CRUST: Cryptographic Remote Untrusted Storage without Public Keys

Erel Geron[*] Avishai Wool[†]

July 10, 2007

Abstract

This paper presents CRUST, a stackable file system layer designed to provide secure file sharing over remote untrusted storage systems. CRUST is intended to be layered over insecure network file systems without changing the existing systems. In our approach, data at rest is kept encrypted, and data integrity and access control are provided by cryptographic means. Our design completely avoids public-key cryptography operations and uses more efficient symmetric-key alternatives to achieve improved performance. As a generic and self-contained system, CRUST includes its own in-band key distribution mechanism and does not rely on any special capabilities of the server or the clients. We have implemented CRUST as a Linux file system and shown that it performs comparably with typical underlying file systems, while providing significantly stronger security.

1 Introduction

Network-based storage systems necessarily reduce the trust that can be placed in the storage infrastructure. For instance, the storage server may be outsourced or shared with other individuals. Moreover, the server may be vulnerable to network-based and physical attacks. Despite this, many existing solutions rely on the remote file server for data integrity and access control. The data in these solutions is often stored unencrypted, and the users rely on the server's access control. This means that users effectively trust the file server's administrators, and data may be exposed from backup copies or stolen hard-disks.

Eliminating the trust in the storage server introduces several security problems. End-to-end security is required, including data secrecy, data integrity, authenticity and access control. Existing solutions that take on these challenges, such as SiRiUS [GSMB03] and Plutus [KRS$^+$03], rely heavily on the use of public-key cryptography. It is widely known that public-key cryptography algorithms are orders of magnitude slower than their symmetric-key counterparts. This fact encourages the design of a new file system that avoids using public-key cryptography and uses symmetric-key alternatives instead.

This paper introduces *CRUST*, a new stackable file system layer designed to provide secure file sharing over untrusted servers. CRUST is intended to be layered over any existing file system, even a system that does not offer file sharing at all. The underlying file system is not modified, but is rather extended through the CRUST layer. Our design completely avoids the use of public-key cryptography in order to achieve

[*] Erel Geron is with the School of Electrical Engineering, Tel Aviv University, Ramat Aviv 69978, Israel. erel.geron@gmail.com

[†] Avishai Wool is with the School of Electrical Engineering, Tel Aviv University, Ramat Aviv 69978, Israel. yash@acm.org

better performance than existing systems, without sacrificing security properties.

CRUST stores the data in encrypted and signed form, so it needs a key distribution mechanism. We chose not to use a Kerberos-like on-line trusted key distribution center (cf. [SNS88]). Instead we use the Leighton-Micali key pre-distribution scheme [LM93]. This approach enables file sharing without the need for a secure server or the intervention of the system administrator, and does not require on-line communication between the users.

Like SiRiUS [GSMB03], CRUST offers flexible sharing policies by maintaining per-user access privileges for each file and differentiates between file ownership, read-only and read-write privileges. However, instead of relying on the asymmetry between the secret signing key and public verification key, we used a MAC-based signature scheme [NSW05]. Furthermore, CRUST supports efficient user revocation, performing random access (both read and write) to different parts in a file and maintaining files of varying sizes. Efficient user revocation is performed through the use of a novel key regression mechanism that avoids the use of long hash chains.

CRUST is designed and implemented in a portable way, requiring a minimal installation process and supporting any underlying file system. The user is required to keep a small number of keys, which can be easily managed. The keys are provided by the user only when he mounts the file system. This allows existing applications to operate normally on the mounted file system without the user's intervention.

We implemented CRUST over the FUSE framework [Sze] on Linux. We then performed an extensive performance evaluation. Our results show that CRUST performs very well with only 2% overhead for reading large files and 8% for writing.

The rest of this paper is organized as follows. In Section 2 we define our goals and assumptions. Section 3 presents the mechanisms and design of CRUST. Section 4 describes the data organization used in CRUST. Section 5 presents the implementation of CRUST, and Section 6 evaluates its performance. Section 7 presents several extra features and security enhancements that can be added to CRUST. We discuss related work in Section 8 and summarize our results in Section 9. A novel key regression method is described in Appendix A. The details of basic CRUST operations appear in Appendix B.

2 Design Requirements

2.1 System Considerations

Stackable file system. CRUST must function as an add-on, providing secure file sharing over existing, unmodified file systems. It should support any system with the basic capability of storing files. This requirement is crucial when the user has no control over the file system, as in the case of using storage services on the web. Moreover, it also enables utilizing CRUST for a wider variety of uses, such as securing removable storage devices.

File sharing. CRUST provides flexible per-file sharing, based on the taxonomy of [RKS02], where every user can act as one of the following players:

- ***Owner*** — The owner, who creates the file, can read, modify and delete it. The owner provides read and write permissions to other users, and may also revoke users' privileges.
- ***Reader*** — The readers are permitted to read the file's data.

- ***Writer*** — The writers are permitted to read and modify the file's data.

This is similar to the approach taken by SiRiUS [GSMB03], which relied heavily on public-key cryptography in order to distinguish between readers and writers. Our goal was to provide the same functionality using only symmetric cryptography.

Performance. CRUST should perform comparably to its underlying file system, minimizing access time and storage space overheads. In particular, we require random access, which is the ability to access arbitrary parts of a file without processing the entire file. User revocation is another operation that needs special care in order to be done efficiently in cryptographic file systems.

Key management. CRUST users should be able to access the file system securely using just a small number of keys — and still allow per-file, per-user access control with a read / write / own granularity. Key exchange should not require any on-line message exchange.

2.2 Security Considerations

Data confidentiality, integrity and authenticity. File data must be unreadable to unauthorized users, despite having access to the physical storage device. No entity should be authorized unless explicitly granted by the file owner. In particular, the storage server's administrator is not trusted. Note that an attacker with access to the physical storage device can, of course, erase or modify the encrypted data. However, unauthorized modifications should be detectable by the CRUST clients.

Meta-data confidentiality, integrity and authenticity. File meta-data must also be protected against unauthorized modifications. Unprivileged modifications to meta-data should be detectable, as in the case of file data. However, we do not require complete confidentiality of meta-data, assuming that the non-confidential part does not reveal any sensitive information. Specifically, file names, directory structure and file sizes are not regarded as secret in the current version of CRUST. Access lists and file names can be encrypted using the methods described in Section 7.1. Implementing such features in CRUST is left for future work.

Cryptographic access control. CRUST must enforce per-file access control. Users should not rely on any access control performed by the server. Access control is enforced by cryptographic means, and involves the three security requirements mentioned above: confidentiality, integrity and authenticity of file data.

Key distribution. CRUST is based on secret key cryptography that requires key distribution. However, we chose not to rely on out-of-band mechanisms (unlike Plutus [KRS$^+$03]), and not to use a Kerberos-like key distribution center [SNS88]. Instead, each user shares a long-term key with every other user. These keys are distributed efficiently using the Leighton-Micali key pre-distribution scheme [LM93], which uses public meta-data stored on the server. File-specific keys are provided by the file owner to the file readers and writers via the file's meta-data (encrypted separately using the keys shared by the owner and each other user). Thus, a secure on-line channel between the users is not required. Moreover, this approach allows privilege modifications to be done when the target users are off-line.

Untrusted storage server. In CRUST, users do not rely on the server to provide any level of security. We assume that an adversary may be able to read, change or destroy arbitrary data stored on the server. Confidentiality, integrity and authenticity are achieved using cryptographic means performed by the CRUST clients. Availability of stored data, however, cannot be assured in our threat model. Our work does not present new methods for improving availability; existing methods are described in Section 2.3.

Trusted client machine. Users trust their local machines to securely handle their data and keys. However,

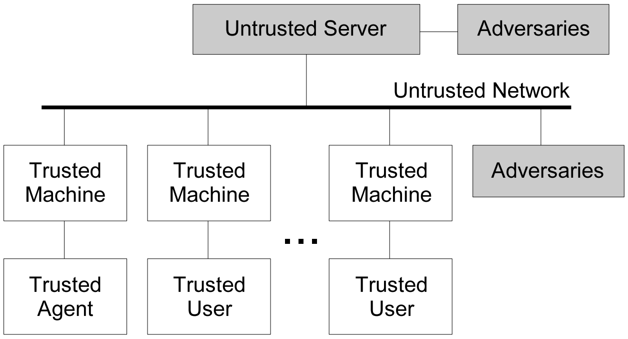

Figure 1: The CRUST trust model. Untrusted entities are grayed.

we do not rely on secure communication between different machines.

A trusted agent during setup. Since CRUST provides its own key distribution mechanism that does not involve direct communication between the users, a trusted agent must set up the system. The agent is trusted to securely keep his keys and to not deceive the users with incorrect keys. Once the system is set up, the trusted agent is not needed any more. In fact, he can forget his keys as long as new users do not need to be added. The agent's possession of the keys allows him to access any information shared between the users; however, we reduce the trust in the agent by preventing him from accessing users' private information. Thus, the agent may only access files that are shared between the users.

We summarize our trust model in Figure 1.

2.3 Adversary Model

Since the server is untrusted in our threat model, it may be fully controlled by an attacker. This control includes reading, writing and destroying arbitrary data stored on the server. The possible attackers include malicious users and other network users, the server administrators and any other people that may have physical access to the server. Since all these entities have the same abilities in our model, even when they collude in a joint attack, we refer to them in the common name — adversaries.

Full control of the server by an adversary allows rollback attacks [MS01]. In these attacks, an adversary replaces all data and meta-data of a file with earlier, valid versions of them. Following this attack, the file users would read stale data and the file access privileges would be restored to their earlier state. Restoring the access privileges prevents recently privileged users from accessing the file, but also revives the privileges of revoked users. We have not implemented counter-measures against rollback attacks — see Section 7.2 for some possible solutions.

Recall that data availability is an open problem in our threat model, because adversaries may destroy the stored data at any time. Preservation of data may be improved, for instance, by replicating it on multiple servers. Currently, CRUST does not address availability issues. Nevertheless, since CRUST is a stackable file system layer, it may easily be layered over existing systems that include availability techniques, such as FARSITE [ABC$^+$02] or OceanStore [KBC$^+$00].

3 System Design

3.1 Overview

Our basic design follows the direction of SiRiUS [GSMB03], but uses the methods of Naor et al. [NSW05]. The latter work identified several symmetric-key alternatives for expensive public-key operations. We incorporate some of the suggested techniques and introduce a complete, practical design of a cryptographic file system for securing untrusted storage without public-key operations.

The data structures in our design can be divided to global structures and per-file structures. The global data structures maintain the list of users in the system and allow key agreement between the users. These structures are organized by the trusted agent, but are not secret. The per-file structures are maintained in two parts. Each part is stored in a different file on the underlying file system: the *data file* and the *meta-data file*. The data file keeps the encrypted file data. The meta-data file contains additional information required for key management and for authentication purposes. The details of all these structures are explained below.

3.2 The User Table

CRUST users refer to each other by their *user names*. Such references occur when, for instance, a file owner decides to grant access privileges to another user. However, it is a common practice in file systems to efficiently represent users in meta-data by using specialized, unique numbers called *user IDs*. The users are normally associated with their IDs upon their registration, by the administrator or by the file server. Translations between user names and IDs often rely on the server.

However, CRUST users do not trust the file server for any purpose, including user ID translations and listing the registered users. Furthermore, we wanted to support file systems with no user management capabilities. Therefore, we decided to supply our own secure method for listing the users and translating between user names and IDs. We call this mechanism the *user table*.

The user table structure consists of a list of entries. Each entry contains a user name and its associated ID. This information is publicly stored on the server, in plain text form, by the trusted agent. We believe that this meta-data does not have to be confidential. However, it must be authenticated in order to prevent attackers from masquerading as arbitrary users.

The trusted agent shares a secret key, $K_i^{\text{User table MAC}}$, with each user i. This key is assumed to be securely exchanged between the trusted agent and each user during system initialization. The key is derived from a master key, denoted by $K_{\text{Master}}^{\text{User table MAC}}$, that is generated and maintained by the trusted agent. The shared key is derived using a one-way function based on the user's ID i, i.e., $K_i^{\text{User table MAC}} = h(K_{\text{Master}}^{\text{User table MAC}}, i)$. The user table is authenticated by the trusted agent for every user, by performing a MAC on the table contents with their shared key. The array of MACs is appended to the stored table. Each user checks the authenticity of the table by verifying his own MAC. Note that the symmetry of MACs requires that each user shares a separate key with the administrator, so that one user cannot be deceived by other users who can modify the table and calculate the correct MACs.

The user table is dynamic, i.e., the trusted agent may add or revoke users at any time. However, we require that the same ID is never reused for different users, even if users are revoked, in order to prevent conflicts between different versions of the user table. This requirement is also crucial for the security of the users' key derivation procedure that is described in the next section. CRUST assures this requirement by allocating monotonically increasing IDs to new users. For this purpose, the trusted agent keeps track of the last allocated ID and increases it every time a new user is added.

3.3 Key Distribution

Sharing encrypted files between several users requires some form of key distribution. Whereas Plutus [KRS+03] depends on an external out-of-band mechanism for key distribution, we argue that such a requirement is not practical in many cases. We prefer the alternative concept of in-band key distribution, which was used in SiRiUS [GSMB03].

This solution uses a file's meta-data for distribution of all the keys relevant to that file. Each user with some access rights to the file is allocated meta-data space, which we call a *lockbox*. The lockbox contains encrypted information shared with the file's owner. In SiRiUS, the information is encrypted using the target user's public key, and relies on a public-key infrastructure (PKI) for exchanging users' public keys. CRUST uses an alternative symmetric-key method, in which every pair of users shares a *common secret key*. For exchanging these keys, we use one of the methods of Leighton and Micali [LM93]. A brief description of the protocol follows (see also [Rub00], [NSW05]):

The trusted agent generates two *master secret keys*, K and K'. Using these keys and a pseudo-random function $h(\cdot)$, the trusted agent computes and provides each user i with his *exchange key* K_i and his *individual authentication key* K'_i, where:

$$K_i = h(K, i), \; K'_i = h(K', i).$$

In practice, we use the HMAC algorithm as the pseudo-random function. These keys are given to each user during system setup in a secure, out-of-band method. Additionally, a public database is published by the trusted agent. It contains two matrices, P and A, including *pair keys* and *authentication keys* respectively, such that:

$$P_{i,j} = h(K_i, j) \oplus h(K_j, i), \; A_{i,j} = h(K'_i, h(K_j, i)).$$

Each user i that wants to encrypt information for user j computes the common secret key $K_{i,j}$ according to:

$$K_{i,j} = P_{i,j} \oplus h(K_i, j) \;\; (= h(K_j, i)) \, .$$

User i can verify the authenticity of the key by ensuring that:

$$h(K'_i, K_{i,j}) = A_{i,j}.$$

The decrypting user j can calculate $K_{i,j} = h(K_j, i)$ without reading the matrices.

In our scenario, when a file owner, user i, shares a file with user j, he creates a lockbox for user j in the meta-data file, encrypted with the key $K_{i,j}^{\text{Enc}}$ and authenticated by applying a MAC with $K_{i,j}^{\text{MAC}}$, where:

$$K_{i,j}^{\text{Enc}} = h(K_{i,j}, \text{“Enc”}), \; K_{i,j}^{\text{MAC}} = h(K_{i,j}, \text{“MAC”}).$$

User j derives these keys from $K_{i,j}$ in the same manner. The lockbox contains file-specific meta-data needed by user j in order to access the file, as we further explain in Section 4.2.

A file owner also needs to store private information that is not shared even with privileged writers. For example, Section 3.6 shows that the owner stores some information allowing derivation of future encryption keys to be used when a reader or writer is revoked. Therefore, the owner maintains a private lockbox as well. Encryption and MAC keys are used for encrypting and authenticating the owner's lockbox, just as any other lockbox. However, although an owner i could use the keys $K_{i,i}^{\mathrm{Enc}}$ and $K_{i,i}^{\mathrm{MAC}}$, which are a part of the key distribution scheme, we prefer that every user generates his own keys for this purpose. These keys, denoted by $K_i^{\mathrm{Self\ enc}}$ and $K_i^{\mathrm{Self\ MAC}}$ respectively, are generated by the user upon initialization and are kept securely. They allow storing secure private information that is unavailable even to the trusted agent.

3.4 Access Control

Data confidentiality in CRUST is maintained by encryption of data at rest performed by the client. Each file is associated with its own encryption key. This key is generated by the file owner upon the file's creation and is securely distributed to the file readers and writers through their encrypted lockboxes. Unprivileged users cannot obtain the key, since they cannot decrypt other users' lockboxes.

Data integrity and authenticity in earlier cryptographic file systems, such as SiRiUS [GSMB03] and Plutus [KRS$^+$03], are provided by public-key digital signatures. A signature on the hash of a file proves that the file was written by an authorized writer having the private key. Readers and writers are distinguished by the asymmetry between the signature's private and public keys. In CRUST, we instead use an efficient MAC-based symmetric-key signature scheme, as suggested in [NSW05]. Note that although public-key signatures achieve the additional property of *non-repudiation*, we believe that it is not a common requirement in file systems.

Since MACs are symmetric by themselves, we need a special construction to provide writer-reader differentiation. In CRUST, the encrypted file data is signed by the file writers and is verified by both the file readers and writers. One MAC is stored in the meta-data for the writers, and an additional MAC is stored for each reader. A signature operation involves storing all the MACs, whereas a verification operation involves calculating and comparing a single MAC. The file owner randomly generates a *file master MAC key*, denoted as $K^{\mathrm{File\ MAC}}$, during file creation. This key is distributed to the writers through their lockboxes. Each reader is handed a separate *file reader MAC key*, denoted as $K_i^{\mathrm{File\ MAC}}$, where i is the reader's user ID. The reader keys are derived from the master key by applying a one-way function, i.e., $K_i^{\mathrm{File\ MAC}} = h(K^{\mathrm{File\ MAC}}, i)$. This scheme simplifies key management, because writers can compute the MAC keys for all readers, based on the master key and the list of readers.

Signature time and space grow with the number of readers, while verification time is always as short as one MAC computation. The space consumption for the signature is acceptable in our case, because we maintain a per-reader lockbox anyway.

3.5 Random Access

Efficient random access is an important feature in file systems, because it improves the performance of the system in some applications by orders of magnitude. In order to support random access, both encryption and authentication need to operate in smaller chunks than the entire file. For this purpose, CRUST divides the file data into blocks of a predetermined size (we used 4096 bytes). Encryption is performed on each data block independently. More details about the encryption techniques are given in Section 4.1.

Authentication of file data by signing the hash of the entire file prevents efficient random access since it requires the file to be entirely obtained in order to verify or update a single block of data. Instead, we use a

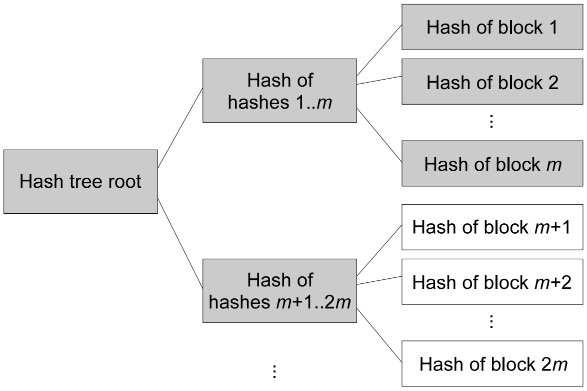

Figure 2: The hash tree structure, where each internal node has m children. The grayed nodes need to be accessed for verifying any block in the range $1..m$.

more efficient scheme by hashing the data in a *hash tree* construction; each leaf of the tree stores the hash of a single block of data, and each internal node stores the hash value computed on the concatenation of the values in its children. The hash tree is stored as a part of the file's meta-data and is not secret because the data blocks being hashed are in encrypted form. The file data is authenticated by signing the tree's root instead of the hash of the entire file. Thus, verifying a block of data only requires: (a) calculating its hash, (b) comparing it to the hash in the relevant leaf of the hash tree, (c) checking the consistency of the hashes on the branch connecting the leaf to the root, and (d) verifying the signature of the root. The hash tree structure is demonstrated in Figure 2.

3.6 Revocation without Re-encryption

The event of user revocation modifies the file's access permissions. Thus, keys currently used for encrypting the file must be replaced. However, re-encryption of the entire file with the new keys is an expensive process. The concept of *lazy revocation* states that re-encryption may be delayed until the next update to the file. Moreover, re-encryption may be performed partially, just for the file blocks that are being updated. This concept makes sense because the revoked users may already have read all data stored prior to their revocation.

In CRUST, the lifetime of a file is divided into *epochs*, where each revocation begins a new epoch. New file encryption and signature keys are generated at the beginning of a new epoch. Updating the file is always done by writing blocks that are encrypted with keys of the latest epoch, whereas reading may require older keys, depending on the epochs when the relevant blocks were written. Therefore, older keys must be available to the users as long as they are being used. However, associating each epoch with its own independent key implies that the space required for maintaining the keys would grow linearly with the number of revocations. A more efficient scheme would use a relation between the keys, so that the most recent key can be used to derive the key of any earlier epoch.

Plutus [KRS$^+$03] introduced the *key rotation* technique, where the owner uses a dedicated private key to generate a new encryption key from the current one. The most recent encryption key is handed to the users of the file. The users can derive all previous keys from the current one by using the dedicated public key. Plutus uses key rotation for a different purpose than ours, where the epoch (or version) was a property of each file in a certain group of files.

The approach we take in CRUST is based on the more generalized paradigm of *key regression* [FKK06]: instead of providing the users with the latest encryption key directly, we provide them with a *state*, which is a small structure that allows computation of the current key and all earlier keys. We use this paradigm to introduce a novel key regression mechanism. Our mechanism maintains states of size $O(\log n)$ and requires $O(\log n)$ symmetric-key operations for deriving keys of arbitrary epochs, where n is the maximal number of epochs. In CRUST, a 2^{28}-epoch system is used. The state size is 140 bytes (7 SHA-1 hash values) and at most 105 keyed-hash computations are needed to initialize the key regression system and to derive any legitimate key or state. Full details of our key regression method appears in Appendix A.

The infrastructure for using a key regression mechanism in CRUST involves sharing the latest key regression state with each of the users that are privileged to access the file. The owner provides the key regression state in the users' encrypted lockboxes. Moreover, in order to keep track of each block's epoch, an array of epoch identifiers is maintained as a part of the file's meta-data. This array is not secret, but it requires authentication; see Section 4.2. Existing blocks' encryption keys are derived from the current state. Newly written blocks are encrypted with the key of the most recent epoch, hence a revoked user cannot decrypt them.

3.7 Owner-Identifying File Names

The owner of a file must be known in order to access the file properly. For instance, a user must know the identity of the owner so he can select the keys used for decrypting and authenticating his lockbox.

Naive methods of owner identification are vulnerable to the following attack, which was mentioned in SiRiUS [GSMB03]: a malicious user can replace an existing file with a file that the malicious user owns. Unless the true owner is known in advance, users can not detect this situation. Furthermore, any data written by a non-suspecting file writer, following the replacement, would be readable by the malicious user. One solution suggested in SiRiUS requires publishing a small amount of information on a secure server, i.e. deviating from the untrusted storage model. Another solution requires every owner to occasionally verify the ownership of his files. The latter solution is more practical, but is still vulnerable to attacks. For example, a malicious file server may hide the replacement from the true file owner by pretending to store the original file's contents only when being accessed by the owner. Thus, the replacement may not be detected by the file owner.

CRUST uses the following solution, which we call *owner-identifying file names*. The CRUST root directory contains a sub-directory per each user, so that all the files owned by a user are located under a specific directory named after that user. For example, if a Linux user *alice* mounts the CRUST file system at her `~/crust` and she is logged into CRUST as Alice, then she can access a typical file owned by her using the file name `~/crust/Alice/foo.c`. Another Linux user *bob*, logged into CRUST as Bob, can mount CRUST at his own `~/crust` and then he can access Alice's file using the same name, `~/crust/Alice/foo.c`. The per-user directories are created when users are added (i.e., during system setup).

This way, when a file name is known, the owner is automatically determined. Once the true owner is known, a file user can verify the owner's identity by verifying the authenticity of his lockbox, because the authentication key is secretly shared between the file user and the owner. Note that *creating* a file in another user's directory is forbidden by CRUST since the ownership of such a file would be ill-defined.

3.8 Key Management

A CRUST user i only needs to store five long-term keys: K_i, K'_i, $K_i^{\text{User table MAC}}$, $K_i^{\text{Self enc}}$ and $K_i^{\text{Self MAC}}$. The first three keys (the Leighton-Micali keys and the user table authentication key) are securely provided to the user by the trusted agent, whereas the two other keys (the personal encryption and authentication keys) are generated by the client and are kept secret. No further keys are required once the file system is mounted. This way, applications can access the file system transparently, without the user's intervention.

The trusted agent is required to keep track of three long-term keys. The trusted agent's keys are K, K' and $K_{\text{Master}}^{\text{User table MAC}}$ (the Leighton-Micali master keys and the master key for user table authentication), which are generated and kept secret by the agent. The agent can discard (or lose) the keys at any time, but this will prevent adding new users to the system after that moment.

Since each CRUST file maintains its own short-term keys and access control information, a file system backup can be performed by unprivileged users. A backup program can detect the modified files and perform the backup over the underlying file system, without mounting CRUST.

4 Data Organization

4.1 Data File Structure

Recall that a CRUST file is maintained using two files in the underlying file system: the data file and the meta-data file. The data file contains the encrypted file data. To allow random access, the file is divided into blocks, where each block is encrypted independently. The encryption keys are specific to each file. To allow efficient user revocation without file re-encryption, we use block-specific encryption keys (unlike SiRiUS [GSMB03], which encrypts the entire file with the same key). However, unlike Plutus [KRS$^+$03], we do not use *independent* per-block keys and we do not use a stream cipher. Instead, all the block keys are derived from a single key regression state, based on the block's epoch. The key regression state and the array of block epochs are provided in the meta-data file. Our approach yields revocation time that is independent of the file size.

Any block cipher can be used for encrypting the data file. However, since the same key may be used for encrypting multiple cipher blocks, and encrypting the same plain text under the same key produces the same output, not every *mode of operation* assures confidentiality. Our chosen mode of operation follows NIST's recommendation [Dwo01] to use CBC, performed on every file block separately, where the *initialization vector* for encrypting each block is the encryption of the block's offset in the file, with the same encryption key. This mode also allows efficient random access of the file blocks.

CBC mode requires the plain text size to be a multiple of the cipher block size (16 bytes for AES [NIS01]), so CRUST uses the following padding method. Let s denote the plain text size, in bytes, and let $k = s \bmod 16$. We first pad the plain text with $16 - k$ binary '0' bytes, and encrypt the result in CBC mode. To allow determining the original plain text size, the cipher text is then padded with k additional (unencrypted) '0' bytes. The only exception is when $k = 0$; in this case, no padding is used. For example, a

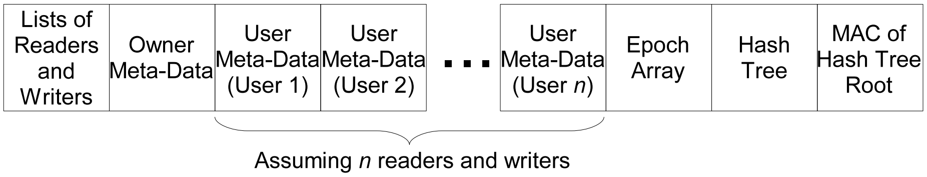

Figure 3: Meta-data file structure. The hash tree root's MAC is calculated using $K^{\text{File MAC}}$.

2-byte file is stored as an 18-byte CRUST data file, in which the last 2 bytes are nulls. This approach allows fast calculation of the original data size, because the actual file size is stored in the data file's *inode*.

4.2 Meta-Data File Structure

All the meta-data needed to maintain a CRUST file is stored in its associated meta-data file. The meta-data files are omitted by the CRUST client when files are being listed, so that the directory structure appears to the user as containing only the data files.

In order to support file data encryption using key regression, the master key regression state, denoted by $\mathit{KRS}_{\text{Master}}$, is privately kept in the owner's lockbox. The current key regression state, denoted by $\mathit{KRS}_{\text{Current}}$ is distributed through the file users' lockboxes. The state includes the current epoch identifier. Additionally, the epoch array, containing the epoch identifier of each data block, is stored publicly in the meta-data file. The epoch array is not secret, but its integrity and authenticity must be verified. Therefore, it is signed along with the file data, as described below.

File signatures that allow efficient random access are achieved by using the hash tree. The tree is stored publicly in the meta-data file. The hash tree root is signed by storing the writers' MAC and an additional MAC for each reader, according to the scheme described in Section 3.4. The MAC keys are also distributed according to the scheme, through the lockboxes.

The user privileges for accessing the file are maintained by storing two lists of user IDs, describing the readers and writers respectively. The lists are stored publicly in the meta-data file, since we assume that they are not secret (but if needed, they can be protected as described in Section 7.1). The access lists, denoted by $\mathit{ACL}_{\text{Readers}}$ and $\mathit{ACL}_{\text{Writers}}$, are signed using a MAC. The lists are also used for ordering the lockboxes in the meta-data file, so that each lockbox can be associated with its user.

File names in CRUST are maintained by the underlying file system, without using encryption. However, the file names must be verified in order to detect file-swapping attacks. Each file's (full) name is signed using a MAC, along with the access lists and each user's lockbox. This implies that CRUST files cannot be renamed or moved except by their owner.

We summarize the meta-data file structure in Figure 3. A part of it includes user-specific information, stored separately for the owner and for each reader and writer. This user-specific meta-data is depicted in Figure 4. Note that a user's meta-data includes his lockbox, which contains secret information. Each lockbox is encrypted and signed with secret keys shared by the owner and the addressed user. The rest of the meta-data in the file is stored once and is not secret.

Note that the only parts of the meta-data that a (non-owner) writer is privileged to modify, and are used

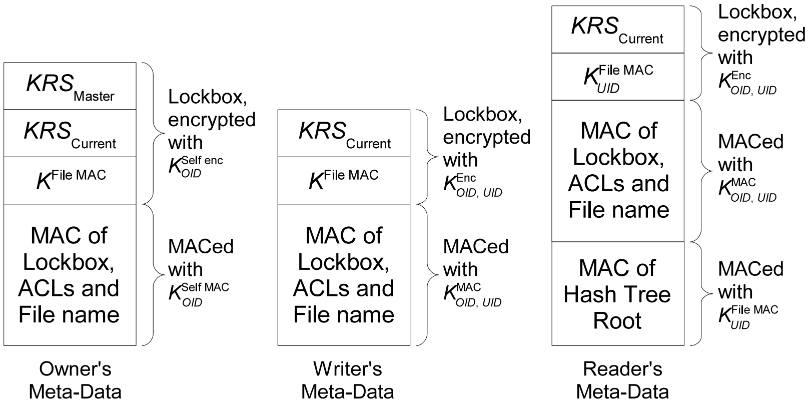
KRS_{Master}
$KRS_{Current}$
$K^{File\ MAC}$
MAC of Lockbox, ACLs and File name
Lockbox, encrypted with $K^{Self\ enc}_{OID}$
MACed with $K^{Self\ MAC}_{OID}$
Owner's Meta-Data
$KRS_{Current}$
$K^{File\ MAC}$
MAC of Lockbox, ACLs and File name
Lockbox, encrypted with $K^{Enc}_{OID,\ UID}$
MACed with $K^{MAC}_{OID,\ UID}$
Writer's Meta-Data
$KRS_{Current}$
$K^{File\ MAC}_{UID}$
MAC of Lockbox, ACLs and File name
MAC of Hash Tree Root
Lockbox, encrypted with $K^{Enc}_{OID,\ UID}$
MACed with $K^{MAC}_{OID,\ UID}$
MACed with $K^{File\ MAC}_{UID}$
Reader's Meta-Data

Figure 4: User-specific meta-data. OID denotes the owner's ID, and UID denotes the addressed user's ID.

by other users, are the epoch array, the hash tree and the MACs of the hash tree root. Readers, on the other hand, cannot feasibly modify any meta-data that is used by another user, without the modification being detected by that user. The owner can detect all modifications to the lockboxes, because they can be derived from information in the owner's lockbox.

4.3 Global Meta-Data Structures

CRUST stores two public data structures in a constant location on the file server, which are accessed on demand by the users. These structures are the user table and the Leighton-Micali structure (recall Sections 3.2 and 3.3). The user table structure consists of user name and ID pairs, and an array containing a MAC of the table for each user. The structure for operating the Leighton-Micali protocol consists of two matrices, containing the pair key and authentication key for every pair of users.

Note that storing and maintaining these Leighton-Micali and user management structures are not necessary if dynamic user management is not required. The system will keep functioning properly even if the structures are not available, as long as new users are not being introduced to the system. In a static user management situation, the structures can be distributed to the users during system initialization, instead of storing them on the server.

5 Implementation Details

5.1 Overview

We implemented CRUST on Linux using the FUSE (Filesystem in Userspace) framework [Sze] and the OpenSSL cryptographic library [Ope]. FUSE provides an interface for user-level programs to export a virtual file system to the Linux kernel. It also enables non-root users to mount their own file systems. FUSE consists of a kernel module, a user-level library and a mount utility. The user-level file system communicates with the FUSE kernel module through a dedicated device file.

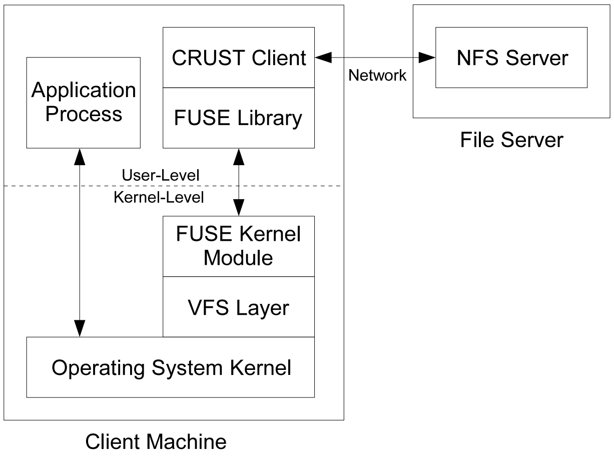

Figure 5: Architecture of CRUST layered over NFS. Note that NFS is just an example. CRUST can be layered over any file system.

CRUST is implemented as a user-level client providing the data and meta-data of the file system through the FUSE interface. Installing the CRUST client requires the FUSE kernel module to be installed by the root user. However, since Linux kernel versions 2.6.14 (released in 2005) and later include FUSE support out of the box, CRUST can also be installed by a non-root user on recent enough Linux operating systems.

We designed CRUST as a stackable file system layer [HP94]. It appears to the user as an ordinary file system, while functioning as a layer over another file system. Requests are processed by the user-level client by communicating with the underlying file system. As an additional layer, CRUST provides strong security to any insecure file system. The underlying system is not known in advance and it is not modified. It is only assumed to have basic functionalities. In fact, it may be any local or network file system, such as ext2 or NFS. Figure 5 shows the general architecture of CRUST in a typical scenario.

The OpenSSL library supports many cryptographic primitives that can be used for the various mechanisms of CRUST. The default options used by CRUST are the following widely-used primitives: SHA-1 [NIS04] as the cryptographic hash function, HMAC [KBC97] based on SHA-1 as the MAC algorithm and AES-128 [NIS01] as the block cipher.

5.2 Software Components

CRUST was implemented according to the design presented in Section 3. The implementation includes a mount utility (`crust-mount`), a file sharing utility (`crust-chmod`) and an administration tool `crust-admin` for the trusted agent. A user logs into CRUST by executing `crust-mount` and providing the CRUST mount point (e.g., `~/crust`) and the underlying file system's mount point (e.g., `/nfs`). The mount utility runs a CRUST client daemon in the background, which intercepts and handles Linux file requests on the CRUST mount point. FUSE guarantees that the mount point is private for the mounting Linux user; additional users on the same machine must mount CRUST on other mount points by running the mount utility (and daemon) independently. Note that the mounting Linux user can identify himself as any CRUST user, as long as he keeps possession of the required keys.

Per-file access permissions are managed using `crust-chmod`. This tool resembles the Linux `chmod` command, but is more expressive, because the file access privileges of each user are controlled individually in CRUST. Command-line options of this tool include `add-writer`, `add-reader` and `revoke-user`.

The system is initialized by the trusted agent using the `crust-admin` tool. The same tool also provides user management functionalities and publishes the public structures of the key distribution protocol when a user is added or removed from the system. Command-line options of this tool include `init`, `add-user` and `remove-user`.

The client daemon (which is started by `crust-mount`) consists of two main components: user interaction and file management. The user interaction component is responsible for securely listing the system's users and providing translations between user names and IDs, as well as computing the common secret keys shared with each user. This functionality is handled by reading and authenticating the public meta-data structures stored by the trusted agent. The user interaction component also manages the user's master keys, as described in Section 3.8.

The file management component of the client implements all the file-specific operations. Standard operations such as file opening, reading and writing are directed to FUSE through the *Virtual File System* (VFS) interface [Kle86]. File sharing operations in CRUST, however, are more expressive than the standard interface and are therefore directed in a special manner; `crust-chmod` expresses sharing operations as specially-formatted Set Extended Attribute (SETXATTR) operations, which are available in the VFS interface; `crust-chmod` encodes the various parameters of a file sharing request, as given in the command-line, into one string. This string is used as an "attribute" being set for the relevant file by initiating a SETXATTR system call. When the CRUST client intercepts this call, it decodes the file sharing command parameters from the attribute string and finally executes the command.

5.3 Optimizations

5.3.1 Caching

Caching of data and meta-data has a great impact on the performance of file systems. It can save precious time by preventing redundant cryptographic and file operations. For instance, consider an application reading a 4 KB CRUST file, one byte at a time. Since cryptographic operations are performed on 4 KB blocks, reading each byte separately involves reading, decrypting and authenticating the entire block every time. A simple data cache would perform this lengthy process only for the first byte in the block, and keep the obtained data for instantly handling the next requests.

CRUST maintains several caches. Note that all caches are maintained by the client software, so each user utilizes his own caches. We employ a data cache that keeps recently read blocks of file data, as demonstrated in the above example. It is a *write-back* cache, in which modified data is kept in memory and is not written or even encrypted until absolutely necessary. This increases the probability of writing full blocks of data, and thus reducing redundant operations.

The hash tree is equipped with a specially designed cache. It caches entire blocks of hashes in every level of the tree, while verifying the consistency of every cached block. Recall that the consistency of each node in the tree depends on the matching of its value to the hash of its children's values. The integrity of each data block involves the consistency of the corresponding leaf and all of its ancestors. Since ancestors are common to many nodes, caching them prevents many redundant verifications.

The key regression mechanism keeps track of recently derived keys, and thus improves the performance for reading blocks of the same epoch. The user interaction component caches the common keys, so that each key is only calculated once. It also caches the entire user table. This makes sense because the table must be entirely read anyway in order to verify its integrity and authenticity.

5.3.2 Multi-Threading

Another important feature of our file system implementation is allowing multiple processes to efficiently access numerous files at the same time. Handling multiple requests concurrently is achieved by multi-threaded programming of the CRUST client. CRUST handles each request by a separate thread. Multi-threading allows for better utilization of the system's resources, because a thread that is waiting for a file operation to complete can be switched with a thread having intensive computations. Furthermore, modern computer systems that have multiple CPUs, or CPUs with multiple cores, allow the threads truly concurrent execution.

The basis for multi-threading support is implemented by the FUSE library, which allows separate requests to run in separate threads. CRUST synchronizes the different threads by enforcing mutual exclusion on the data structures that are shared between them (using POSIX *pthread* mutexes). Coherency between threads accessing the same file is assured in CRUST by sharing all the in-memory data structures that are relevant for the file. By sharing the caches, threads accessing the same file gain a performance improvement, since redundant operations are omitted. Note that the sharing of caches is legitimate because the different application processes addressing the client represent the same CRUST user, thus they share the same privileges.

5.3.3 Meta-Data Organization

Further optimizations were achieved by improving the meta-data organization. The epoch array and the hash tree leaves are accessed correspondingly, because both are required for reading and writing any block of file data. Therefore we decided to store them interleaved, such that a single block of meta-data contains both epoch information and hash tree information relevant for several data blocks. This scheme improves the delay of our system because fewer file operations are necessary in order to read a single byte. It may also improve the system's throughput due to the increase in space locality of the meta-data access sequences.

The meta-data file contains several dynamically sized structures, such as the array of lockboxes and the hash tree. Storing such structures adjacently in the same file introduces an organization problem. The location of each structure in the file must be managed. Moreover, changing the size of a structure requires a reorganization, which may involve moving a large amount of data from one position to another. When frequent changes occur, such as the case of constantly appending data at the end of a CRUST file, reorganization of meta-data can become very inefficient.

Our solution uses a small index of locations and sizes for organizing the various structures. This index is saved at the beginning of the meta-data file. To avoid frequent reorganizations, the structures are allocated more space than immediately needed. A doubling technique multiplies the allocated space by two when additional space is needed and by half when only a quarter of the space is utilized. Note that the organization problem can also be solved by storing each structure in a separate meta-data file. However, this simplistic solution adds the overhead of maintaining multiple file handles, and does not scale for systems containing large numbers of files.

5.3.4 Optimized Constants

The information in the data file is always read in chunks of a single block, because the whole block is necessary for decrypting and authenticating each byte in the block. A similar situation occurs for writing. Having a large block size increases the delay required for handling a single read or write request, whereas having a small block size increases the space overhead of our meta-data structures. CRUST uses a block size of 4 KB, to match the default block size in most Linux file systems; this way, we reach a reasonable balance between the delay and the space overhead of our system.

Similarly, the hash tree nodes (except the root) are also accessed in chunks, because verifying each internal node requires obtaining all of its children, which are stored adjacently on the disk. We defined the internal nodes' maximal number of children (the fan-out) so that each node's children fit in a 4 KB block. For instance, when SHA-1 is used for hashing and when the hash tree leaves are interleaved with the blocks' 32-bit epoch identifiers, their parents' fan-out equals to 170.

The number of nodes on the paths from the root to the leaves (the height) of the hash tree also affects performance, because an entire path is accessed in order to authenticate a single block of data. CRUST limits the height of the hash tree to three nodes, including the root and the leaves. Note that the height limitation is possible because the root's fan-out is unlimited. CRUST caches the root and all of its children. Having this information cached and verified in advance, only a single block of hashes needs to be read and hashed in order to verify any randomly read data block.

5.4 Coherency

Caching improves performance, but it comes at the price of coherency problems between clients accessing the same file simultaneously. For instance, when writing is performed, a reader may read stale data due to caching on his side or due to the writing being delayed by the writer. However, a coherency problem would occur in our case even without caching, because CRUST reading and writing are not atomic operations. Each of those operations requires performing several actions on the underlying file system, which may get mixed with other clients' actions.

To resolve such problems, CRUST locks the files while such operations are performed. The files are locked using the `fcntl` system call, which is supported by many file systems. File reading in CRUST involves acquiring a shared lock, whereas writing involves an exclusive lock. Shared locks can be acquired by an unlimited number of clients at the same time, but an exclusive lock cannot coexist with another lock of any kind. A reading client waits for writers to finish their tasks, whereas a writing client waits for every client currently accessing the file.

Locking and unlocking files for every read and write operation can be very inefficient, especially when the files are stored on a remote server. Thus, CRUST acquires locks when the files are opened and releases the locks only when the access is finished. The file locking procedure is hidden from the CRUST users; an operation issued by the user completes after the necessary waiting is done. The use of file locking assures coherency as long as the system is legitimately used; however, it is impossible to guarantee coherency when various kinds of attacks occur.

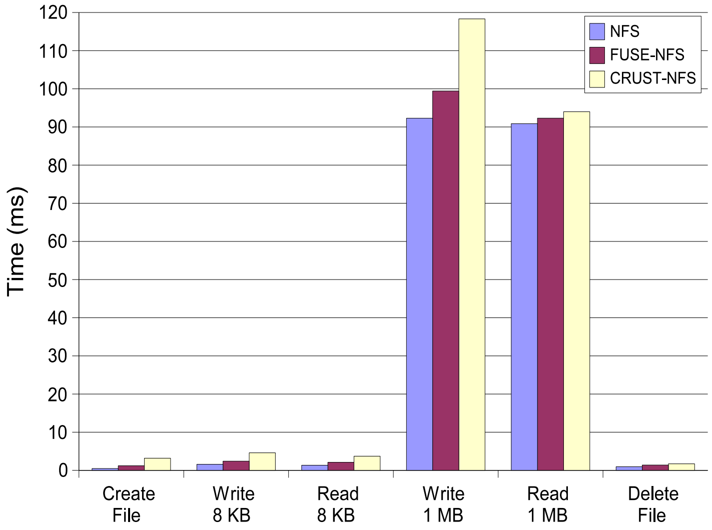

Figure 6: Micro-benchmark results. The NFS columns show the times for a standard NFS system. FUSE-NFS means a pass-through FUSE over NFS system, and CRUST-NFS shows our system's times.

6 Performance Evaluation

We performed a series of tests in order to compare the performance of CRUST layered over NFS to native NFS. Some of the tests also evaluated a simple FUSE pass-through file system client. The pass-through file system was used for estimating how much of the overhead of CRUST is due to implementing it in user-level.

Our experimental setup included a server machine and a client machine. The NFS server (using NFS protocol version 3) was run on a 1.8 GHz Pentium 4-M machine with 256 MB memory. The clients were run on a 2.4 GHz Pentium 4 machine with 512 MB memory. The operating systems of the server and the client were Fedora Core 4 Linux (version 2.6.11) and Fedora Core 6 (version 2.6.18), respectively. The machines were connected by a 100 Mbps Ethernet link. The CRUST code was written in C++ and includes about 5000 lines of code.

6.1 Micro-Benchmarks

The micro-benchmarks were primarily designed to compare between CRUST and NFS. They also allow us to compare the performance of CRUST to the extrapolated performance of two related works — SiRiUS [GSMB03] and Plutus [KRS$^+$03]. Our benchmarks included similar operations to those tested in the SiRiUS and Plutus micro-benchmarks. Our tests included creating files, deleting files and sequentially reading and writing files of different sizes. Each test was executed a hundred times on different files. The average results are shown in Figure 6.

The figure shows that CRUST reads a 1 MB file with a time overhead of as low as 3% relative to NFS. Writing a 1 MB file takes about 28% more time than on NFS. In comparison, SiRiUS performs the same operations with a slowdown of $2.3\times$ for read and $6.3\times$ for write over NFS. Although the latter slowdown is mostly affected by technical limitations of the protocol (NFSv3) used in the SiRiUS file system interface, we can still see a clear performance improvement in CRUST. Plutus incurs an overhead of about 3 seconds in comparison to its underlying file system for reading and writing a 40 MB file. As an estimation, we divide this time by 40 to match the 1 MB file we used, and assume that our processors are about 2× as fast. This optimistic estimation concludes that Plutus would have achieved a time overhead of about 40 ms, or 44%, if it was run in our experimental setup. Thus, these results demonstrate a considerable improvement.

For small files, reading and writing operations in CRUST are about 2.8 times slower than the NFS equivalents. This overhead mostly consists of meta-data access times and context switches. The cryptographic overhead is small due to avoiding public-key operations. Optimizations included in CRUST, such as improved meta-data organization for small files, also contribute to the performance measured in these tests. These results demonstrate a considerable improvement relative to SiRiUS, which performs the same operations with approximately a 20× slowdown, as compared with NFS. The authors of Plutus did not include tests on small files.

Figure 6 shows that CRUST file creation is 6.5 times slower than NFS, because it involves key generation and initialization of both the data file and the meta-data file. However, it outperforms the 36× slowdown factor of SiRiUS as a result of avoiding public-key operations. File deletion in CRUST is slightly slower than in NFS because it performs deletion of two files, compared to just one when using NFS. The figure also shows that the overhead of the pass-through FUSE file system in all tested operations is small relative to our system's overhead. This observation suggests that the user-level implementation of CRUST has little effect on the results.

6.2 The Bonnie Benchmark

Bonnie [Bra96] is a well-known file system benchmark. It performs a series of tests on a single large file. The tests are:

- *Sequential output.* The file is created and written one character at a time. Then, it is written again block by block. Subsequently, each block is read, modified and rewritten.
- *Sequential input.* The file is read one character at a time. Then, it is read again block by block.
- *Random seeks.* Several processes concurrently seek to random locations in the file and read one block at a time. The block is modified and rewritten in 10% of the cases.

We executed Bonnie with a file size of 1 GB, which is twice the amount of memory on the client machine. This reduces the amount of read requests that are trivially resolved using the operating system's page cache. Bonnie measures the throughput of the operations, as well as the percentage of CPU usage. The CPU usage results are ignored in our evaluation because they do not include the overhead of the CRUST client process.

The sequential input and output results are shown in Figure 7. For each test, we compare NFS and NFS via CRUST. The performance is measured in transfer rate, as reported by Bonnie.

CRUST performed block reads with only a 2% time overhead, as compared to NFS. Block writes were performed with an 8% overhead. These results are indeed better than those presented in the micro-benchmarks, because the larger file size increases the effect of caching and optimizations. Character reads and writes in CRUST were performed 3% and 13% slower, respectively. Rewriting was performed by CRUST with a 2% time overhead.

The random seek results of the Bonnie benchmark are presented in Table 1. CRUST performed random

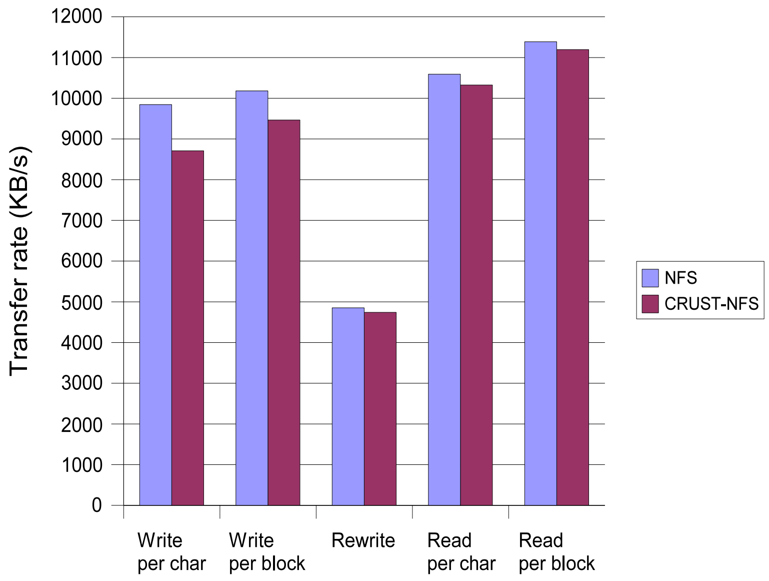

Figure 7: Bonnie benchmark results.

System	Operations per second
NFS	93.4
CRUST-NFS	22.8

Table 1: Bonnie random seek benchmark results.

seeks with a $4.1\times$ slowdown. Random seeks were expected to have worse performance than sequential access because they eliminate the effect of most caches used by our system. Moreover, random seeks in a large CRUST file translate to random seeks in both the data file and the meta-data file, accumulating overhead from both seeks. Therefore, the above result is satisfying, and proves that the random access optimizations included in CRUST pay off.

6.3 Privilege Modifications

In this section we present a performance evaluation of the operations that modify file access privileges in CRUST. Since our per-user privileges are more expressive than standard UNIX file permissions, a concrete comparison with NFS is not possible for these operations. Instead, we compare the performance of permission modifications relative to other CRUST operations.

Write permissions are expected to be slightly more expensive to change than read permissions, because changing them involves additional actions. Specifically, a new master MAC key is generated only when a writer is revoked. This change requires both modifying all the lockboxes and recalculating the hash tree MACs, and thus may incur a longer delay. Therefore, we focus on evaluating modifications of only write privileges.

Our test creates a 1 MB file, grants 1000 users write permissions to the file and then revokes each one of them. The time for granting permissions to each user is measured and the times are averaged. Revocation time is measured in the same way. The results are presented in Table 2.

Operation	Time (ms)
Permission granting	31.5
Revocation	32.3

Table 2: Time for changing permissions in CRUST.

The results show that both modifications can be performed more than 30 times per second. Moreover, their performance is comparable to the time required to read a few hundreds of kilobytes from an ordinary CRUST file. Thus, the results are acceptable.

Granting access permissions to a user requires CRUST to calculate and add a new record to the meta-data file, recompute meta-data MACs and rarely reorganize the meta-data file. Revocations, on the other hand, involve several additional actions, such as changing keys and performing key regression operations. Thus, we indeed expected that revocations would require more time than permission granting. The fact that the time difference between the two operations is only 3% confirms that the key regression mechanism is efficient and valuable for the system.

6.4 Lazy Revocation

Lazy revocation using key regression has the benefit of making revocation operations extremely efficient. However, it comes at the expense of possibly harming a file's access performance following a large number of revocations. This phenomenon may occur when each block in the file is modified in a different epoch, and the epoch variance throughout the file is large. Calculating the encryption keys for each block would take a longer time for such a file than for a file entirely written in a single epoch, where the keys are equal for each block.

In this section we compare the access performance of a revocation-free file with a revocation-rich file. Our test begins by creating the two files, each filled with 100 MB of data. The first file is created normally, without special permission modifications. The other file is manipulated as follows: a specific user is alternately granted write permissions and revoked from the file; this process is repeated a million times, while writing data to a random block in the file following each revocation. The random blocks are chosen in advance such that the resulting revocation-rich file has blocks of random epochs picked uniformly from a wide range of a million epochs. After initialization, we compared the two files for performance of sequential reading and overwriting. The results are shown in Figure 8.

The revocation-rich file had only a 1% slowdown in read performance, as compared to the revocation-free file. The performance difference for overwriting is negligible — less than 0.1%. This makes sense because overwriting entire blocks does not involve decrypting the original blocks, but instead encrypts the new data with the current key. As we expected, the variance of the blocks' epochs had a negligible effect on the performance of accessing the file. This benchmark further demonstrates that lazy revocation using key regression is a practical solution, even in extreme conditions.

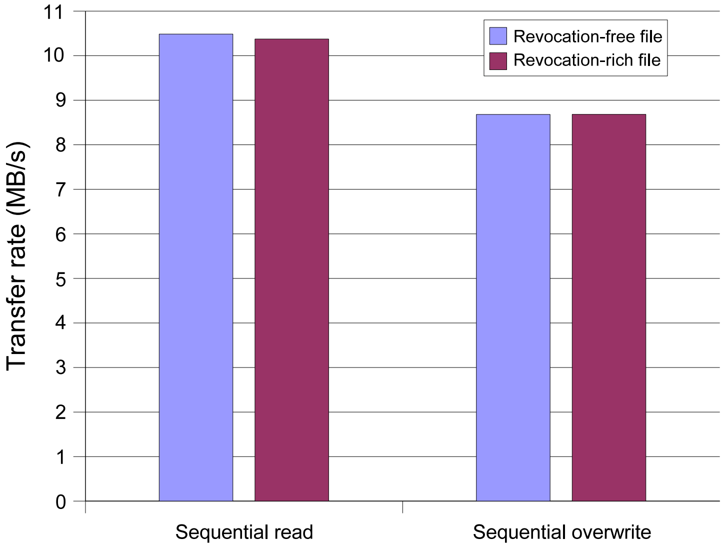

Figure 8: Performance of CRUST for sequentially accessing revocation-free and revocation-rich 100 MB files.

7 Extensions

In this section we present several extensions for the basic CRUST functionality that we did not yet implement.

7.1 Supplemental Confidentiality

Confidentiality of file names and access lists is not provided by the basic CRUST design. These properties can be achieved using methods similar to those included in Plutus [KRS$^+$03]. File name confidentiality can be achieved by encrypting each file's name with a different random key that is generated by its owner. To allow file name decryption by the file's users, the key should be included in their encrypted lockboxes that are stored in the meta-data file.

An access list can also be encrypted with an owner-generated key, provided that the key is included in the lockboxes and that an additional method is used to distinguish the lockboxes, so that each file user can detect his own lockbox. A possible method is to include the target user's ID inside his encrypted lockbox. In this method, a user can identify his lockbox by decrypting its contents (using his common secret key with the owner) and verifying that the decrypted user ID matches his own ID.

7.2 Freshness

Recall that an adversary with full control of the server can mount rollback attacks [MS01]. The following solution for meta-data freshness was proposed in SiRiUS [GSMB03], and can be adopted in CRUST. The owner's file hierarchy is periodically timestamped and the files' meta-data is signed in a tree construction. This solution can be converted to use symmetric-key operations by replacing the public-key signature of the tree's root with a MAC for each user granted access to any file in the hierarchy.

Note that the above solution does not provide freshness guarantees for the file data, since the file writers need the ability to sign the file contents without the owner's intervention. Solutions for data freshness are left for future work.

7.3 File Links

UNIX file systems usually implement a symbolic link as an ordinary file containing a textual reference to the target of the link, and an indicator marking it as a symbolic link. CRUST can implement symbolic links cryptographically, by having a CRUST file whose data contains the textual reference. The file is encrypted and involves access control, as ordinary files in our system. An indicator for symbolic links should be added to the meta-data file, unless support for symbolic links exists in the underlying file system. In any case, the indicator must be specifically verified for integrity and authenticity, as any other part of the meta-data, by including the indicator's value in the existing meta-data MACs.

Hard links allow a file to be referenced by multiple names. The basic CRUST design does not support hard links since only one file name is signed in the meta-data file.

7.4 Crash Recovery

Since CRUST operations are not atomic, a system crash on the client or server side may leave a file in an unreadable state. Moreover, the write caching mechanism increases the probability for the loss of some written data following a crash. Well-known crash recovery solutions that conform to our security model usually maintain non-volatile logs of the actions waiting to be carried out, which can be resumed following a crash. Since the actions performed on the underlying file system are not secret in our design, the logs can be saved in an insecure location, and they may be operated with only a little overhead to the overall performance of the system. Implementation of such mechanisms in CRUST is left for future work.

8 Related Work

The popularity of networked storage systems is constantly growing. However, the nature of these systems exposes them to a vast variety of security threats. A framework for evaluating the security of storage systems is presented in [RKS02]. Additional surveys of security services provided by storage systems are given in [KK05], [Sta04]. These works examine and compare existing systems, architectures and techniques.

CFS [Bla93] is the first widely-known file system that performs file encryption. It is a virtual file system that encrypts each protected directory with a secret key. CFS was primarily designed for securing local file systems, and thus only relatively simple confidentiality issues were addressed. File sharing in CFS is limited; for instance, sharing a protected file with another user involves disclosing the encryption key to that user. CryptFS [ZBS98] and TCFS [CCSP01] are variants of CFS. NCryptfs [WMZ03] provides kernel-level encryption services. It allows convenient file sharing between users located on the same machine. However, support for file sharing in more general situations is limited.

More advanced file systems were designed for securing remote storage systems and for allowing more flexible file sharing between users. Most of these systems trust the file servers and concentrate on protecting against malicious users accessing the network. Such systems usually include mechanisms that ensure proper authentication of the users. SFS [MKKW99] also authenticates the server, so that adversaries are prevented from masquerading as the server. Communication with the server is protected by a session key. However, trusting the server with the data enables attacks where adversaries collude with the server.

SGFS [KSLK06] offers secure, efficient and flexible global file sharing. NASD [GGT97] provides security to network attached storage, where the file server is removed from the data path. Improved performance is gained by the direct interaction of users with the storage device. In this architecture, the storage devices themselves participate in cryptographic protocols. However, they are trusted with the data, and the data is stored in the clear. Thus, the attacks mentioned above are still relevant.

The strictest trust model used by related work avoids trusting the entire storage infrastructure. In such systems, the data is encrypted by writers before it is sent to the server, and decrypted by readers after it is received from the server. Access control is achieved by cryptographic means, and does not rely on the file server's mechanisms. Plutus [KRS+03], SiRiUS [GSMB03] and SNAD [MLFR02] are cryptographic file systems that enable secure file sharing over untrusted servers. However, eliminating the trust in the server comes at the expense of performance. Relaxed assumptions, such as those used by some variants of SNAD, allow improved performance in cases where the server is not completely untrusted.

SiRiUS is designed to be layered on top of any existing file system (but was implemented only over NFS). It implements in-band key distribution, while relying on secure public-key servers or IBE [BF01] master-key servers. Plutus provides efficient random access, file name encryption and lazy revocation, but does not relate to key distribution mechanisms. Plutus, SiRiUS and SNAD provide end-to-end encryption of data, and thus prevent adversaries from accessing files, despite having access to the physical storage device. Plutus and SNAD, unlike SiRiUS, place some trust on the server's access control mechanisms by requiring it to perform checks before committing users' requests.

The designs of Plutus, SiRiUS and SNAD rely on public-key cryptography. SNAD suggests using a symmetric HMAC as an alternative to signatures, as long as the server can be trusted for differentiating readers from writers, thus deviating from the untrusted storage model. Our work follows the research and suggestions of [NSW05] for securing untrusted storage without public-key operations. In our work, we incorporated some of the suggested alternatives into a complete file system design.

There are many other secure file systems whose foci differ from ours. SUNDR [MS01], [LKMS04] addresses important consistency issues of untrusted servers. OceanStore [KBC+00] and FARSITE [ABC+02] are secure large-scale distributed file systems. Availability and fault-tolerance issues are handled in these systems by replicating data and by using cryptographic techniques that allow discarding bad information while preserving useful information.

9 Conclusions

We introduced a new file system layer that allows secure file sharing over untrusted storage systems. Our solution includes: in-band key distribution; flexible control on file access privileges; cryptographic access control, and strong security while assuming that the file server is untrusted and unmodifiable. We achieve all these results without any public-key cryptography. Our approach is especially useful in situations where the users have no control over the underlying file system. It is also useful for sharing files between users that are rarely on-line, because direct communication between the users is not necessary. An additional insight of our approach is that the servers are not required to carry out cryptographic operations and thus the server scalability and the overall performance may be superior.

A significant part of our work focused on performance and usability issues of the system. We have implemented our designed system as a Linux stackable file system and shown that it has very modest overhead, despite the strong security that it provides. We conclude that our approach is convenient to use, performs well in high workloads and supports any underlying file system.

References

[ABC$^+$02] Atul Adya, William J. Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R. Douceur, Jon Howell, Jacob R. Lorch, Marvin Theimer, and Roger Wattenhofer. FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. In *Proc. of OSDI*, 2002.

[BCO06] Michael Backes, Christian Cachin, and Alina Oprea. Secure key-updating for lazy revocation. In *Proc. of ESORICS*, volume 4189 of *Lecture Notes in Computer Science*, pages 327–346. Springer, 2006.

[BF01] Dan Boneh and Matthew K. Franklin. Identity-based encryption from the Weil pairing. In *CRYPTO '01: Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology*, pages 213–229, London, UK, 2001. Springer-Verlag.

[Bla93] Matt Blaze. A cryptographic file system for Unix. In *Proc. of the ACM Conference on Computer and Communications Security*, pages 9–16, 1993.

[Bra96] Tim Bray. The Bonnie home page. Located at `http://www.textuality.com/bonnie`, 1996.

[CCSP01] Giuseppe Cattaneo, Luigi Catuogno, Aniello Del Sorbo, and Pino Persiano. The design and implementation of a transparent cryptographic file system for Unix. In *Proceedings of the FREENIX Track: 2001 USENIX Annual Technical Conference*, pages 199–212, Berkeley, CA, USA, 2001. USENIX Association.

[Dwo01] Morris Dworkin. Recommendation for block cipher modes of operation. Special Publication 800-38A, NIST, 2001.

[FKK06] Kevin Fu, Seny Kamara, and Tadayoshi Kohno. Key regression: Enabling efficient key distribution for secure distributed storage. In *Proc. of NDSS*, 2006.

[GGT97] Howard Gobioff, Garth Gibson, and Doug Tygar. Security for network attached storage devices. Technical Report CMU-CS-97-185, Carnegie Mellon University, October 1997.

[GSMB03] Eu-Jin Goh, Hovav Shacham, Nagendra Modadugu, and Dan Boneh. SiRiUS: Securing remote untrusted storage. In *Proc. of NDSS*. The Internet Society, 2003.

[HP94] John S. Heidemann and Gerald J. Popek. File-system development with stackable layers. *ACM Transactions on Computer Systems*, 12(1):58–89, 1994.

[Jak02] Markus Jakobsson. Fractal hash sequence representation and traversal. In *IEEE International Symposium on Information Theory*, 2002.

[KBC97] H. Krawczyk, M. Bellare, and R. Canetti. HMAC: Keyed-hashing for message authentication. RFC 2104, February 1997.

[KBC$^+$00] John Kubiatowicz, David Bindel, Yan Chen, Steven E. Czerwinski, Patrick R. Eaton, Dennis Geels, Ramakrishna Gummadi, Sean C. Rhea, Hakim Weatherspoon, Westley Weimer, Chris Wells, and Ben Y. Zhao. OceanStore: An architecture for global-scale persistent storage. In *Proc. of ASPLOS*, pages 190–201, 2000.

[KK05] Vishal Kher and Yongdae Kim. Securing distributed storage: challenges, techniques, and systems. In *StorageSS '05: Proceedings of the 2005 ACM workshop on Storage security and survivability*, pages 9–25, New York, NY, USA, 2005. ACM Press.

[Kle86] Steve R. Kleiman. Vnodes: An architecture for multiple file system types in Sun UNIX. In *Proceedings of the USENIX summer conference*, pages 238–247, 1986.

[KRS$^+$03] Mahesh Kallahalla, Erik Riedel, Ram Swaminathan, Qian Wang, and Kevin Fu. Plutus: Scalable secure file sharing on untrusted storage. In *Proc. of FAST*. USENIX, 2003.

[KSLK06] Vishal Kher, Eric Seppanen, Cory Leach, and Yongdae Kim. SGFS: Secure, efficient and policy-based global file sharing. In *Proceedings of the 23rd IEEE / 14th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2006)*, 2006.

[LKMS04] Jinyuan Li, Maxwell N. Krohn, David Mazières, and Dennis Shasha. Secure untrusted data repository (SUNDR). In *Proc. of OSDI*, pages 121–136, 2004.

[LM93] Frank Thomson Leighton and Silvio Micali. Secret-key agreement without public-key cryptography. In *Proc. of CRYPTO*, volume 773 of *Lecture Notes in Computer Science*, pages 456–479. Springer, 1993.

[MKKW99] David Mazières, Michael Kaminsky, M. Frans Kaashoek, and Emmett Witchel. Separating key management from file system security. In *Proceedings of the 17th ACM Symposium on Operating System Principles*, pages 124–139, 1999.

[MLFR02] Ethan L. Miller, Darrell D. E. Long, William E. Freeman, and Benjamin Reed. Strong security for network-attached storage. In *Proc. of FAST*, pages 1–13, 2002.

[MS01] David Mazières and Dennis Shasha. Don't trust your file server. In *Proc. of HotOS*, pages 113–118. IEEE Computer Society, 2001.

[NIS01] NIST. Advanced encryption standard. Federal Information Processing Standards, FIPS PUB 197, 2001.

[NIS04] NIST. Secure hash standard. Federal Information Processing Standards, FIPS PUB 180-2, 2004.

[NSW05] Dalit Naor, Amir Shenhav, and Avishai Wool. Toward securing untrusted storage without public-key operations. In *StorageSS '05: Proceedings of the 2005 ACM workshop on Storage security and survivability*, pages 51–56, New York, NY, USA, 2005. ACM Press.

[Ope] The OpenSSL project. Located at `http://www.openssl.org`.

[RKS02] Erik Riedel, Mahesh Kallahalla, and Ram Swaminathan. A framework for evaluating storage system security. In *Proc. of FAST*, pages 15–30, 2002.

[Rub00] Aviel D. Rubin. Kerberos versus the Leighton-Micali protocol. *Dr. Dobb's Journal of Software Tools*, 25(11):21–22, 24, 26, November 2000.

[SNS88] Jennifer G. Steiner, B. Clifford Neuman, and Jeffrey I. Schiller. Kerberos: An authentication service for open network systems. In *Proc. of the USENIX Winter Conference*, pages 191–202, 1988.

[Sta04] Paul Stanton. Securing data in storage: A review of current research. *CoRR*, cs.OS/0409034, 2004.

[Sze] Miklos Szeredi. Filesystem in userspace. Located at `http://fuse.sourceforge.net`.

[WMZ03] Charles P. Wright, Michael C. Martino, and Erez Zadok. NCryptfs: A secure and convenient cryptographic file system. In *Proceedings of the Annual USENIX Technical Conference*, pages 197–210, 2003.

[ZBS98] E. Zadok, I. Badulescu, and A. Shender. Cryptfs: A stackable vnode level encryption file system. Technical Report CUCS-021-98. Computer Science Department, Columbia University, 1998.

A The Hash Matrix

A.1 Overview

In this appendix, we present an efficient key regression scheme using only symmetric-key cryptography. In comparison, the *key rotation* mechanism presented in Plutus [KRS$^+$03] (which can be extended to a key regression mechanism [FKK06]) is based on public-key cryptography; it uses a dedicated private key to generate a new encryption key from the current one. A straight-forward symmetric-key alternative mentioned in Plutus, is the *hash chain*. It is a more efficient method, due to the use of symmetric-key cryptography, but is limited to supporting a predefined maximal amount of revocations.

The hash chain method uses a one-way hash function for deriving an older encryption key from the current one. In order for the owner to compute the first key, the entire chain of n keys must be calculated upon initialization. The initialization begins by generating K_{n-1}, a random master key. The other keys are then computed according to the formula $K_i = h(K_{i+1})$, where $h(\cdot)$ is a cryptographic hash function and K_i is the file encryption key to be used after the i-th revocation. The number i is associated with every block, and represents the block's epoch. The last key of the chain, K_0, is the first one to be used.

In the hash chain method, when a revocation occurs, the owner calculates the next key and provides it to the users. The most recent key can be used by any user to derive all earlier keys, by applying the hash function for an appropriate number of times. The owner keeps private information that allows calculating the new keys after revocations occur. He can either store the entire chain or as little as the master key alone. The amount of information stored affects the key derivation time. More specifically, there is a trade-off between the two, where the memory-times-computational complexity is $O(n)$ per key derivation. Jakobsson [Jak02] presented an improved technique for efficiently deriving the keys without storing the entire chain; however, all hash chain-based techniques suffer from an inefficient initialization process, because it requires calculating the entire chain. Since the length of the chain limits the number of lazy revocations, which is desired to be as large as possible, the hash chain technique is inefficient for our purposes.

We present a new key regression mechanism called the *hash matrix*. It enables both initialization and key derivation to be done efficiently, in aspects of memory and computational complexity. The hash matrix can be seen as a generalization of the hash chain, where multiple keys are derived from every single key, by using multiple one-way functions. This multiplicity of one-way functions shortens the computation paths dramatically, in comparison to a hash chain supporting the same number of total available keys.

Our mechanism is reminiscent of the binary-tree key-updating scheme (TreeKU) of [BCO06], where the keys are derived in a tree-like manner. The general goals of the two mechanisms are similar, though the mechanism to be used in our system has a stricter space limit. In TreeKU, a key can be derived into two other keys by using it as a seed for a pseudo-random generator, and taking two parts of its output as the derived keys. The keys are conceptually arranged in a complete binary tree of depth d, where a master key is used as the tree's root. The edges indicate which part of the pseudo-random generator's output is taken for deriving each key from its parent. The maximal number of pseudo-random function applications required in order to compute any key in the tree is d. The total number of available keys is $n = 2^{d+1} - 1$. TreeKU requires storing a table of up to $d+1$ keys for maintaining the key-updating state; each key can be computed by applying the pseudo-random function at most d times. Both space and time bounds are logarithmic in n, but we seek for a technique that allows storing less than $\log_2 n$ keys, without harming the computational complexity significantly. The scheme we suggest in the next section achieves these goals.

A.2 Scheme Details

In our technique, the keys are arranged in a d-dimensional matrix of uniform length m. The total amount of keys in our key regression system is $n = m^d$. Each key K_i is identified by a base-m number i indicating its position in the matrix (i can also be seen as the epoch identifier; recall Section 3.6). We use d different one-way functions $f_k(\cdot)$ for $k = 0, 1, \ldots, d-1$ in order to define the keys. The master key K_{n-1} is generated randomly when the key regression system is initialized. Each key K_i can be derived from the master key by a unique *computation sequence*, in which each step includes an application of a one-way function on the previous step's result. Let $b_k(i)$ for $k = 0, 1, \ldots, d-1$ denote the k-th base-m digit of i, where $k = 0$ refers to the least significant digit. The computation sequence of K_i from K_{n-1} consists of several consecutive applications of each one-way function, starting with $f_{d-1}(\cdot)$ and ending with $f_0(\cdot)$. For each k, $f_k(\cdot)$ is applied $t_k(i)$ consecutive times during the sequence, where:

$$t_k(i) = (m-1) - b_k(i).$$

For example, for $m = d = 3$, $K_{012_3} = f_1\left(f_2\left(f_2\left(K_{n-1}\right)\right)\right)$ and $K_{122_3} = f_2\left(K_{n-1}\right)$. Mapping the keys in a Cartesian coordinate system, as demonstrated in Figure 9, results in the so-called hash matrix. Note that some keys have a computation sequence that is a prefix of other keys' computation sequences. This allows to derive some keys from others.

After the i-th revocation, our scheme provides the users with a state, S_i, which is a structure allowing computation of the keys $K_0, K_1, \ldots, K_i$. The state S_i consists of a small group of keys, including K_i and

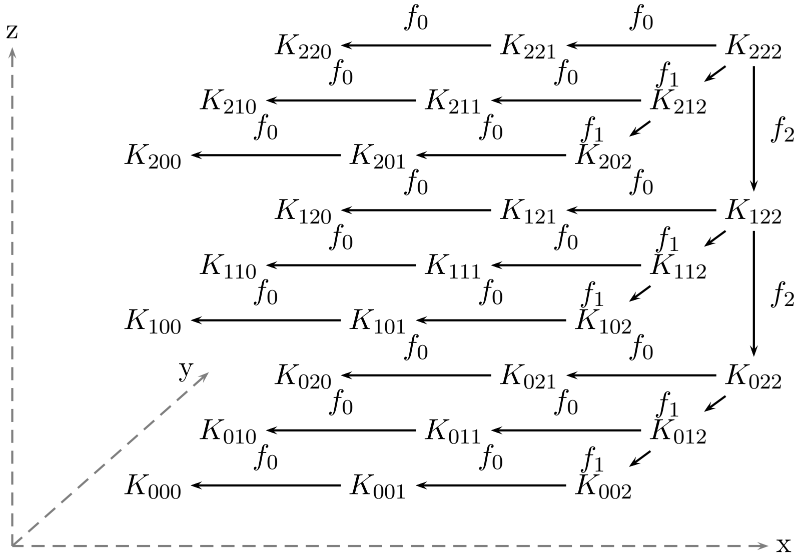

Figure 9: The hash matrix for $m = 3$, $d = 3$. Arrows indicate the key computation sequences.

some earlier keys, that enable the computation of $K_0, K_1, \ldots, K_i$. More specifically, the state S_i consists of d keys:

$$S_i = \langle K_i, K_{\mathrm{sub}(i,1)}, K_{\mathrm{sub}(i,2)}, \ldots, K_{\mathrm{sub}(i,d-1)} \rangle,$$

where $\mathrm{sub}(i,k)$ is the base-m number obtained by subtracting one from the k-th digit of i, and maximizing all less significant digits. If the digit to be subtracted is zero, $\mathrm{sub}(i,k)$ is defined as zero, and $K_{\mathrm{sub}(i,k)}$ is not necessary for the state. In any case, $\mathrm{sub}(i,k) \le i$.

The hash matrix mechanism includes three algorithms: an *initialization* algorithm, a *state derivation* algorithm and a *key extraction* algorithm. The initialization routine constructs S_{n-1} by simply choosing the master key K_{n-1} randomly, and computing $K_{\mathrm{sub}(n-1,k)} = f_k(K_{n-1})$ for $k = 1, 2, \ldots, d-1$. The state derivation and key extraction algorithms compute S_i and K_i accordingly, given a more recent state, S_j, where $j >= i$. For $j = i$, the algorithms are trivial, since $S_i = S_j$, and K_i is included in S_j. We now explain the algorithms for the other case, $j > i$.

First, we show the key extraction algorithm, which is later extended to the state derivation algorithm. Let k denote the position of the most significant digit differing i and j. Obviously, $b_k(i) < b_k(j)$. If $k = 0$ then K_i and K_j differ only in the number of applications of $f_0(\cdot)$, hence $K_i = f_0^{j-i}(K_j)$ (i.e., $j - i$ applications of $f_0(\cdot)$). Otherwise, observe $K_{\mathrm{sub}(j,k)}$, which is included in the given state S_j. We claim that this key's computation sequence is a prefix of K_i's computation sequence.

This follows from the fact that the $d - (k+1)$ most significant digits of i, j and $\mathrm{sub}(j,k)$ are identical, the k-th digit of i is smaller or equal to that of $\mathrm{sub}(j,k)$, and each of the less significant digits of $\mathrm{sub}(j,k)$ is maximal (i.e., equals to $m-1$). The maximality of the less significant digits means that $f_{k-1}(\cdot), f_{k-2}(\cdot), \ldots, f_0(\cdot)$ were not applied at all for computing $K_{\mathrm{sub}(j,k)}$. Therefore, what is left to do for completing the computation sequence of $K_{\mathrm{sub}(j,k)}$ to that of K_i, is applying $f_k(\cdot), f_{k-1}(\cdot), \ldots, f_0(\cdot)$ for the appropriate number of times; first, $f_k(\cdot)$ should be applied $b_k(\mathrm{sub}(j,k)) - b_k(i)$ times; then, each of the functions $f_{k-1}(\cdot), f_{k-2}(\cdot), \ldots, f_0(\cdot)$ should be activated $t_{k-1}(i), t_{k-2}(i), \ldots, t_0(i)$ times accordingly.

The key extraction algorithm is extended to the state derivation algorithm by computing all the keys contained in S_i, as follows. For $\hat{k} = k+1, k+2, \ldots, d-1$, which represent the most significant digits that are equal in i and j, $\mathrm{sub}(i, \hat{k}) = \mathrm{sub}(j, \hat{k})$, so $K_{\mathrm{sub}(i,\hat{k})} = K_{\mathrm{sub}(j,\hat{k})}$. For $\hat{k} = 0, 1, \ldots, k$, we relate to a specific intermediate key computed during the original algorithm, that was the result of applying $f_{\hat{k}}(\cdot)$ for the appropriate number of times. We apply $f_{\hat{k}}(\cdot)$ on that key one additional time. This additional transformation is equivalent to the subtracted digit in $\mathrm{sub}(i, \hat{k})$, and thus calculates $K_{\mathrm{sub}(i,\hat{k})}$. Note that each of these calculations should be avoided if $b_{\hat{k}}(i) = 0$.

Finally, we remark that an additional hash function should be applied on each of the extracted keys K_i in order to derive the actual encryption keys to be used in our system: $K_i^{\text{Enc key}} = h(K_i)$. This final step of the key extraction algorithm ensures the pseudo-randomness of the encryption keys, which is a required property of key regression schemes [FKK06].

We prove the security of the hash matrix by proving that it is computationally infeasible to determine K_i given $K_0, K_1, \ldots, K_{i-1}$ (note that this also implies that it is infeasible to determine K_i given $S_0, S_1, \ldots, S_{i-1}$ as well); an informal proof follows. For every $j < i$, let k be the maximal number such that $b_k(i) \neq b_k(j)$. Obviously, $b_k(i) > b_k(j)$ and all the more significant digits of i and j are equal. Therefore, $t_k(i) < t_k(j)$. These facts imply that the unique calculation sequences of K_i and K_j share a common prefix. At the first step where the sequences differ, the calculation sequence of K_j contains an additional application of $f_k(\cdot)$ that does not appear in K_i's. This one-way transformation makes the computation of K_i given K_j infeasible. Since such a difference exists for every $j < i$, it is infeasible to determine K_i given $K_0, K_1, \ldots, K_{i-1}$.

A.3 Complexity Analysis

The maximal space needed for storing a hash matrix state is equivalent to storing d keys. The time for computing keys and states is dominated by the one-way transformations. Initialization requires $d-1$ transformations. Extracting K_i involves applying the d one-way functions, no more than $m-1$ times each. Thus, it involves at most $d(m-1)$ one-way transformations. Deriving S_i involves at most $d-1$ additional one-way transformations, but is also bounded by $d(m-1)$ total one-way transformations.

For a given number of keys n and some small constant p:

$$n = 2^{\log_2 n} = 2^{\frac{p}{p}\log_2 n} = (2^p)^{\frac{1}{p}\log_2 n} = m^d,$$

so a possible choice of parameters for the hash matrix is $m = 2^p$ and $d = \frac{1}{p}\log_2 n$. This choice of parameters implies that the maximal size of a state is equivalent to $\frac{1}{p}\log_2 n$ keys, and deriving a state or a key involves no more than $\frac{1}{p}(2^p - 1)\log_2 n$ one-way transformations. For example, choosing $p = 4$ yields space consumption of at most $\frac{1}{4}\log_2 n$ keys, while deriving a state or a key involves less than $4\log_2 n$ one-way transformations. These results indeed achieve the desired goal of having space consumption of less than $\log_2 n$ keys, without significantly harming the logarithmic computational complexity achieved in previously known constructions.

In CRUST we used $m = 16$, so that each base-m digit can be stored in a 4-bit nibble. We used $d = 7$, giving a maximum of $n = m^d = 2^{28}$ keys and $n-1$ revocations without re-encryption. We used HMAC based on SHA-1 as the keyed one-way function. The state S_i contains $d = 7$ hash values (20 bytes each), totaling as little as 140 bytes. Initialization, as well as key or state derivation, requires at most

$d(m-1) = 105$ HMAC operations. The 20-byte values extracted from the key regression mechanism are truncated to 16 bytes in order to be used as AES-128 encryption keys.

B Protocols

B.1 Creating a File

CRUST creates a file anywhere in the user's directory hierarchy by taking the following steps.

1. Generate a random master key regression state for the file, as well as a file master MAC key.
2. Create the meta-data file and initialize the file's access lists.
3. Set the current epoch to zero.
4. Derive the current key regression state and encryption key for the current epoch from the master key regression state.
5. Save the owner's lockbox, encrypted with his encryption key, $K_i^{\text{Self enc}}$, and a MAC of the owner's meta-data using his MAC key, $K_i^{\text{Self MAC}}$.
6. Create the data file and encrypt the file data using the derived encryption key from step 4. Update the hash tree on-the-fly, while encrypting. Set the block epochs to zero.
7. MAC the hash tree root using the file's master MAC key (from step 1).

B.2 Sharing a File

CRUST takes the following steps when a file owner, Alice, wishes to share the file with another user, Bob.

1. Alice reads her lockbox, verifies it using her MAC key, $K_{\text{Alice}}^{\text{Self MAC}}$, and decrypts it using her encryption key, $K_{\text{Alice}}^{\text{Self enc}}$. The hash tree root is verified using the file's master MAC key obtained from the lockbox.
2. Alice obtains Bob's user ID. If the cached version of the user table is not sufficient, a fresh version of the table is read and verified using Alice's user table MAC key, $K_{\text{Alice}}^{\text{User table MAC}}$.
3. Alice adds Bob's ID to the file's access list (as a reader or a writer) and re-signs the list.
4. Alice derives her common encryption and MAC keys for communicating with Bob, using the Leighton-Micali scheme.
5. Alice prepares Bob's encrypted and authenticated lockbox. It contains the current key regression state and Bob's file reader MAC key, $K_{\text{Bob}}^{\text{File MAC}}$, if Bob is granted read-only access. If Bob is given write permission, the file's master MAC key is given instead of the reader MAC key.
6. Alice calculates a MAC of the meta-data for Bob using her common MAC key with him. If Bob is granted read-only access, Alice also calculates a MAC of the hash tree root using $K_{\text{Bob}}^{\text{File MAC}}$.

B.3 Writing to a File

The following procedure is performed by CRUST when Bob writes to a file owned by Alice (it is located under Alice's directory). Note that a similar procedure is carried out if the file is owned by Bob instead.

1. Bob obtains Alice's user ID. If the cached version of the user table is not sufficient, a fresh version of the table is read and verified using Bob's user table MAC key. Bob also obtains his common encryption and MAC keys with Alice according to the Leighton-Micali scheme.

2. Bob locates and reads his writer lockbox. The lockbox is decrypted and verified using his common encryption and MAC keys with Alice. The hash tree root is also verified using the file's master MAC key.

3. Bob derives the current encryption key from the current key regression state.

4. For every data block to be written:

   (a) If the data block is only partially updated, the stored block is firstly read:

      i. The encrypted block and its epoch are read and verified using the hash tree.

      ii. The block's encryption key is derived from the current key regression state, based on the block's epoch identifier.

      iii. The decrypted block is updated with the written data.

   (b) The block is encrypted with the current encryption key.

   (c) The block's epoch is updated to the current epoch.

   (d) The hash tree is updated.

5. Bob updates the MACs of the hash tree root with the file master MAC key and with each reader's MAC key.

B.4 Reading a File

The following procedure is performed by CRUST when Bob reads a file owned by Alice.

1. This step is identical to the first step for file writing.

2. Bob locates and reads his lockbox. The lockbox is decrypted and verified using his common encryption and MAC keys with Alice. The hash tree root is also verified using either the file's master MAC key or Bob's reader MAC key, according to Bob's permissions.

3. For every data block to be read, the encrypted block and its epoch are read and verified using the hash tree. Each block is decrypted using an encryption key derived from the current key regression state, based on the block's epoch identifier.

B.5 Revoking a File Sharing

CRUST takes the following steps when Alice wishes to revoke the sharing of a file owned by her with another user, Bob.

1. This step is identical to the first step for file sharing.
2. Alice obtains Bob's user ID.
3. Alice removes Bob's ID from the file's relevant access list and re-signs it.
4. Alice increases the current epoch and derives the new key regression state. If Bob had write permissions, a new random file master MAC key is generated. All the lockboxes are updated with these modifications, and re-signed.
5. If a new file master MAC key was generated, the MACs of the hash tree root are recalculated.