

Encrypted Container File: Design and Implementation of a Hybrid-Encrypted Multi-Recipient File Structure

Tobias J. Bauer and Andreas Aßmuth

Ostbayerische Technische Hochschule Amberg-Weiden
Faculty of Electrical Engineering, Media and Computer Science
Kaiser-Wilhelm-Ring 23, 92224 Amberg, Germany
Email: {t.bauer|a.assmuth}@oth-aw.de

Abstract—Modern software engineering trends towards Cloud-native software development by international teams of developers. Cloud-based version management services, such as GitHub, are used for the source code and other artifacts created during the development process. However, using such a service usually means that every developer has access to all data stored on the platform. Particularly, if the developers belong to different companies or organizations, it would be desirable for sensitive files to be encrypted in such a way that these can only be decrypted again by a group of previously defined people. In this paper, we examine currently available tools that address this problem, but which have certain shortcomings. We then present our own solution, Encrypted Container Files (ECF), for this problem, eliminating the deficiencies found in the other tools.

Keywords—Cloud-based software development; hybrid encryption; agile software engineering.

I. INTRODUCTION

Software development undergoes a permanent change and, occasionally, long-lasting trends emerge, which influence the choices made in terms of software architectures, technologies, programming languages and frameworks used. Current trends involve the development of Cloud-native distributed software components which are deployed automatically via Continuous Delivery and Continuous Deployment [1].

This implies that these components, often running in separate containers, must communicate with each other. Furthermore, there is an interest in securing such communication links because very often confidential data is transmitted. This in turn places demands on the software development process: in order to secure (digital) communications these must be encrypted. This is also true for storing confidential data. In common cases, e.g., running a web server or storing confidential data in a database, means of authentication must be kept secret. Such means of authentication include, but are not limited to, passwords, private certificate keys, and symmetric encryption keys.

Modern software development takes place in teams whose members are in constant exchange with each other. Often, version control systems, e.g., *git* [2] are used to manage the source code and other artifacts. Also with regard to the practice of Continuous Integration (see [3]), which is a preliminary step to the aforementioned Continuous Delivery and Continuous Deployment, it is necessary to check-in *all* artifacts into the version control system. This would be grossly negligent for

confidential data provided that no protective measures against unauthorized access are taken.

In this paper, we address the issue of access to an encrypted file structure in the cloud by different people in a software development team. With the Encrypted Container File (ECF), we present our own solution for a cloud-based, encrypted data storage for software development teams, in which the functionality of currently available tools is extended and their shortcomings are eliminated.

This paper is structured as follows: in Section II, there is a brief introduction to two existing solutions before the requirements are presented in Section III. In Section IV, we present an example of use and describe the structure and operations of the ECF. Following that, Section V describes implementation details. Finally, Section VI concludes the paper and gives an outlook on future work.

II. RELATED WORK

There are different solutions to address the issue we described in Section I. In this section, we give an overview of two of these tools, *jak* and *git-crypt*, and discuss their features and shortcomings.

The tool *jak* [4] is written in Python and allows symmetric encryption of files using Advanced Encryption Standard (AES). Using the tool, one can generate keys and store them in a keyfile, which is not encrypted. To enable automatic encryption and decryption with a single command *jak* uses a special text file that contains a list of file names. This special text file can be added to the repository [4].

The practical use is limited because of sole symmetric encryption as the key distribution problem remains unsolved. Especially with growing team sizes distributing confidential data results in disproportionate effort.

Another issue with *jak* is that the confidential files' content stays unencrypted on the developers' computers. This is because *jak* decrypts these files during checkout and re-encrypts them before committing. This implies that only externals with reading access to the repository at maximum and no access to any of the developers' computers are unable to access the confidential data. A common application scenario are projects that are developed on a public repository platform.

The tool *git-crypt* [5] allows symmetric encryption of files within a *git* repository using AES, too. It shares the same limitations as *jak* in terms of access restrictions to externals.

However, *git-crypt* offers a solution to the key distribution problem by using GNU Privacy Guard (GPG) [6]. Public GPG-keys, the recipients' keys, can be added to the repository. When encrypting the confidential files within the repository *git-crypt* generates an asymmetrically encrypted keyfile for each recipient. Every recipient therefore gets access to the symmetric key and because of that is capable of decrypting the confidential files in that repository.

The tool *git-crypt* is implemented in a way that all confidential files are encrypted with the same symmetric key and this very key must therefore be shared with all recipients added to the repository. This results in coarse grained access control as there is no way to restrict access to some confidential files to a subset of the recipients. Consider the case that, e.g., secret information about the production environment should only be accessible to the production team.

Furthermore, *git-crypt* does not secure the confidential files' content on the developers' computers. This is analogous to *jak* because both tools decrypt the confidential files during checkout. The integration of *git-crypt* into the mechanisms of *git* is optional but recommended [5].

Another shortcoming of *git-crypt* is the lacking feature to remove recipients. Ayer justifies this by stating that by using a version control system a removed recipient can still access old versions of the repository and, therefore, the confidential data stored within [5]. This argument is correct as far as it goes – nevertheless, it seems sensible to implement such a mechanism into the to-be-designed ECF format since the confidential data should be updated regularly regardless. For example, certificates and passwords expire and symmetric keys should be changed regularly with regards to staff turnover.

III. REQUIREMENTS ENGINEERING

From the features and shortcomings of the *jak* and *git-crypt* tools presented in Section II, some requirements for the ECF format can be derived:

- 1) Mandatory encryption of confidential data,
- 2) possibility to modify confidential data (content is writable),
- 3) key distribution is no prerequisite,
- 4) decryption not during checkout but on demand,
- 5) support for multiple recipients,
- 6) addition and removal of recipients,
- 7) minimal information gain for external parties, and
- 8) customizable set of recipients per file.

Based on these requirements, we have decided to use the following design goals for our solution:

- Use of hybrid encryption (Items 1, 3 and 5),
- inclusion of recipient information to allow re-encryption on changes (Items 2, 5, 6 and 8),
- obfuscation of recipient information for respective external parties (Items 7 and 8), and
- delivery of the associated software as a library for embedding into existing applications (Item 4).

IV. STRUCTURE AND OPERATIONS OF THE ENCRYPTED CONTAINER FILE

This section gives an overview over the use of the ECF format in Subsection IV-A. The following subsections describe the structure of the ECF format in detail. Figure 1 shows an overview of the components of an ECF, how they are connected and related to each other. Subsection IV-B describes the general structure, components, and storage formats of an ECF. The publicly accessible fields are described in Subsection IV-C and the private fields in Subsection IV-D. The following Subsections IV-E and IV-F describe the decryption and encryption process, respectively. Finally, Subsection IV-G concludes this section with further operations that can be performed on an ECF.

A. Usage in Practice

In this subsection, we walk through the following scenario: Alice wants to encrypt a file using the ECF format and operations in such way that her friend Bob will be able to read the content, while Charlie should not be able to.

First, Alice needs access to Bob's public information, which comprises among others his public key. Bob must have created his public information beforehand. Next, Alice creates an ECF using, e.g., the CLI tool described in this paper and provided via GitHub, and adds the confidential data. After that, she can add Bob as a recipient using his public information. To retain access to the content, Alice should add herself as a recipient to the ECF. Alice can now save the ECF within a public repository and only Bob and herself are able to decrypt the file's content. Charlie, on the other hand, cannot retrieve the encryption key as he is not a recipient of that ECF and has therefore no access to the confidential data stored inside.

B. General Structure and Data Type Storage Format

Each ECF consists of three parts: A public part and two non-public/private parts. In Figure 1, the whole ECF is framed yellow, whereas the public part is colored purple. Both private parts are treated as a single datum by the symmetric encryption and are colored in blue. The following list describes the data types used in the following subsections and their storage format within an ECF:

- Unsigned Integer: 4 Bytes, Little Endian
- Byte Array [x]: x Bytes, sequential
- String: 4 Bytes, little endian (Length), then UTF-8 bytes without byte order mark (BOM)

The ECF format is designed to be flexible with regards to the used cipher suite. In order to allow future extensions, it allows more algorithms and cipher suites. For this paper and also for our Proof of Concept (PoC) implementation, a selection for the cipher suite was made, which is the basis for the rest of this paper:

- Key Agreement/Exchange: X25519 [7]
- Symmetric Encryption: AES-256-GCM [8]
- Signature: Ed25519 [9]
- Hash Function: SHA-512 [10]

C. Public Fields

Each ECF must provide enough information for all authorized recipients to decrypt the file. Information for encrypting, however, is not required to be public because only recipients should be able to modify the confidential data within the ECF. Hence, the public part of an ECF contains just the information required for decryption. It comprises a general part and then m identically constructed recipient-specific parts.

The general part contains the following data (in this order):

- Container Version (Unsigned Integer)
ECF format version; intended for future extensions
- Cipher Suite (Unsigned Integer)
Information about used algorithms
- Public Header Length (Unsigned Integer)
Length of the public part in Bytes
- Private Length (Unsigned Integer)
Length of the private part in Bytes
- Recipient Count (Unsigned Integer)
Number of recipients in the public part (m)
- Salt (Byte Array [16])
Salt value (usage described below)
- Symmetric Nonce (Byte Array [12])
Symmetric nonce value (usage described below)

The first two fields, Container Version and Cipher Suite, are used to make the ECF format flexible and future-proof. However, we discuss only the cipher suite selected in Subsection IV-B.

The recipient-specific decryption information is yet to be defined. In total, m such blocks – one for each recipient – are stored after the general part. To obfuscate the number of recipients towards externals, $m \geq n$ can be chosen freely with n the true number of recipients. Random blocks, which belong to no recipient, may be inserted, which is not evident to externals. Each recipient-specific block consists of two fields: an Identification Tag (Byte Array [16]), which is used to assign a block to a recipient, and Key Agreement Information that contains recipient-specific decryption information.

The field Identification Tag is colored orange in Figure 1. It is the hash value truncated after 16 Bytes from the concatenation of the bit strings of the public key of the respective recipient and the Salt value introduced above. Shortening the hash value saves storage space and allows with $(2^8)^{16} = 2^{128}$ possible values for practically unlimited unique recipients. An authorized recipient can calculate their Identification Tag based on the knowledge of their own public key and the public Salt.

The second field, Key Agreement Information, contains recipient-specific information for the decryption process and is highly dependent on the used cipher suite. For the selected cipher suite, an Ephemeral X25519 Public Key (Byte Array [32]) and an AES Pre Key (Byte Array [32]) is stored. The first is used in the key agreement phase to obtain a second AES pre key, the latter is the first AES pre key. Subsection IV-E describes the combination

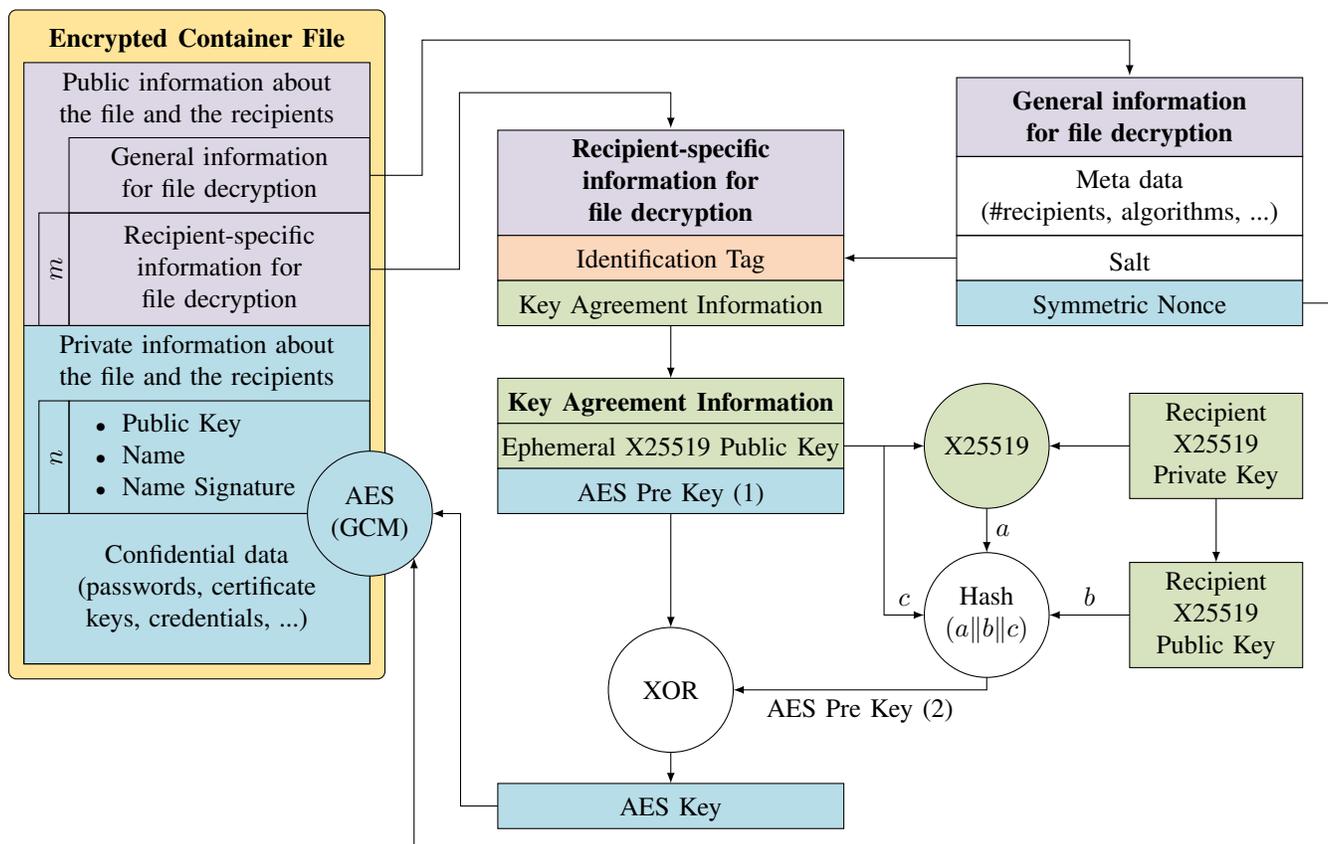


Figure 1. Diagram of the most important components of an Encrypted Container File and visualization of the interrelationships.

of the available recipient-specific information to obtain the symmetric key in more detail. The process is depicted in Figure 1 as well.

D. Private Fields

The private part of an ECF consists of two segments: First, there is information about the ECF and its recipients, and second, there is the encrypted confidential data itself. The private part is completely encrypted symmetrically and, therefore, not accessible for external parties. The following fields are stored in the private part:

- Content Type (Unsigned Integer)
Describes the type of the confidential data
- Public Header Hash (Byte Array [64])
Hash value of the public part
- Recipient Count (Unsigned Integer)
Number of true recipients (n)
- Recipient Information (Array [n])
Information about the recipients (n blocks)
- Content Length (Unsigned Integer)
Length of the confidential data in Bytes (len)
- Content (Byte Array [len])
Confidential data
- Private Hash (Byte Array [64])
Hash value of the private part so far

The first field, Content Type, is intended for future use and should characterize the type of confidential data stored in Content. It seems reasonable that future applications using the ECF library will define and handle their own content types.

The field Public Header Hash contains the hash value over the whole public part of an ECF. The value for the public field Public Header Length (c.f. Subsection IV-C) is unknown at the time of encryption because the length of the symmetric encryption algorithm's output is not necessarily known in advance. Therefore, this field is set to the constant value of 0xECFFCODE (Encrypted Container File Format Code) during the calculation of the hash value. The Public Header Hash is used to detect unauthorized or unintended modifications of the public part, e.g., non-destructive changes of recipient-specific information of other recipients.

The fields Recipient Count and Content Length specify the number of true recipients and the length of the confidential data, respectively. The field Recipient Information consists of n blocks of variable length, which in turn consist of three fields: the Public Key (Byte Array [32]) of the recipient, a Name (String) which contains a self-chosen name of the recipient (variable length), and the Name Signature (Byte Array [64]) over the self-chosen name.

Every block of Recipient Information contains information about a recipient, so that re-encrypting the ECF is possible, e.g., after modifying confidential data. These information blocks about the recipients are stored within the private part of the ECF in order to hide them from externals. The block field

Public Key contains the recipient's public key, which is a public Ed25519 key as specified in Subsection IV-B. One can convert an Ed25519 public key into an X25519 public key as described in [11][12]. The next block field Name holds a text of variable length that describes the recipient. It may contain the name of the related person or their email address. This field is for human legibility and information purposes only, e.g., when displaying the recipients or when removing existing recipients. The last block field, Name Signature, contains a signature over the content of Name. The signature is used first and foremost to ensure, that the person owing the associated private key has chosen the name, and that no changes have been made to the name by other recipients afterwards.

The field Content encloses the confidential data and has a theoretical limit of $2^{32} - 1$ Bytes ≈ 4 GiB. In practice, this limit should never be reached because an ECF is designed primarily to be used with passwords, certificate keys, credentials and similar confidential data.

The last field, Private Hash, takes the hash value over the private part up to this point. This field is inside the private part of an ECF and, therefore, the hash value is calculated before encryption. A more detailed description of the encryption process can be found in Subsection IV-F.

E. Decryption Process

This subsection describes the processes of calculating the AES key according to Figure 1. To decrypt an ECF a recipient needs both, their private X25519 key and their public Ed25519 key. Both can be calculated from the recipient's private Ed25519 key [11][12].

Nomenclature. The following notation is used: Alice is the recipient and sk_A^{Ed} denotes her private Ed25519 key, pk_A^X denotes her public X25519 key, analogously. The used cryptographic hash function is denoted by H , $a||b$ denotes the concatenation of two bit strings a and b , and $a \oplus b$ denotes the bitwise exclusive OR (XOR) operation on two bit strings a and b of the same length. $a[0, \dots, n]$ denotes the truncation of the bit string a to the first n Bytes. The ephemeral public X25519 key contained in the recipient-specific decryption information is denoted by pk_c^X . The function $X25519(a, B)$ describes the multiplication of scalar a with point B on the elliptic curve *Curve25519* [7].

Alice performs the following steps to obtain the AES key:

- (1) Compute $identification_tag = H(pk_A^{Ed} || Salt)[0, \dots, 16]$.
- (2) Load the decryption information (pk_c^X, k_{pre1}^{AES}) with matching $identification_tag$.
- (3) Execute the key agreement algorithm with Alice's private X25519 key and the public ephemeral X25519 key:
 $k_{shared}^X = X25519(sk_A^X, pk_c^X)$.
- (4) Compute $k_{pre2}^{AES} = H(k_{shared}^X || pk_A^X || pk_c^X)[0, \dots, 32]$.
Shortening the hash value to 32 Bytes is necessary because of the used symmetric encryption algorithm AES-256
- (5) Compute $k_{pre1}^{AES} = k_{pre1}^{AES} \oplus k_{pre2}^{AES}$.

In Step 4 the hash function gets evaluated on the concatenation of the shared key and both public keys to obtain

the second AES pre key. The reason for this is a recommendation in [13] to not use the shared key k_{shared}^X directly but to transform it with a hash function first. The question arises to why a simple hash function is used and not a Key Derivation Function (KDF). Primarily, the reason is to speed up the encryption process, because using a KDF is resource-intensive and it must be computed n times (separately for each of the n recipients). This would result in a far slower encryption process for large n . Furthermore, the input data in Step 4 is substantially longer than the symmetric pre key to be computed, which makes key stretching not required and, therefore, the use of a cryptographic hash function seems sufficient.

Finally, the private part of an ECF can be decrypted by using the computed AES key k^{AES} and the public AES nonce. In this paper, the Galois/Counter Mode (GCM) [14] was chosen for the symmetric encryption algorithm AES. Therefore, one is not required to check the authenticity of the decrypted data separately. Furthermore, instead of the field `Public Header Hash` the public part of the ECF could have been authenticated with AES-GCM. However, when supporting different modes of operation this field would have been required anyway. Hence, the field `Public Header Hash` was not removed and no additional data (*Associated Data*) was added to the AES-GCM encryption algorithm.

F. Encryption Process

The encryption process consists of an initial key and nonce generation step and an m -wise computation of the public X25519 ephemeral keys and AES pre keys. For each of the $n \leq m$ true recipients exactly one public recipient-specific decryption information block must be generated. The remaining $m - n$ blocks serve as obfuscation and may be generated using a special process as proposed in Appendix A.

Nomenclature. The same nomenclature applies as in Subsection IV-E. It gets extended by the following functions: $\text{Gen}^{\text{AES}}(256)$ and Gen^X denote functions to generate AES-256 keys and X25519 key pairs, respectively. $\text{RandomBytes}(x)$ denotes a function to generate a random bit string of length x Bytes.

For each recipient *Bob*, their public Ed25519 key pk_B^{Ed} is known by every recipient of that ECF because of the (private) block field `Public Key` (see Subsection IV-D). Based on pk_B^{Ed} one can calculate Bob's public X25519 key pk_B^X [11][12].

First, a symmetric AES key $k^{\text{AES}} \leftarrow \text{Gen}^{\text{AES}}(256)$, an AES nonce $\text{nonce}^{\text{AES}} \leftarrow \text{RandomBytes}(12)$ and a bit string $\text{Salt} \leftarrow \text{RandomBytes}(16)$ must be generated at random (randomness indicated by the left arrow " \leftarrow ").

Then, the following steps are performed n times to generate the key agreement information for each recipient Bob:

- (1) Compute $\text{identification_tag} = \text{H}(\text{pk}_B^{\text{Ed}} \parallel \text{Salt})[0, \dots, 16]$.
- (2) Generate an ephemeral X25519 key pair:
 $(\text{sk}_c^X, \text{pk}_c^X) \leftarrow \text{Gen}^X$.
- (3) Execute the key agreement algorithm with the private ephemeral X25519 key and Bob's public X25519 key:
 $k_{\text{shared}}^X = \text{X25519}(\text{sk}_c^X, \text{pk}_B^X)$.

- (4) Compute $k_{\text{pre2}}^{\text{AES}} = \text{H}(k_{\text{shared}}^X \parallel \text{pk}_B^X \parallel \text{pk}_c^X)[0, \dots, 32]$.

Shortening the hash value to 32 Bytes is necessary because of the used symmetric encryption algorithm AES-256.

- (5) Compute $k_{\text{pre1}}^{\text{AES}} = k^{\text{AES}} \oplus k_{\text{pre2}}^{\text{AES}}$.

Steps 2 and 3 correspond to a "half" Diffie-Hellman key exchange [15] that gets completed during decryption (see Subsection IV-E) in Step 3.

For each recipient Bob the recipient-specific information can be written into the public part of the ECF. This information per recipient consists of `identification_tag`, public ephemeral X25519 key pk_c^X and AES pre key $k_{\text{pre1}}^{\text{AES}}$.

The values `Salt` and $\text{nonce}^{\text{AES}}$ are valid for all recipients and are written into their respective fields (see Subsection IV-C).

G. Further ECF Operations

This subsection introduces more ECF operations which are based on the elementary operations Decryption (Subsection IV-E) and Encryption (Subsection IV-F). The same nomenclature is used as in the specified subsections. It gets extended by the function $\text{Dec}^{\text{ECF}}(\text{sk}_A^{\text{Ed}}, \mathcal{E})$ which denotes the decryption of an ECF \mathcal{E} with Alice's private Ed25519 key sk_A^{Ed} . This function returns a tuple (R, p) after successful decryption, with R being the set of all n recipients $R = \{r_1, r_2, \dots, r_n\}$ and p being the bit string of the decrypted confidential data. Analogous to this, the function $\text{Enc}^{\text{ECF}}(R, p)$ encrypts the confidential data p for the recipients R and returns an ECF \mathcal{E} .

1) Modification of Confidential Data:

Let $p' = \text{modify}(p)$ be the new bit string created by modification of the original confidential data p . The replacement of the confidential data within an ECF \mathcal{E} is done by these steps:

- (1) $(R, p) = \text{Dec}^{\text{ECF}}(\text{sk}_A^{\text{Ed}}, \mathcal{E})$
- (2) $p' = \text{modify}(p)$
- (3) $\mathcal{E}' \leftarrow \text{Enc}^{\text{ECF}}(R, p')$

2) Addition of a New Recipient:

Recipient Alice wants to add a new recipient Bob to an existing ECF. Bob's public Ed25519 key is denoted by pk_B^{Ed} , the bit string of his name by name_B . The signature over Bob's name is denoted by $s = \text{signature}^{\text{Ed}}(\text{sk}_B^{\text{Ed}}, \text{name}_B)$. Alice performs the following steps to add Bob to the recipient list:

- (1) Alice verifies the Signature s :
 $\text{verify}^{\text{Ed}}(s, \text{pk}_B^{\text{Ed}}) \stackrel{?}{=} \text{Valid}$.
- (2) If the signature is invalid, abort the operation, if the signature is valid, proceed.
- (3) $r_B = (\text{pk}_B^{\text{Ed}}, \text{name}_B, s)$
- (4) $(R, p) = \text{Dec}^{\text{ECF}}(\text{sk}_A^{\text{Ed}}, \mathcal{E})$
- (5) Alice checks whether Bob is already in the recipient list:
 $R \cap \{r_B\} \stackrel{?}{=} \emptyset$ (Compare Ed25519 public keys).
- (6) If Bob is already a recipient, abort the operation, if Bob is not a recipient, proceed.
- (7) Optionally, Alice can check if name_B already exists in one r_i and abort the operation if necessary.
- (8) $R' = R \cup \{r_B\} = \{r_1, r_2, \dots, r_n, r_B\}$

(9) $\mathcal{E}' \leftarrow \text{Enc}^{\text{ECF}}(R', p)$

3) Removal of a Recipient:

Recipient Alice wants to remove a recipient Bob from an existing ECF. Bob's public Ed25519 key pk_B^{Ed} and/or the bit string of his name name_B must be known. If only his name is known, it must be unique within the ECF \mathcal{E} . Alice performs the following steps to remove Bob from the recipient list:

- (1) $(R, p) = \text{Dec}^{\text{ECF}}(\text{sk}_A^{\text{Ed}}, \mathcal{E})$
- (2) Alice searches for r_B in R based on his public key pk_B^{Ed} or his name name_B .
- (3) If r_B does not exist (Bob is not a recipient of \mathcal{E}), abort the operation, if r_B exists proceed.
- (4) $R' = R \setminus \{r_B\}$
- (5) $\mathcal{E}' \leftarrow \text{Enc}^{\text{ECF}}(R', p)$

When removing recipients, one does not require any private keys during the encryption process. This implies that recipients of an ECF can remove themselves. It is therefore the task of the implementation to warn the user or abort the operation if the user attempts to do this. Additionally, the implementation should also realize additional security functions if, for example, only the creator of the confidential data stored in an ECF is allowed to add or remove recipients. Finally, it must be noted that the restriction explained in Section II is still true: Former recipients are still able to access old versions of an ECF when using a version control system.

V. IMPLEMENTATION DETAILS

A PoC was implemented using *C#* and the *.NET 6.0* runtime. All cryptographic primitives were provided by the portable library *Sodium* [16], which is a fork of the *NaCl* [17] library. To use *Sodium* with *.NET* a wrapper is needed. For this PoC the wrapper library *NSec* [18] was used.

A. Implementation of ECF Functionality

The implementation in *C#* was subdivided into two projects: *ECF.Core* and *ECF.CLI*. The *ECF.Core* project contains all functionality of the ECF and helps with managing private keys (see Subsection V-B). *ECF.Core* is a library and can be included into other projects (according to requirements in Section III). It is used by the *ECF.CLI* project which provides a command line interface to the ECF functionality.

The `Create(CipherSuite, ContentType)` function in class `EncryptedContainer` implements the ECF creation process using the given cipher suite and content type. For this PoC the aforementioned cipher suite (see Subsection IV-B) is implemented as well as a single content type: *BLOB*. Because of the GCM mode of operation, the execution platform must support the instruction set extension *AES-NI*. As a rule, this can only lead to problems when using very old processors or virtual machines.

An object of type `EncryptedContainer` can be encrypted using the function `Write(Stream)`. The output is written into the parameter `Stream`. Analogously, one can obtain an unencrypted object of this type using the class function

`Load(Stream, ECFKey)`. It is necessary to provide a private Ed25519 key of a recipient to that function. Per default all name's signatures are verified. This can be disabled to achieve better runtime performance during decryption. The property `ContentStream` of an `EncryptedContainer` object provides read and write access to the confidential data.

To protect the private key and the confidential data the implementation uses protected memory spaces, if possible. The library *Sodium* provides suitable functions for this [19], which in turn are used by *NSec*. Furthermore, heap allocations are replaced by stack allocations wherever possible using the *C#* keyword `stackalloc` [20]. Alternatively, when using memory that cannot be protected via *Sodium* or stack allocation, the implementation pins these memory regions in memory to prevent the Garbage Collector from arbitrarily copying them. Furthermore, used memory regions are actively deleted before they are freed.

B. Private Key Management

Using ECFs requires private keys that should never be stored unencrypted. Therefore, the *ECF.Core* project uses AES-256 (GCM) to encrypt the private keys. The encryption key is derived from a user-provided password using Argon2id [21][22]. Argon2id aims to enforce costly calculations that cannot be parallelized or otherwise shortened (in time) by an attacker. The algorithm can be configured arbitrarily in order to keep the required computing time variable. For the PoC implementation the following settings were chosen:

- Degree of parallelism: 1 (Limit by *Sodium*)
- Memory requirements: 2 GiB
- Number of iterations: 5

This results in an approximate run time of 5s on an *Intel Core i5-6600K* with newer processors being presumably faster.

The private key is needed when decrypting an ECF and when creating an ECF. For the latter it is necessary to add oneself as a recipient to that ECF, which includes signing the name. Therefore, *ECF.CLI* always prompts the user's password to load the encrypted private key.

VI. CONCLUSION AND FUTURE WORK

In this paper we introduced Encrypted Container File, a hybrid-encrypted multi-recipient file structure aimed to store confidential data and share it with a customizable set of recipients. Full examples of basic and advanced operations recipients can perform on an ECF were presented in this paper. Although we were using a single cipher suite as described in Subsection IV-B, the file format supports multiple cipher suites which can be implemented analogously. The PoC implementation demonstrates this by implementing both SHA-512 and SHA-256 as cryptographic hash functions resulting in two different cipher suites.

The full code of the PoC implementation and unit tests for that code are available at:

<https://github.com/Hirmoder/ECF>

For the future, we plan to add additional cipher suites to ECF. Additional functionalities are also possible depending on the feedback we get from the community.

REFERENCES

- [1] Intellectsoft. “Top 8 software development trends in 2022.” Intellectsoft. (Dec. 3, 2021), [Online]. Available: <https://www.intellectsoft.net/blog/software-development-trends/> (visited on 06/05/2023).
- [2] Git. “Git.” (2023), [Online]. Available: <https://git-scm.com/> (visited on 06/05/2023).
- [3] M. Fowler. “Continuous integration.” (May 1, 2006), [Online]. Available: <https://www.martinfowler.com/articles/continuousIntegration.html> (visited on 06/05/2023).
- [4] Dispel LLC. “Jak – simple git encryption,” Dispel LLC. (2017), [Online]. Available: <https://jak.readthedocs.io/en/latest/> (visited on 06/05/2023).
- [5] A. Ayer. “Git-crypt – transparent file encryption in git.” (2023), [Online]. Available: <https://www.agwa.name/projects/git-crypt/> (visited on 06/05/2023).
- [6] W. Koch, N. Ellmenreich, M. Ashley, *et al.* “The gnu privacy guard,” The GnuPG Project. (May 31, 2023), [Online]. Available: <https://gnupg.org/> (visited on 06/05/2023).
- [7] D. J. Bernstein, “Curve25519: New diffie-hellman speed records,” in *Public Key Cryptography - PKC 2006*, Springer Berlin Heidelberg, 2006, pp. 207–228.
- [8] D. A. McGrew and J. Viega. “The galois/counter mode of operation (GCM),” National Institute of Standards and Technology. (May 31, 2005), [Online]. Available: <https://csrc.nist.rip/groups/ST/toolkit/BCM/documents/proposedmodes/gcm/gcm-revised-spec.pdf> (visited on 06/05/2023).
- [9] D. J. Bernstein, N. Duif, T. Lange, P. Schwabe, and B.-Y. Yang, “High-speed high-security signatures,” *Journal of Cryptographic Engineering*, vol. 2, no. 2, pp. 77–89, Aug. 2012.
- [10] Q. H. Dang, “Secure hash standard,” Tech. Rep., Jul. 2015. DOI: 10.6028/nist.fips.180-4. [Online]. Available: <https://doi.org/10.6028/nist.fips.180-4>.
- [11] “Squeamish Ossifrage”. “Curve25519 over Ed25519 for key exchange? Why?” Crypto Stack Exchange. (Mar. 19, 2019), [Online]. Available: <https://crypto.stackexchange.com/a/68129> (visited on 06/05/2023).
- [12] The Sodium Authors. “Ed25519 to curve25519 – libsodium.” (2023), [Online]. Available: <https://doc.libsodium.org/advanced/ed25519-curve25519> (visited on 06/05/2023).
- [13] The Sodium Authors. “Point*scalar multiplication – libsodium.” (2021), [Online]. Available: https://doc.libsodium.org/advanced/scalar_multiplication (visited on 06/05/2023).
- [14] M. J. Dworkin, “Recommendation for block cipher modes of operation: Galois/counter mode (gcm) and gmac,” Tech. Rep., 2007. DOI: 10.6028/nist.sp.800-38d. [Online]. Available: <https://doi.org/10.6028/nist.sp.800-38d>.
- [15] W. Diffie and M. E. Hellman, “New directions in cryptography,” *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.
- [16] The Sodium Authors. “Introduction – libsodium.” (2022), [Online]. Available: <https://doc.libsodium.org/> (visited on 06/05/2023).
- [17] D. J. Bernstein, T. Lange, and P. Schwabe. “Nacl: Networking and cryptography library.” (Mar. 15, 2016), [Online]. Available: <https://nacl.cr.yt.to> (visited on 06/05/2023).
- [18] K. Hartke. “Nsec – modern cryptography for .net core.” (2022), [Online]. Available: <https://nsec.rocks/> (visited on 06/05/2023).
- [19] The Sodium Authors. “Secure memory.” (2022), [Online]. Available: https://doc.libsodium.org/memory_management (visited on 06/05/2023).
- [20] Microsoft Docs. “stackalloc expression (c# reference),” Microsoft. (Apr. 12, 2023), [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/csharp/language-reference/operators/stackalloc> (visited on 06/05/2023).
- [21] A. Biryukov, D. Dinu, and D. Khovratovich. “Argon2: The memory-hard function for password hashing and other applications.” version 1.3. (Mar. 24, 2017), [Online]. Available: <https://raw.githubusercontent.com/P-H-C/phc-winner-argon2/master/argon2-specs.pdf> (visited on 06/05/2023).
- [22] The Sodium Authors. “The pwhash* api.” (2022), [Online]. Available: https://doc.libsodium.org/password_hashing/default_phf (visited on 06/05/2023).
- [23] “fgrieu”. “Using sha2 as random number generator?” Crypto Stack Exchange. (Jun. 20, 2020), [Online]. Available: <https://crypto.stackexchange.com/a/81459> (visited on 06/05/2023).

APPENDIX

A. Generating $m - n$ Obfuscation Blocks

In Subsection IV-F the generation process for the n public recipient-specific blocks was described. The remaining $m - n$ blocks serve as obfuscation blocks to hide the true number of recipients to externals. These obfuscation blocks should not be random bit strings because there is a possibility that the outputs of the used algorithms are subject to statistical effects. This would allow an external party to distinguish between real blocks and obfuscation blocks and therefore determine n .

To avoid this, we suggest that the $m - n$ obfuscation blocks are constructed using randomly generated Ed25519 and X25519 key pairs. The function Gen^{Ed} denotes the creation of an Ed25519 key pair and the function $\text{Convert}^{\text{X}}(\text{sk}^{\text{Ed}})$ converts an Ed25519 private key into an X25519 key pair. The following steps are performed for each obfuscation block:

- (1) Generate a random key pair:
 $(\text{sk}_r^{\text{Ed}}, \text{pk}_r^{\text{Ed}}) \leftarrow \text{Gen}^{\text{Ed}}, (\text{sk}_e^{\text{X}}, \text{pk}_e^{\text{X}}) = \text{Convert}^{\text{X}}(\text{sk}_r^{\text{Ed}}).$
- (2) Compute $\text{identification_tag} = \text{H}(\text{pk}_r^{\text{Ed}} \parallel \text{Salt})[0, \dots, 16]$.
- (3) Generate an ephemeral X25519 key pair:
 $(\text{sk}_e^{\text{X}}, \text{pk}_e^{\text{X}}) \leftarrow \text{Gen}^{\text{X}}.$
- (4) Generate a random AES-256 key:
 $k_r^{\text{AES}} \leftarrow \text{Gen}^{\text{AES}}(256)$ or $k_r^{\text{AES}} \leftarrow \text{RandomBytes}(32).$
- (5) Execute the key agreement algorithm with the private ephemeral and the random public X25519 keys:
 $k_{\text{shared}}^{\text{X}} = \text{X25519}(\text{sk}_e^{\text{X}}, \text{pk}_r^{\text{X}}).$
- (6) Compute $k_{\text{pre2}}^{\text{AES}} = \text{H}(k_{\text{shared}}^{\text{X}} \parallel \text{pk}_r^{\text{X}} \parallel \text{pk}_e^{\text{X}})[0, \dots, 32]$.
Shortening the hash value to 32 Bytes is necessary because of the used symmetric encryption algorithm AES-256.
- (7) Compute $k_{\text{pre1}}^{\text{AES}} = k_r^{\text{AES}} \oplus k_{\text{pre2}}^{\text{AES}}.$

Provided that the used cryptographic hash function generates truly random looking bit strings, one can simplify the generation process to increase runtime performance. The assumption of true random looking bit strings is justified with the input lengths used in Subsections IV-E and IV-F, see [23].

- (1) Generate an ephemeral X25519 key pair:
 $(\text{sk}_e^{\text{X}}, \text{pk}_e^{\text{X}}) \leftarrow \text{Gen}^{\text{X}}.$
- (2) Generate $\text{identification_tag} \leftarrow \text{RandomBytes}(16)$.
- (3) Generate $k_{\text{pre1}}^{\text{AES}} \leftarrow \text{RandomBytes}(32)$.

The shortened generation process is used in the PoC implementation. The number m is randomly chosen in dependence on n , such that $\max\{8, 2n\} \geq m \geq n$.