ŤUDelft

ID-based self-encryption via Hyperledger Fabric based smart contract

Ilya Grishkov
Supervisor(s): Kaitai Liang, Roland Kromes
EEMCS, Delft University of Technology, The Netherlands
22-6-2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

This paper offers a prototype of a smart-contract-based encryption scheme meant to improve the security of user data being uploaded to the ledger. A new extension to the self-encryption scheme was introduced by integrating identity into the encryption process. Such integration allows to permanently preserve ownership of the original file and link it to the person who originally encrypted it. Moreover, self-encryption provides strong security guarantees under the condition that the encrypted file and the key are safely stored.

1 Introduction

The modern world is increasingly adopting blockchain technology. The first significant market adoption of blockchain happened in 2009 when Bitcoin was introduced [1]. Interest in blockchain solutions grew over the years and led to the invention of Ethereum - Bitcoin peer but with support for smart contracts [2]. The introduction of smart contracts leads to further development in the field of blockchain and creates demand for more industry-friendly solutions that allow identifying users of the system (Know-Your-Customer, Anti-Money-Laundering). Hyperledger Fabric was introduced as a highly modular permissioned blockchain that allows great customization to suit particular industrial needs [3]. Given its customizability and modularity, Hyperledger Fabric (HLF) is a perfect platform for extending it with various trust and privacy preservation solutions.

According to Huang et al. [4] the main component of a blockchain being attacked the most is a smart contract. A high frequency of attacks on a component designed to handle private user data suggests a need for an alternative approach to handling sensitive information other than just sending it raw to the ledger. Local data encryption prior to sending data to the smart contract could be a solution.

Self-encryption was introduced as a means of encrypting files that "requires no user intervention or passwords" [5]. This algorithm can be used for local encryption of files, encrypted chunks of which will be later uploaded to the distributed file system. Pointers to the encrypted chunks are then sent to the ledger. While this solution allows keeping file content private, the file itself is not linked to its owner. A variant of identity-based encryption can tackle this problem. If a file is self-encrypted with the owner's identity used during the encryption process, this file remains linked to the person who initially uploaded it to the blockchain. This way, original ownership can be preserved.

This paper aims to explore trust and privacy-preserving solutions in the Hyperledger Fabric blockchain. More specifically, the goal is to further investigate the utility of a combination of identity-based encryption and self-encryption to improve data security in the HLF; extend the previously done research by Park [6] and implement ID-based self-encryption via Hyperledger Fabric smart contract. Hence the main research question is: "How can the security of Hyperledger Fabric smart contracts be improved using ID-based self-encryption?"

Within this paper, an approach to integrating ID-based self-encryption is presented. Moreover, a detailed description of the prototype implementation is given.

This work is structured as follows. Section 2 describes the methodology used to achieve the goal. Section 3 discusses the inner workings of ID-based self-encryption as a means of increasing the security of Hyperledger fabric. Section 4 discusses security implications and presents the analysis of the implemented algorithm. Section 5 concludes the work.

2 State of the Art

The topic of security and privacy of the Hyperledger Fabric has been thoroughly studied [7], [8], [9]. Moreover, research has been conducted this year by a student of the Delft University of Technology [6] addressing the similar issue of improving HLF security using self-encryption.

The concept of self-encryption was introduced by Yu Chen [10]. The approach of the original paper involves converting a file into a bit stream, extracting the key by randomly selecting bits from the stream, and then doing the encryption using that key. After the encryption, the key and the encrypted file should be stored separately, e.g., the key can be stored locally, while the encrypted file can be sent to a server.

The original encryption scheme was also extended by Moch Rezky Debby Rahardjo [11]. According to the paper, "The modification is located in dividing the plaintext and ciphertext into 1024-bit chunks at XOR process and using the date when encryption process starts as a seed. The modification also adds the database for the key management function". Storing the key and the encrypted chunks in separate places makes it computationally not feasible to get the original data.

The later industrial adaptation of the self-encryption scheme happened when a team led by David Irvine made self-encryption the core of his company's (MaidSafe) product - SAFE Network [5]. Irvine's implementation of the self-encryption scheme will be the basis of this work. Hence a more detailed explanation of the implementation of the algorithm will be given.

Figure 1 shows the encryption process. First, the original file is split into a minimum of 3 file chunks. After the file is split into chunks, the algorithm creates a data map, where the key needed for decryption will be stored. Each chunk is then hashed, and those hashes are written to the data map. Parts of those hashes are used as a key and initialization vector for AES 128 algorithm that encrypts each file chunk. When encryption is done, each encrypted chunk is obfuscated with the previously computed hash values by applying the XoR function. At the end of the process, the encryption scheme returns encrypted file chunks and a data map that contains keys for decryption.

3 Methodology

The original implementation of the self-encryption schema by David Irvine [12] was modified and used for this research. The use of the rayon library (which adds parallelization to the code) was removed from the algorithm since the compilation target (WebAssembly) only supports single-threaded code. Additionally, the code base was modified to include an
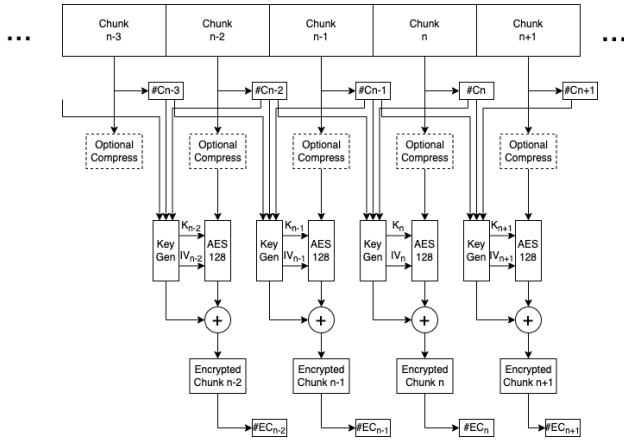
Figure 1: Self-encryption process [12]

interface for communication with the external code. Changes were also done to the Cargo.toml to make the code compatible with the target. The modified self-encryption algorithm was compiled to WebAssembly and run in a virtual machine (VM), and invoked from the code of the developed local application (which allows the interaction with the Hyperledger Fabric Smart Contract). A more detailed description of the process will be given in subsection 4.3. The benchmarks of this implementation will be provided in section 5.

Hyperledger Fabric test network v2.4.3 was used. Test network was deployed to Docker with the following command[1]:

```
$ ./network.sh up createChannel –ca
```

This command creates a new test-network with a single channel and also uses Certificate Authorities. A smart contract was then deployed (detailed in subsection 4.3).

Encrypted file storage is handled by the InterPlanetary File System (IPFS)[2], which is a distributed Torrent database that uses hashes of files to address its content. IPFS node was also deployed to Docker. For IPFS deployment, two directories (staging and data) were mounted on the host file system to persist the stored data when the container is stopped. Hyperledger Fabric provides an official software development kit (SDK) for three languages: Go, Java, and Javascript. Go was chosen for the implementation due to the ease of integration with Hyperledger Fabric and the IPFS. The encryption library is written in Rust and is compiled to WebAssembly; hence a way to call WebAssembly was needed. Go also provides support for the Wasmer library that calls the WebAssembly function directly from Go code.

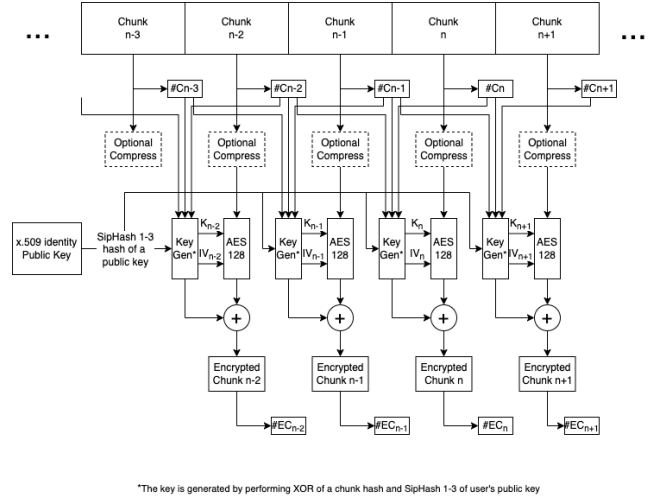## 4 ID-based self-encryption

4.1 Integrating identity into the encryption

This paper offers an extension to the algorithm proposed by Irvine [5]. The encryption step in the original algorithm is

modified to include the identity of a person running the algorithm in the encryption process. Instead of using part of the chunk hash as a key for AES 128, the result of the XoR of the hashed identity and the chunk hash is used as a key. The identity can be any string of any length. If the length of this string is shorter than the length of the key, the cycle function is applied to the string, which repeats the iterator of the string. The hashing function SipHash 1-3 is used to hash the identity of a user before passing it to the XoR function. Figure 2 demonstrates the process of encrypting a file using the modified version of self-encryption with identity integrated into the encryption process.

*The key is generated by performing XOR of a chunk hash and SipHash 1-3 of user's public key

Figure 2: ID-based self-encryption process

The decryption of the file encrypted using ID-based self-encryption is similar to that of a regular self-encryption, with the key for AES 128 being the only different part. The decryption calculates the key the same way the encryption does it by applying the XoR function to the hash of identity and the chunk hash from the data map.

The implementation of the encryption scheme can be found on GitHub [3]

4.2 Connecting the encryption algorithm and the local application

The implementation of the identity-based self-encryption is written purely in Rust, while the rest of the project is written in Go. This approach demands a way to integrate the Rust library into Go code. Among the solutions to tackle the problem are:

1. Use Go tools to assemble the Go code and compile Rust code into a static library. Then link compiled code using additional assembly "glue-code" [14].

2. Compile Rust to a static library and call it from the Go code using Go build-in pseudo-library C for interacting with native interfaces.

---

3. Compile Rust to WebAssembly (WASM) code and call it from Go using Wasmer library [1].

All of the methods have been successfully tried. The first two methods do not allow cross-compilation because both require compiling Rust to a static library, which is platform-specific. Additionally, the first method requires the use of assembly language, which is different on different processor architectures and operating systems. The second method also uses the C pseudo-library, which does not allow cross-compilation of the Go code. Overall, both methods are very *platform-specific*, making them a less preferable choice.

The third method was chosen for connecting the Rust library to Go code. Compiling Rust to WASM to use as a standalone application or a library can be done using the following command:

```
$ cargo build --target=[chosen_target]
```

where *chosen_target* is a WebAssembly target that can be either wasm32-unknown-unknown or wasm32-wasi. The latter was used because it compiles using WASI API [2], which is a system API that provides access to multiple operating system functionalities, such as access to the file system.

The resulting WASM file is then placed in a hidden folder in the user's home directory, so the Go code can later load it. As the WASM code is used within a virtual machine (VM), it is independent of the operating system it will run on, so it requires compiling only for one target.

Calling WASM from Go using Wasmer

In order to call the WASM code, a VM needs to be used. Wasmer library provides such VM that can also be initialized from within Go code. The process of calling WASM code requires main steps.

1. Load WASM code into a Wasmer VM

   (a) A directory on the host operating system that will be accessible in the VM needs to be specified

   (b) Optionally, the standard output of the WASM library can be inherited.

2. Invoking a function by its name and passing arguments to it

The communication between the Go code and WASM library and passing arguments for function invocation is happening using C types, which means that types like strings are not supported directly and need to be converted to corresponding C types. If a string is passed as an argument, it needs to be written to memory and end with a zero byte. The pointer to the first byte of this string is then passed to the invoked function as an argument.

As the host operating system memory is inaccessible to the VM, allocation and deallocation of memory must happen inside the VM itself. In order to facilitate the allocation and deallocation, two dedicated Rust functions were developed as a part of the id-based encryption library interface: *allocate* and *deallocate*.

---

[1]https://wasmer.io/

[2]https://wasi.dev/

If the called function requires a string as an argument, the allocation must be performed before passing the pointer to that string. The allocate function has to be called to allocate memory inside the VM. The memory is then accessed from the Go code, and each byte of the string argument is written to the newly allocated memory. The pointer to the memory and the length must be preserved to deallocate the memory before the program terminates. The pointer to the first memory cell containing the string argument is then passed as an argument to the function being called.

Wrapper code for Wasmer calls

A wrapper code has been written to simplify the invocation of WASM functions. The significant simplification this code provides is the ability to pass Go native-type argument to the wrapper, performing all the necessary processing and allocation if needed. The pointers to string or array types and their lengths are stored, so when the program terminates, the memory is getting deallocated.

Moreover, the developed wrapper code allows to pass simple numerical Go types (integers, floats, bytes, etc.) as pointers to the WASM library, so the changes happening to them when WASM functions run are also reflected in Go code, without the need to return anything.

Additionally, the wrapper requires a return type parameter argument (represented as an enumerator) when calling the invocation function through the wrapper. It uses the return type to case the return of the WASM function to the corresponding Go type. In cases when a pointer to a string is returned, the wrapper reads bytes from the VM memory until the zero byte and creates a Go string from it. The return type of the wrapper's invocation function is a generic *interface{}*, which requires additional type casting. For example, if the called function returned a pointer to a string, a Go string will be built from the pointer, but a user will still have to dynamically convert the returned value as it will be *interface{}*.

4.3 Smart contract

The smart contract in Hyperledger Fabric allows for defining assets on the ledger. This paper defines an asset containing three fields: ID, Owner, and CID. The code below shows the definition of an asset written in Go.

```
type Asset struct {
        ID string `json:"ID"`
        Owner string `json:"Owner"`
        CID []string `json:"CID"`
}
```

The ID is a universally unique identifier (UUID) generated when the new asset is created. The Owner is a string of hexadecimal numbers representing a public key of a user who created the asset. The CID is an array of unique immutable identifiers referencing encrypted file chunks saved in the IPFS. The references are used for retrieving the stored data and can also be used for verifying if it was manipulated (the hash of data is a unique value).

Additionally, the smart contract defines a list of functions for creating, deleting, and updating assets. The implementa-

tion can be found on GitHub[1].

4.4 Design implementation

The encryption and decryption process and interactions with the IPFS and the Hyperledger Fabric are orchestrated by a local application, a command-line interface (CLI) tool written in Go. The prototype of the tool is accessible from GitHub[2]

Executing any command starts with creating a new instance of a WASM wrapper and loading the encryption library. When the command requires interaction with the Hyperledger Fabric, the presence of the wallet containing identity (which is necessary to enable interaction with the smart contract) is being checked. If the wallet is missing, it is populated based on a user's certificates and keys. When this preparation is done, the execution of the command starts.

At the end of the program execution, the wrapper iterates over all allocated memory pointers and individually deallocates them.
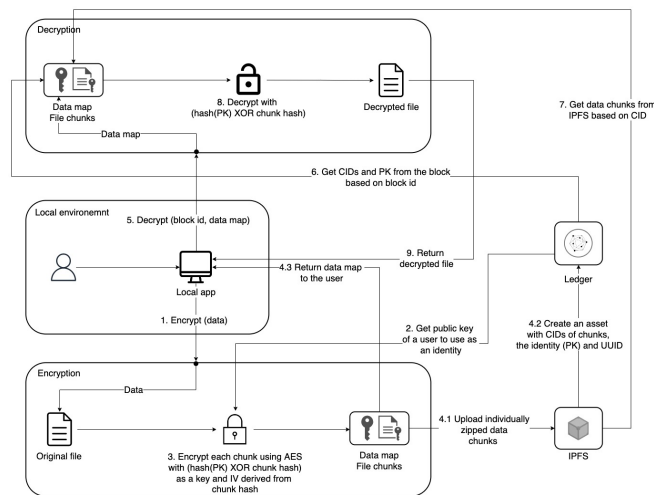

Figure 3: Workflow

There are two major parts of the system - encryption and decryption. Figure 3 demonstrates the workflow of both of them.

Encryption

The first part, encryption, that deals with encrypting a file and uploading data to the Hyperledger Fabric starts when the following command is called:

```
$ ibse add [file] [key_output_path]
```

where *ibse* is the name of the local app, *file* is the absolute path to the file that needs to be encrypted, and *key_output_path* is the absolute path to the location where the key will be stored.

The original file is uploaded to the directory that was mapped during the VM initialization. From there, it can be

---

[1]https://github.com/ilyagrishkov/ib-self-encryption-smart-contract

[2]https://github.com/ilyagrishkov/ib-self-encryption

read by the WASM code. The encryption function is then called, and the output is written to a new directory inside the mapped one. The output consists of multiple encrypted file chunks and a data map. The data map is moved to the location specified by the user and can later be shared via a secure channel. Each encrypted file chunk is being put into a zip archive to preserve their names when uploading to the IPFS and sent to the IPFS. The unique identifier corresponding to each chunk (Content Identifiers or CIDs which are the hash values of the files) is returned. A smart contract function is then called that creates a new asset with all CIDs.

Decryption

The second part of the system, decryption, is invoked using the following command:

```
$ ibse get [block] [key] [destination]
```

where *block* is the UUID of an asset in HLF blockchain that contains CIDs of encrypted chunks, *key* is the absolute path to the data map, and *destination* is the absolute path to the location where the decrypted file should be written.

The UUID allows identifying an asset containing CIDs of encrypted file chunks. Each chunk is downloaded from the IPFS, unarchived, and written to the directory accessible from the VM. The data map is then copied to the same directory. After collecting all the necessary files for decryption, the decryption function is called, and the restored file is written to a user-specified destination.

5 Results

5.1 Performance analysis

Benchmarking the system was done on the iMac 2019, 3,6 GHz 8-Core Intel Core i9 with 32 GB of memory running on MacOS 12.3.1.

Benchmarking of the implemented id-based self-encryption scheme was done. As the encryption itself is not implemented in the same language as the rest of the project (the encryption is implemented in Rust, and the rest of the project is in Go), the execution time can differ when Rust functions are called from Go compared to pure Rust execution time.

Files of sizes 100-, 250-, 500-, 750 kilobytes, 1 megabyte, 10-, 25-, 50-, 75-, and 100 megabytes were created to benchmark the pure Rust implementation as well as the WASM + Go implementations. Moreover, for this benchmark, the Rust code and the WASM library were optimized using the maximum level of optimization provided by the Rust compiler.

The initial benchmark was performed on the encryption function only and was measuring the execution time of the pure Rust implementation. Figure 4 shows the results of the benchmarking.

The chart shows near-linear dependence between the size of the file and the time it takes to encrypt it. This dependence can be explained by the fact that the most demanding computational is the AES 128 encryption process and hashing, and with the increase of the file size, the number of chunks it is split to increases. Each chunk of the original file needs to be
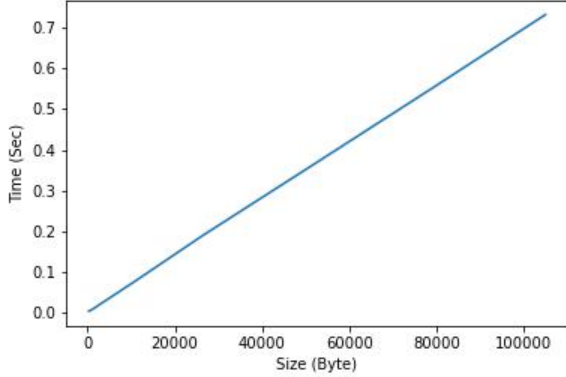
Figure 4: Dependence of the execution time of id-based self-encryption algorithm in pure Rust from the file size

Figure 5: Run time distribution for 50MB file encryption using pure Rust and WebAssembly + Go implementations

individually encrypted. Hence the computation time grows linearly with the size of the file.

As the encryption function execution time grows linearly due to the computational demand of the AES 128 and the hashing algorithms, the decryption process will be identical because it uses the same algorithms for decryption.

In order to achieve more objective benchmark results, a file of each size has been encrypted 100 times, and the average calculated. In order to visualize execution time, a chart in Python using MatPlotLib[1] was created. The chart contains a 25-bin histogram, each representing density of a particular measurement. Following the central limit theorem, the distribution of the execution time measurements was assumed to be normal, so the mean and the standard deviation were calculated, and the distribution was plotted over the histogram. Figure 5 shows an example of combined charts for pure Rust and WASM + Go execution times when encrypting 50 megabytes file. The blue histogram on the left-hand side shows the results of the 100 measurements of the execution time of the Rust implementation; on the right-hand side - of the WASM + Go implementation.

The results of execution time measurements for various file sizes are summarized in Table 1. The execution time shown in the table is the average number of seconds it takes a corresponding implementation to encrypt a file of a corresponding size. In addition to the average execution time, the overhead of the WASM + Go implementation is calculated for every pair of measurements.

It is visible from the table that the overhead has a clear downwards trend (except for the spike when encrypting a 250KB file). When the execution time of a WASM + Go encryption implementation is less than 0.01 seconds, the overhead falls in the range between 70% and 85%. When the execution time is longer than 0.1 seconds, the overhead goes down to 50% - 55% and stays in that range when the file size increases. Figure 6 demonstrates the overhead of WASM + Go encryption of files of different sizes.

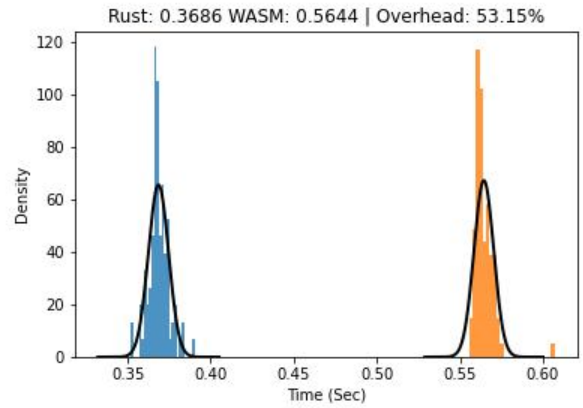Such a decrease in the overhead when the execution time

---

[1]https://matplotlib.org/

File size	Average execution time (Sec)		Overhead
(Byte)	Rust	WASM + Go	(%)
100KB	0.0024	0.0042	75.89
250KB	0.0038	0.007	84.51
500KB	0.005	0.0088	75.48
750KB	0.0069	0.0117	71.4
1MB	0.0081	0.0139	71.29
10MB	0.0747	0.117	56.55
25MB	0.1885	0.2851	51.22
50MB	0.3686	0.5644	53.15
75MB	0.5492	0.8447	53.81
100MB	0.7317	1.1201	53.08

Table 1: Average execution time and overhead when encrypting files of different sizes using id-based self-encryption

becomes longer is explained by the presence of the Wasmer library invocation overhead, which occurs every time a call is made to the Wasmer VM. When execution time is less than 0.01 seconds, the invocation overhead is significant compared to the execution time. At the same time, when the execution time becomes longer, the overhead from invocation becomes insignificant, and measurements start to approximate the actual WASM VM overhead, which is around 50% - 55%.

5.2 Security analysis

The designed app has multiple surfaces of attack. The IPFS nodes where the encrypted file chunks are stored can be attacked. Also, an adversary can gain unauthorized access to the ledger with references to files on the IPFS. Both of those possibilities are analyzed below.

Firstly, the security of IPFS nodes (assuming the encrypted file chunks were stored individually on multiple nodes) can be compromised, in which case encrypted files will be leaked to the malicious user. As encrypted file chunks have been stored on different nodes, the probability that all of them are compromised is negligible and should not be considered. Additionally, individual files do not link to each other, so matching multiple encrypted chunks needed for successful decryption is not computationally feasible.
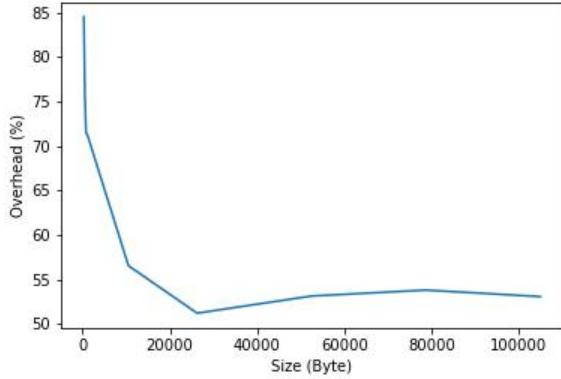
Figure 6: WebAssembly + Go implementation overhead measurements over for files of different sizes compared to pure Rust implementation

Secondly, the security of smart contracts can be compromised. If a malicious third party gains access to the ledger, then all assets' data (ID, Owner, CID) becomes available to the attacker. A list of CID-s will allow the attacker to download encrypted file chunks from the IPFS. Data integrity, in this case, relies on the security of the self-encryption scheme. In this case, the missing part for the decryption is the data map, which was stored locally by the user who encrypted the file. Without the original keys, such an attack on the self-encryption scheme is computationally not feasible [5].

6 Responsible Research

Multiple research papers, journals, and online resources were thoroughly analyzed. Every use of other authors' materials and resources is referenced and cited accordingly. All data used for benchmarking the implementation was created by the author of the paper. Benchmarking itself was also implemented by the author. The source code used throughout the research is open-source and cited accordingly. The implementations of the smart contract, encryption library, and the local application have been cited accordingly and can be found in GitHub repositories.

7 Discussion

The results show high-security guarantees of the id-based self-encryption scheme when used for encrypting data stored on the Hyperledger Fabric blockchain, which allows using the implemented prototype as a secure medium for saving and retrieving information from the ledger.

Additionally, the study demonstrates a relatively low overhead and high performance of WebAssembly library integration with the Go code base compared to the pure Rust implementation. The relatively small overhead of WASM run time creates possibilities for developers to use WASM integration with Go and other languages that support Wasmer library as a cross-platform solution that achieves high degrees of performance while also being deterministic.

In future works, full Go implementation of self-encryption should be compared with the design proposed in this paper.

It was also beyond the scope of this study to create a standardized benchmarking for WASM and Rust libraries. It can be done using multiple sample programs that test specific properties of the programming language (e.g., the efficacy of memory allocation and deallocation) or very computationally intensive programs [15]. The objective could be running a containerized version of both libraries against a set of such programs and analyzing the run time.

Moreover, the study can be expanded by analyzing and comparing the CPU and memory load of WASM and Rust libraries. Such benchmark could also be done using sample programs mentioned in the previous paragraph.

8 Conclusion

In this study, a new approach to the storage of files on the Hyperledger Fabric blockchain was presented. The demonstrated approach allows for secure data storage in a decentralized way, with the ability to preserve the original file ownership and information about the person who encrypted it. The ownership preservation allows to uniquely identify a user who has encrypted data even outside of the blockchain context. At the same time, self-encryption guarantees that file content will not be known to malicious third parties. This approach can be used for storing information on the Hyperledger Fabric ledger in applications, where data security and integrity are critical. Moreover, integrating a user's identity into the encrypted data makes it possible for this approach to be used in systems where parties can not be fully trusted.

The prototype uses Rust implementation of id-based self-encryption that is compiled to WebAssembly and invoked from Go code with 55% overhead.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Dec 2008, accessed: 2015-07-01. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[2] V. Buterin, "Ethereum: A next-generation smart contract and decentralized application platform," 2014, accessed: 2016-08-22. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper

[3] C. Cachin, "Architecture of the hyperledger blockchain fabric," 2016, accessed: 2016-08-10. [Online]. Available: https://www.zurich.ibm.com/dccl/papers/cachin_dccl.pdf

[4] Y. Huang, Y. Bian, R. Li, J. L. Zhao, and P. Shi, "Smart contract security: A software lifecycle perspective," *IEEE Access*, vol. 7, pp. 150 184–150 202, 2019.

[5] D. Irvine, "Self encrypting data," 2010.

[6] C. Park, "Using self-encryption to safeguard data security in fabric's smart contract," 2022.

[7] K. Yamashita, Y. Nomura, E. Zhou, B. Pi, and S. Jun, "Potential risks of hyperledger fabric smart contracts," in *2019 IEEE International Workshop on Blockchain*

Oriented Software Engineering (IWBOSE). IEEE, 2019, pp. 1–10.

[8] A. Dabholkar and V. Saraswat, "Ripping the fabric: Attacks and mitigations on hyperledger fabric," in *International Conference on Applications and Techniques in Information Security*. Springer, 2019, pp. 300–311.

[9] C. Stamatellis, P. Papadopoulos, N. Pitropakis, S. Katsikas, and W. J. Buchanan, "A privacy-preserving healthcare framework using hyperledger fabric," *Sensors*, vol. 20, no. 22, p. 6587, 2020.

[10] Y. Chen and W.-S. Ku, "Self-encryption scheme for data security in mobile devices," in *2009 6th IEEE Consumer Communications and Networking Conference*. IEEE, 2009, pp. 1–5.

[11] M. R. D. Rahardjo and G. F. Shidik, "Design and implementation of self encryption method on file security," in *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)*. IEEE, 2017, pp. 181–186.

[12] Maidsafe, "self_encryption," https://github.com/maidsafe/self_encryption, 2022.

[13] Hyperledger, "Using the fabric test network," 2020, accessed: 2022-03-19. [Online]. Available: https://hyperledger-fabric.readthedocs.io/en/release-2.2/test_network.html

[14] F. Valsorda, "Rustgo: Calling rust from go with near-zero overhead," Feb 2019. [Online]. Available: https://words.filippo.io/rustgo/

[15] I. Gouy, "Toy benchmark programs." [Online]. Available: https://benchmarksgame-team.pages.debian.net/benchmarksgame/why-measure-toy-benchmark-programs.html