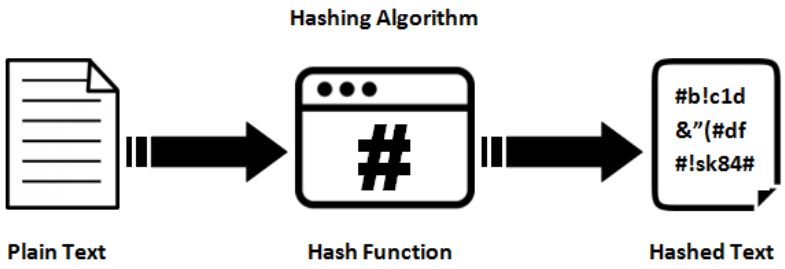
Vladimir Avdeev

SHA-1:

Introduction:

Secure Hashing Algorithm is what SHA stands for and the 1 stands for the first version. Now what is hashing and what is its significance. Hashing allows you to take in an input string and apply a mathematical formula splitting this string up into numerous pieces. Before it was used widely as a security method it was used to index through large amounts of data. A good way to think of this is thinking of a word you want to define in a dictionary book. Now thinking of hashing in a security standpoint we apply difficult formulas making it difficult to reverse the original string.

Example of Hashing:



Hashing takes in a string input then returns a fixed length output. Hashing is a highly effective method of securing data by converting it into a fixed-length, unique string (hash) that cannot be reversed to its original form. It ensures data integrity, as even minor changes to the input produce a completely different hash. Hashing is widely used for password storage, data verification, and digital signatures due to its irreversible and secure nature.

For example SHA-256 will always return a 64 character hash (or string). Hashing allows data to be secure by making it difficult to revert to the original string. In the real world we can

commonly see this used as a storage method of passwords. Where only the hash of the password is being stored instead of the raw text of the password. An example on how this works with authentication is that environments will compare hashes instead of the raw text of the password.

History of SHA-1:

SHA-1 was developed by the US government back in 1993 and was a project that was developed by US government standard agencies NIST. Later the 1993 version was superseded and it's now called SHA-0 and was replaced in 1995 with a small change. The only change was a single bitwise rotation in the message schedule. The reason for this was that the original version contained a flaw which affected “cryptographic security”. [8]

SHA-1 was officially removed from major products as the method of hashing data was now deemed insecure. Its successor (SHA-256) is still commonly used, the main difference being that the hash sizes are larger. The biggest worry with smaller hash sizes is collisions which is when two different inputs create the same hash.

SHA-1 was widely used for many applications and protocols before it was superseded. It was used widely for TLS, SSL,PGP,SSH,S/MIME, and IPsec. This method of hashing was required by law for certain US government applications. This hashing algorithm played a key role in the security infrastructure and still is with its successors.

Intuition:

Hashing Functions:

I decided to conduct research on two different hashing algorithms, one allowing us to go over the concept in a little bit more simpler perspective. And the other being more sophisticated giving more of a real world example of how hashing really works.

DJB2:

The DJB2 hash function is a great example of hashing, where raw text is converted into a numeric representation that cannot be reversed. It initializes a hash value to 5381 and processes each character by shifting the hash 5 bits to the left (effectively multiplying by 32), adding the original hash (to make it a multiplication by 33), and then adding the ASCII value of the character. This simple formula efficiently produces a unique (but not cryptographic) representation of the input text.[6]

Bitwise Operations:

Before we begin explaining the steps of how SHA-1 works I think it would be useful to go over some bitwise operations. SHA-1 uses bitwise AND, OR, XOR, NOT, and circular shift operations. Which are used to mix bits, maintain fixed block sizes, and utilize non-linear transformations. [9]

AND : Returns true if both conditions are true.

OR : Returns true if atleast one condition is true.

NOT: Negates the value. True becomes false vice versa.

XOR : Returns true if exactly one condition is true.

Left shift : Shifts bits to the left by filling the vacant bits with 0's.

AND operation	OR operation	NOT operation	XOR operation	Left shift
1 & 1 = 1 1 & 0 = 0 0 & 0 = 0	1 | 1 = 1 1 | 0 = 1 0 | 0 = 0	~ 1 = 0 ~ 0 = 1	1 ^ 1 = 0 1 ^ 0 = 1 0 ^ 0 = 0	1111 << 2 = 111100

SHA-1:

Sha-1 first works by taking in a string input then begins to pad the message. Padding the message is "to ensure that length is congruent to 448 modulo 512"[5]. What is padding? Padding is the processing step where the input message is suitable for an algorithm to process the input. Allowing the algorithm to divide the message into 512 bit blocks. Then to "divide into 16 words of 32 bits" [5]. Then these words will be "expanded into 80 32-bit words"[5]. After they are split into 80 words then SHA-1 uses five constant values which are applied to a specific range of rounds (0-79). After all rounds the values are all added into their original hash values in order to produce the final hash.

Steps of How SHA-1 Works:

1. First the message needs padding the message by adding 0's until it's a multiple of 512 (or 64 bytes). Then append the original length of the message at the end.
2. Then we will divide the padded message into 16 32 bit "words"
3. Extend the 16 32 bit words to 80 32 bit words. First we are just initializing the 16 words from the message block. By extending these words we will run XOR operations on words before the word we are iterating on and then rotate the bits left by 1 bit.

   ```
   W[i] = RotateLeft( W[i - 3] XOR W[i - 8] XOR W[i - 14] XOR
   W[i - 16], 1)
   ```

4. We set up the five working variables with the given five constant values; A=H0 ,B=H1, C=H2, D=H3, and E=H4. We are cycling through 80 rounds applying different operations to each word. For this we have a round function 'F' which determines how intermediate hash values B,C,D are combined each round. As well we have a different constant value of 'K' which is applied differently to each group of 20 rounds.

Values of K ;

Rounds 0-19: K=0x5A827999

Rounds 20-39: K=0x6ED9EBA1

Rounds 40-59: K=0x8F1BBCDC

Rounds 60-79: K=0xCA62C1D6

Function of F:

Rounds 0-19 : `F = (B & C) | ((~B) & D)`

Rounds 20-39: `F = B ^ C ^ D`

Rounds 40-59: `F = (B & C) | (B & D) | (C & D)`

Rounds 60-79: `F = B ^ C ^ D`

Key: & = and , | = or , ~ = not, ^ = XOR

For each round we store these values into a temp value. F adds the logic and complexity and K introduces the round specific constants. By combining this with words (word[iteration]) this ensures that hash computation is resistant to prediction.

```
TEMP = RotateLeft( A, 5) + F + E + W[iteration] + K
```

```
E = D , D = C , C =  RotateLeft( B, 30), B = A, A = TEMP
```

5. After 80 rounds we get the final values for A,B,C,D,E. We now update our constant the five constant values of H by adding the final values. H0 += A ,H1 += B, H2 += C, H3 += D, and H4 += E. After we add to all the blocks H0-H4 contain all the final hash values. To create the final hash we concatenate them into one 160-bit single value. Concatenate means joining each binary value end to end.

Pseudocode:

DJB2:

```
FUNCTION djb2Hash(inputString):
    hash = 5381 // Initial seed value
    FOR each character c IN inputString:
        hash = (hash << 5) + hash + c // Equivalent to hash * 33 + c
 RETURN hash
```

Input

inputString: A string of characters to be hashed. Each character will be processed using its ASCII value.

Output

hash: An integer hash value, which is the result of applying the **djb2 hash algorithm** to the input string.

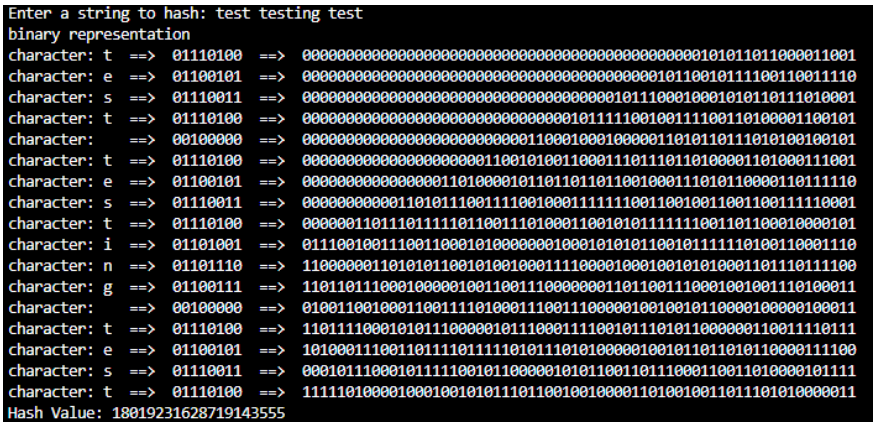
Algorithm Description

1. **Initialize the Hash**:
    - The hash is initialized to a fixed seed value 5381.
2. **Iterate Over the Characters**:
    - For each character c in the input string:
        - Compute the ASCII value of c.
        - hash = (hash << 5) + hash + c
3. **Return the Hash**

Argument for Correctness

1. **Efficient Hashing:**
    - The algorithm uses a fixed multiplier (33) that has been found to provide good distribution for hash values with minimal collisions.

Example of execution:



SHA-1:

padded_message:

```
func padded_message(message):
    padded_message = []  // Initialize an empty list for the padded message

    // Append the original message to the padded message
    padded_message += message
```

```
    // Append a single '1' bit (0x80 in hexadecimal is 10000000 in binary)
    padded_message += 0x80

    // Add zero bytes (0x00) until the message length is 64 bits shy of a multiple of 512
    while ((size(padded_message) * 8) % 512 != 448):
        padded_message += 0x00

    // Append the original message length as a 64-bit big-endian integer
    original_length = size(message) * 8  // Calculate the length of the original message in
bits
    for i from 7 to 0:  // Iterate from 7 to 0 for the 8 bytes of the 64-bit integer
        padded_message += (original_length >> (i * 8)) & 0xFF  // Extract each byte

    return padded_message
```

Input

- message: A list of bytes (or an equivalent iterable), representing the original message to be padded. Each element of the message should have a value in the range [0, 255].

Output

- padded_message: A list of bytes that includes:
    1. The original message.
    2. A single 1 bit followed by zero bits to make the length 64 bits shy of a multiple of 512.
    3. The original message length encoded as a 64-bit big-endian integer.

Algorithm Description

1. **Initialize the Padded Message**:
    - Start with an empty list (padded_message) and copy the original message into it.
2. **Append the '1' Bit**:

- Append the hexadecimal value 0x80 to signify the binary sequence 10000000 (a single '1' bit followed by seven '0' bits). This marks the end of the message.

3. **Add Padding Bytes**:
    - Use a while loop to append 0x00 bytes (binary 00000000) to the message until the total bit-length of padded_message is 64 bits shy of a multiple of 512.
4. **Encode the Original Message Length**:
    - Calculate the original message length in bits (size(message) * 8).
    - Split this 64-bit value into 8 bytes (big-endian order) by iterating from the most significant byte to the least significant byte and appending them to padded_message.
5. **Return the Padded Message**

Example of execution:

```
Enter a message: test testing sha-1!
Padded Message (Binary):
01110100 01100101 01110011 01110100 00100000 01110100 01100101 01110011 01110100 01101001
01101110 01100111 00100000 01110011 01101000 01100001 00101101 00110001 00100001 10000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 10011000
```

parse_message:

```
func parse_message(padded_message):
    blocks = []  // Initialize an empty list to store 32-bit words

    // Iterate over the padded_message in chunks of 4 bytes (32 bits)
    for i from 0 to size(padded_message) - 1 step 4:
        block = 0  // Initialize the block as 0
```

```
    // Combine 4 consecutive bytes into a single 32-bit word
    for j from 0 to 3:
        block = (block << 8) | padded_message[i + j]

    blocks += block  // Add the 32-bit word to the list of blocks

return blocks
```

Detailed Description

Input

- padded_message: A list of bytes (or equivalent iterable) representing a padded message. Each element in the list should be an integer in the range [0, 255].

Output

- blocks: A list of 32-bit words. Each word is an integer formed by combining 4 consecutive bytes (32 bits) from the input padded_message.

Algorithm Description

1. **Initialize the Blocks List**:
    - Create an empty list block to store the 32-bit words that will be extracted from the padded_message.
2. **Iterate Over the Padded Message**:
    - Use a loop to process the padded_message in chunks of 4 bytes (32 bits) at a time.

- For each iteration, the loop starts at i and processes bytes at indices i, i+1, i+2, and i+3.

3. **Combine 4 Bytes Into a 32-Bit Word**:
    - For each chunk of 4 bytes:
        - Start with block = 0.
        - Use a nested loop to process each byte in the chunk:
            - Shift block left by 8 bits (block << 8) to make space for the next byte.
            - Combine the next byte into block using the bitwise OR (|) operation.
4. **Append the 32-Bit Word to Blocks**:
    - Once the 4 bytes have been combined into a 32-bit word, append the resulting block to the blocks list.
5. **Return the List of Blocks**

Example of Execution:

```
Message Block (Binary):
00000000000000000000000001110100 00000000000000000000000001110011
00000000000000000000000001100111 00000000000000000000000001100001
00000000000000000000000010000000 00000000000000000000000000000000
00000000000000000000000000000000 00000000000000000000000000000000
00000000000000000000000000000000 00000000000000000000000000000000
00000000000000000000000000000000 00000000000000000000000000000000
00000000000000000000000000000000 00000000000000000000000000000000
00000000000000000000000000000000 00000000000000000000000010011000
```

leftRotate:

```
func leftRotate(value , shift):
    return (value << shift) | (value >> (32 - shift))
```

sha1:

```
func sha1(message):
    // Step 1: Preprocessing
    padded_message = pad_message(message)
    blocks = parse_message(padded_message)

    // Step 2: Initialize hash values (H0, H1, H2, H3, H4)
    H0 = 0x67452301
    H1 = 0xEFCDAB89
    H2 = 0x98BADCFE
    H3 = 0x10325476
    H4 = 0xC3D2E1F0

    // Step 3: Process each 512-bit block
    for block in blocks:
        // Step 3.1: Prepare the message schedule (W)
        W = array[80]
        for i from 0 to 15:
            W[i] = block[i]  // First 16 words come from the block
        for i from 16 to 79:
            W[i] = leftRotate(W[i-3] XOR W[i-8] XOR W[i-14] XOR W[i-16], 1)

        // Step 3.2: Initialize working variables (a, b, c, d, e)
        a = H0
        b = H1
        c = H2
        d = H3
        e = H4

        // Step 3.3: Perform the main compression loop
```

```
        for i from 0 to 79:
            if i between 0 and 19:
                f = (b AND c) OR ((NOT b) AND d)
                k = 0x5A827999
            else if i between 20 and 39:
                f = b XOR c XOR d
                k = 0x6ED9EBA1
            else if i between 40 and 59:
                f = (b AND c) OR (b AND d) OR (c AND d)
                k = 0x8F1BBCDC
            else if i between 60 and 79:
                f = b XOR c XOR d
                k = 0xCA62C1D6

            temp = (leftRotate(a, 5) + f + e + k + W[i]) & 0xFFFFFFFF
            e = d
            d = c
            c = leftRotate(b, 30)
            b = a
            a = temp

        // Step 3.4: Add the working variables back into the hash values
        H0 = (H0 + a) & 0xFFFFFFFF
        H1 = (H1 + b) & 0xFFFFFFFF
        H2 = (H2 + c) & 0xFFFFFFFF
        H3 = (H3 + d) & 0xFFFFFFFF
        H4 = (H4 + e) & 0xFFFFFFFF

    // Step 4: Produce the final hash value
    return concatenate(H0, H1, H2, H3, H4) as hexadecimal string
```

Detailed Description

Input

- message: A sequence of bytes representing the input message to be hashed.

Output

- A 160-bit SHA-1 hash value, represented as the concatenation of five 32-bit integers .

Algorithm Description

1. **Preprocessing**:
    - **Padding**: Pad the message to ensure its length (in bits) is a multiple of 512.
    - **Parsing**: Split the padded message into 16 32-bit blocks.
2. **Initialization**:
    - Set the initial hash values (H0, H1, H2, H3, H4) as constants defined by the SHA-1 standard.
3. **Process Each Block**:
    - **Prepare the Message Schedule**:
        - Create an array W of size 80.
            - Process was explained in depth in the intuition.
    - **Initialize Working Variables**:
        - Set a = H0, b = H1, c = H2, d = H3, e = H4.
    - **Main Loop**:
        - For each of the 80 rounds (i = 0 to 79):
            - Compute the function f and constant k based on the round index
                - The process of computing values on each round is

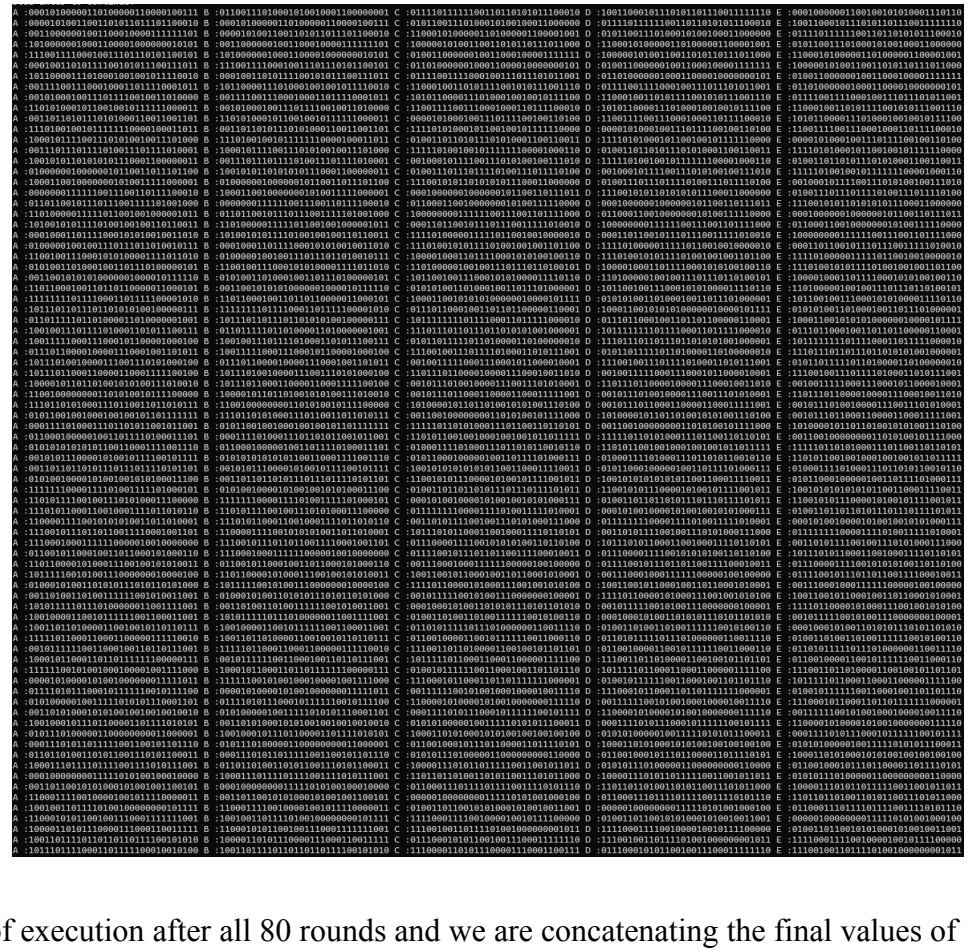
explained in the intuition.

4. **Produce the Final Hash**:
    - Concatenate H0, H1, H2, H3, H4 to produce the final 160-bit hash.

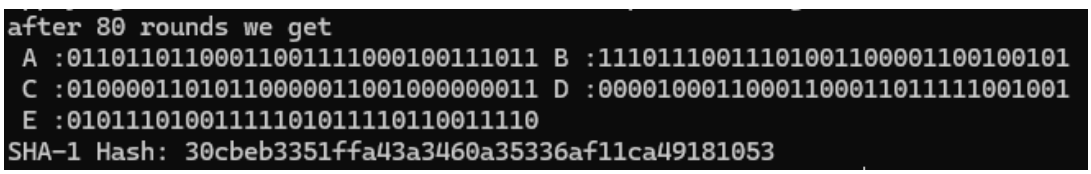
Argument for Correctness

1. **Adheres to the SHA-1 Standard**:
    - The algorithm follows the structure and steps defined by the SHA-1 cryptographic hash standard, including padding, message scheduling, and the main processing loop.
2. **Handles Any Input**:
    - The preprocessing step ensures the input is correctly padded, making it compatible with the SHA-1 block size requirements.
3. **Secure Message Scheduling**:
    - The schedule ensures that each word in the block contributes to the final hash, even for long inputs, by using XOR and left rotation.
4. **Iterative Hashing**:
    - Each block updates the hash values incrementally, preserving all information from previous blocks.
5. **Comparing Hashes:**
    - If we run SHA-1 algorithm on two different occasions with the same input string. We should in theory get the same output hash as we are performing the
    - same operations on both strings.

Example of execution of 80 rounds of operations performed on values A,B,C,D E:



Example of execution after all 80 rounds and we are concatenating the final values of



Asymptotic Runtime Analysis:

DJB2:

Runtime: **O(n)**

This is because we only have one loop that is iterating through the size of the string.

SHA-1:

Runtime: **O(n)**

For the three main functions that are used in this algorithm we are iterating through a given size and the main function iterates to the size of the input string. The *padMessage* function has a while loop that iterates until the size of the array is a multiple of 512. Then *parseMessage* iterates to the given size of the padded message made by the function *padMessage.* The main function *SHA-1* will iterate to the size of the parsed message.

Sources:

1. https://tresorit.com/blog/the-history-of-encryption-the-roots-of-modern-day-cyber-security/#:~:text=The%20first%20encryption%20was%20the,with%20the%20sender%20and%20receiver.
2. https://spectrum.ieee.org/hans-peter-luhn-and-the-birth-of-the-hashing-algorithm
3. https://www.clickssl.net/blog/difference-between-hashing-vs-encryption#:~:text=Hashing%20vs%20Encryption%20%E2%80%93%20Hashing%20refers,the%20reach%20of%20third%20parties.
4. https://en.wikipedia.org/wiki/SHA-1
5. https://www.geeksforgeeks.org/sha-1-hash-in-java/
6. http://www.cse.yorku.ca/~oz/hash.html
7. https://chatgpt.com/
8. https://en.wikipedia.org/wiki/SHA-1#:~:text=SHA%2D1%20was%20developed%20as,Institute%20of%20Standards%20and%20Technology).
9. https://www.geeksforgeeks.org/bitwise-operators-in-c-cpp/