derivation : Transforms a user-provided password into a secure cryptographic key suitable for ChaCha20, which requires a 256-bit key. This transformation is achieved using a Key Derivation Function (KDF), which ensures that the derived key is both strong and resistant to attacks.

A KDF takes the input password and processes it with additional parameters, such as a salt and an iteration count. The salt is a random value added to the password before hashing, ensuring that identical passwords yield different keys. This prevents attackers from using pre-computed tables, like rainbow tables, to crack passwords efficiently.

The iteration count specifies how many times the KDF applies its hashing algorithm, increasing the computational effort required to derive the key. This makes brute-force attacks more challenging, as attackers must perform the KDF for each password guess multiple times.

Common KDFs include PBKDF2, bcrypt, and Argon2, each designed to enhance security through these mechanisms. By deriving a secure key from a password, the ChaCha20 encryption process can effectively protect sensitive data against unauthorized access. Proper key management practices are essential to maintain the confidentiality and integrity of the derived key throughout its lifecycle.

2. Nonce generation is a vital component of the encryption process, particularly in stream ciphers like ChaCha20. A nonce, or "number used once," is a unique value that is generated for each encryption operation. In the case of ChaCha20, the nonce is 12 bytes (96 bits) in length, providing a sufficiently large space to ensure uniqueness across multiple encryption sessions.

The primary purpose of the nonce is to ensure that the same plaintext encrypted multiple times with the same key produces different ciphertexts. This is crucial for maintaining the security of the encryption scheme. If a nonce is reused with the same key, it can lead to vulnerabilities, such as revealing patterns in the ciphertext and potentially allowing attackers to recover the plaintext through various cryptanalytic techniques.

To generate a secure nonce, it is essential to use a reliable source of randomness. Cryptographic random number generators (CSPRNGs) are typically employed to produce nonces that are unpredictable and uniformly distributed. This randomness is critical, as predictable nonces can compromise the security of the encryption.

In practice, the nonce is often stored alongside the ciphertext, allowing the decryption process to retrieve it easily. By ensuring that each nonce is unique and never reused with the same key, the integrity of the ChaCha20 encryption process is upheld, significantly enhancing the overall security of the encrypted data. Proper nonce management is essential to prevent accidental reuse, which could lead to severe security vulnerabilities. This nonce should never be reused with the same key to maintain security.

2.2 Initialization

1. The state setup is a critical step in the ChaCha20 encryption process, as it establishes the initial conditions for generating the keystream. The ChaCha20 state consists of 16 32-bit words, organized into a specific structure that combines fixed constants, the encryption key, the nonce, and a counter.
 - (a) Fixed Constants: The first four words of the state are fixed constants that serve as identifiers for the ChaCha20 algorithm. These constants are predefined values that help distinguish the ChaCha20 cipher from other algorithms, ensuring that the initialization process is consistent and recognizable.
 - (b) Key Integration: The next eight words of the state are derived directly from the 256-bit key. Each 32-bit segment of the key is loaded into the state, providing the necessary cryptographic strength for the encryption process. This integration ensures that the key influences the keystream generation, making it unique to the specific encryption session.
 - (c) Nonce and Counter: The final four words of the state are derived from the nonce and a counter. The nonce, which is unique for each encryption operation, is split into three 32-bit segments, while the counter, initialized to zero, is included as the last word. The counter is incremented with each block processed, ensuring that the state changes with each encryption operation and preventing the reuse of the same keystream. By combining these elements, the ChaCha20 state is initialized in a way that maximizes security and efficiency. This structured approach allows the algorithm to produce a high-quality keystream that is resistant to cryptanalytic attacks. The proper setup of the state is essential for the subsequent keystream generation, which directly impacts the confidentiality of the encrypted data.

2.3 Block Processing

1. Keystream generation is a fundamental aspect of the ChaCha20 encryption process, where the initialized state undergoes a series of transformations to produce a pseudorandom keystream. This keystream is then used to encrypt plaintext by performing an XOR operation with the data.

- (a) Quarter-Round Operations: The core of the keystream generation involves a series of quarter-round operations, which are designed to mix the state variables effectively. Each quarter-round operation takes four 32-bit words from the state and applies a sequence of arithmetic and bitwise operations, including addition, XOR, and rotation. This mixing process ensures that the output is highly dependent on all input values, enhancing the diffusion properties of the algorithm.
 - (b) Loop Iterations: The ChaCha20 algorithm performs a total of 20 rounds of these quarter-round operations, divided into 10 iterations. In each iteration, the state is processed in a specific order, with the quarter-round operations applied to different combinations of state variables. This repeated mixing significantly increases the complexity of the output, making it resistant to cryptanalysis.
 - (c) Keystream Output: After completing the quarter-round operations, the final state is added to the original state used for initialization. This addition ensures that the keystream is unique to the specific key, nonce, and counter combination. The result is a 64-byte keystream, which is then used in the encryption process. The keystream is generated on-the-fly, allowing for efficient encryption of data in blocks, without the need to precompute large amounts of data. This is done in a loop to produce a keystream of 64 bytes.
2. The XOR operation is a crucial step in the ChaCha20 encryption process, where the generated keystream is combined with the plaintext data to produce the final ciphertext. This operation leverages the properties of the XOR (exclusive OR) function, which is fundamental to the security of stream ciphers.
- (a) Combining Keystream and Plaintext: During encryption, the 64-byte keystream generated by the ChaCha20 algorithm is XORed with the plaintext data in 64-byte blocks. The XOR operation is performed on each corresponding byte of the keystream and plaintext, resulting in a ciphertext that is a direct transformation of the original data. Mathematically, this can be expressed as: $\text{ciphertext}[i] = \text{plaintext}[i] \oplus \text{keystream}[i]$ where \oplus denotes the XOR operation.
 - (b) Security Through Randomness: One of the key advantages of using the XOR operation in this context is that it ensures that the same plaintext encrypted multiple times with the same key and nonce will yield different ciphertexts. This is due to the unique nature of the keystream generated for each encryption session. As a result, even if the same plaintext is processed, the output ciphertext will differ, thereby enhancing security and preventing attackers from identifying patterns in the encrypted data.
 - (c) Reversibility for Decryption: The XOR operation is inherently reversible, which is a critical property for decryption. To recover the original plaintext, the same keystream used for encryption is XORed with the ciphertext: $\text{plaintext}[i] = \text{ciphertext}[i] \oplus \text{keystream}[i]$. This property allows for efficient and straightforward decryption, making the ChaCha20 algorithm both secure and practical for real-world applications.

2.4 Message Authentication

The Poly1305 Message Authentication Code (MAC) is a critical component of the ChaCha20 encryption process, providing a mechanism to ensure the integrity and authenticity of the encrypted data. This step is essential for verifying that the ciphertext has not been altered and that it originates from a legitimate source.

1. **MAC Generation:** After the encryption process is complete, the Poly1305 algorithm is employed to generate a MAC. This MAC is computed over the ciphertext and any associated data (AAD) that may be included in the encryption operation. The use of AAD allows for additional context to be authenticated without being encrypted, which is useful in various applications where metadata needs to be verified.
2. **Key Usage:** The same key used for the ChaCha20 encryption is also utilized in the Poly1305 MAC computation. This ensures that the MAC is tightly coupled with the encryption process, enhancing security. The MAC is derived from the ciphertext and the associated data, producing a fixed-length output (typically 16 bytes) that serves as a cryptographic signature for the data.
Integrity and Authenticity: The primary purpose of the Poly1305 MAC is to provide assurance that the ciphertext has not been tampered with during transmission or storage. When the recipient receives the ciphertext, they can compute the MAC using the same key and compare it to the received MAC. If the two MACs match, it confirms that the data is intact and authentic. If they do not match, it indicates potential tampering or corruption, prompting the recipient to reject the data.
3. **Appending the MAC:** The computed MAC is appended to the end of the ciphertext, allowing the recipient to verify the authenticity of the message upon decryption.

2.5 Decryption Process

1. **Nonce Retrieval:** For decryption, the nonce is read from the beginning of the ciphertext.
2. **State Reinitialization:** The ChaCha20 state is reinitialized using the same key, nonce, and counter as during encryption.
3. **Keystream Generation:** The keystream is generated again using the same process as in encryption.
4. **XOR Operation:** The ciphertext is XORed with the keystream to recover the original plaintext.
5. **MAC Verification:** The MAC is extracted from the end of the ciphertext, and the Poly1305 MAC is computed over the decrypted plaintext and associated data. If the computed MAC does not match the extracted MAC, the decryption fails, indicating potential tampering.

Chapter 3

Results and Discussion

Chapter 4

Conclusion and Future Work

Bibliography

- [1] Daniel J Bernstein et al. Chacha, a variant of salsa20. In *Workshop record of SASC*, volume 8, pages 3–5. Citeseer, 2008.