For reliable storage services, we need a way not only to monitor the state of stored data but also to recover the original data when some data loss is discovered. To solve the problem, a novel technique called HAIL has been proposed. Unfortunately, HAIL cannot support dynamic data which is changed according to users’ modification queries. There are many applications where dynamic data are used. So, we need a way to support dynamic data in cloud services to use cloud storage system for various applications. In this paper, we propose a new technique that can support the use of dynamic data in cloud storage systems. For dynamic data update, we design a new data chunk generation strategy which guarantee efficient data insertion, deletion, and modification. Our technique requires $O(1)$ operations for each data update when existing techniques require $O(n)$ operations where $n$ is the size of data.

**key words:** cloud storage, dynamic data, proof of retrievability

1. **Introduction**

In these days, the use of remote storage for maintaining valuable information is increasing due to the rapid progress of network bandwidth. Along with the growth of the use of remote storage, clients’ concern regarding the security of storage services is also increasing [2]. In cloud storage services, clients entrust their data to remote storage servers and delete entrusted data in their local storage. Since clients do not possess the data, the integrity of stored data is very important for reliable cloud storage services.

For cloud storage services, we need to consider two scenarios that can break the security of the services.

**Inside Adversary:** A malicious service provider could be an adversary. Though it is assumed that service providers are not actively breaking the integrity of clients’ data, sometimes, some service providers could lose clients’ data due to their carelessness or the use of cheap storage. The malicious servers are also called as an honest-but-curious adversary. In general, the security against inside adversaries can be ignored only if clients can detect the service provider’s fault.

**Outside Adversary:** Differently from an inside adversary who is a legitimate service provider, outside adversaries can mount any attack since they have nothing to lose even if their attacks are detected by any entity including clients. An outside adversary may try to take clients’ data stored in service provider’s storage or interrupt valid connection between the service provider and clients. The first mentioned attack can be relatively easily countered since we can use an encryption scheme to prevent the adversary from getting visible information. When the goal of an outside adversary is to compromise or delete stored data, simply encrypting stored data is not enough, instead, we need a way to monitor the current state of the stored data to ensure that outsourced data are intact. However, the latter cannot be easily countered since the adversary can easily mount the attack without losing any asset of the adversary. Is the reason why such an attack is harmful and hard to encounter.

In clients’ viewpoint, two techniques are required to guarantee the fact that the stored data can be recovered whenever the owner of the data want to retrieve.

- Verifiable proof of integrity for stored data
- Data recovery when a part of data is damaged

To check the integrity of data stored in remote cloud servers, provable data possession (PDP) and proof of retrievability (PoR) have been proposed to permit users to audit the current state of data stored in remote storage [1], [4]–[7], [9]–[11]. Note that PDPs and PoRs are very similar in terms of the functionality of integrity check. The main difference between them is the retrievability of stored data. In PDPs, only the integrity check is supported, but PoRs support original data recovery when some corrupted data blocks are discovered. However, in this paper, we consider PoRs to be unable to provide the retrievability since an outside adversary can spoil the data so that the original data cannot be recovered. In other words, we say that PoRs also cannot support the retrievability against outside adversaries.

If a storage service meets the first requirement, we can guarantee the security against an inside adversary since clients can test if the service provider maintains the original data. However, we still cannot cover the security against an outside adversary. If an outside adversary breaks the storage system of a cloud service provider and maliciously modifies the original file, the first requirement doesn’t work for the threat. In this case, the only thing the client can do is to identify the attack by verifying the current state of the stored data. To cover the case, we need the second requirement which permits clients to recover the original data from damaged data. To support the functionality, a high-availability
and integrity layer (HAIL) for cloud storage has been proposed [3], which permits users to recover the original data by the help of storage servers.

Until now, a number of schemes have been proposed the security against inside adversaries and outside adversaries. Most of them are designed for static data where the data are not modified after uploading. However, we need to use dynamic data in many cloud-based applications. For example, if we use a cloud system for smart work, we may want to modify document files stored in cloud storage. Due to the complexity of algorithm and protocol design for dynamic data, relatively few schemes have been proposed to cover dynamic data in cloud storage environment [4], [6], [10], [11]. Among existing techniques for the security against outside adversaries [3], [8], only a few schemes support dynamic data [8].

In this paper, we propose a new architecture for maintaining dynamic data in distributed storage servers, which provides (1) verifiable proof of integrity and (2) original data recovery if limited number of storage servers are damaged. For the verifiable proof of integrity, we devise an efficient tag generation function that returns a short tag for a number of data blocks. The tag is an accumulator which is computed as a linear sum of the data blocks. Since the linearity is suitable for dynamic updates, our technique supports efficient verifiable proof of integrity suitable for dynamic data. For the retrievability, we use error-correcting code as in many existing techniques. Differently from existing techniques where a chunk is composed of consecutive data blocks, consecutive data blocks are distributed in our construction.

2. Preliminaries

In this section, we will describe the adversarial model and the system model of our technique. As we explained in the introduction, we consider two types of adversaries, inside adversaries and outside adversaries, which are defined as following:

- **Security against Inside Adversary**: An inside adversary’s malicious behavior can be detected by clients so that the adversary cannot try to mount any attack that causes the loss of stored data.
- **Security against Outside Adversary**: An outside adversary’s attack may not be prevent, but the damage should be recoverable.

Recall that the main goal of our work is to support cloud storage services to resist against inside and outside adversaries.

As seen in Fig. 1, in the proposed system, we consider three types of storage servers, servers for storing original data, servers to maintain some additional information for data recovery, and a server to manage dynamically changed data status. Let \( n \) be the total number of servers in the proposed cloud storage systems. Let \( n' \) be the number of servers for original data and \( n'' \) be the number of additional servers for data recovery. Since we need one server for managing dynamic data status, \( n = n' + n'' + 1 \). Note that the number of additional servers for data recovery is determined by the level of error tolerance against outside adversaries who break a number of servers so that clients cannot recover their original data. In our setting, clients can recover their data until \( n''/2 \) servers are broken by outside adversaries. For this reason, we assume that \( n'' = \) an even number. Differently from some existing works, CSP accesses to \( n' + 1 \) storage servers unless a part of their data stored in \( n' \) storage servers for original data are not harmed.

3. Proposed Construction

3.1 Initial Data Encoding

The goal of the initial step is to generate chunks to distribute the original data to multiple servers. Let \( b \) be the size of a block. We assume that the bit-size of data is always dividable by \( b \). Let \( F \) be the input which is composed of \( \ell \) blocks. We assume \( n \) storage servers where one server is in charge of maintaining information for data dynamic, \( n' \) servers store original data, and \( n'' \) servers maintain additional information for supporting original data recovery. To resists against the corruption of \( cor \) servers, we will set \( n'' = 2cor \). Let \( params \) be the set of input parameters of encoding function. Then, the input file \( F \) is divided into \( \ell \) blocks of size \( b \). We expect that the original file is distributed to \( n' \) servers and they have the same sized chunk. Hence, we firstly test if \( \ell \) is a multiple of \( n' \) or not. If so, we use the input file as is, and set \( M = F \). Otherwise, we append some redundant blocks at the end of the file \( F \) and set it \( M \). Then the number of blocks in \( M \) is a multiple of \( n' \). Here, the same value is assigned to all redundant blocks. In Algorithm 1, we initialize redundant blocks as NULL, but any value can be used for them. \( m_i \) means the \( i \)-th bit of \( M \).

From step 7 to 9, original data blocks are divided into \( n' \) chunks \( CH_i \) for \( i = 1, \ldots, n' \). In our construction, \( \alpha \) blocks are assigned to each chunk and the difference of any two adjacent blocks’ indexes is \( n' \). Then we initialize the chunk \( CH_0 \) which contains some redundant information to support dynamic data and the proof of retrievability. Detailed explanation for the initialization function \( \text{Init}(\cdot) \) will be given in Sect. 3.1.1. From step 16 to 22,
Algorithm 1 Initial Data Encoding

Require: \( F \): input file composed of \( n \) blocks of size \( b \)-bits, \( n' \): total number of servers in the cloud storage systems, \( n'' \): number of servers for original data, \( c_{ori} = n''/2 \): number of additional servers for data recovery, \( corr = n''/(2n' - n) \): robustness against corruption of storage servers, \( params \): input parameters of encoding functions for server corruptions.

Ensure: \( n \) chunks to be stored in distributed cloud storage servers

1: /* Input File Segmentation */
2: if \( t \) is a multiple of \( n' \) then
3: \( a = \lceil t/n' \rceil \) and \( M = F \)
4: else
5: \( a = \lceil t/n' \rceil \) and \( M = F \)
6: end if
7: for \( i \) from 1 to \( n'' \) do
8: \( CH_i = m_i || m_{i+1} || \ldots || m_{i+(n-1)n'} \)
9: end for
10: /* Initialize \( CH_0 \) */
11: for \( i \) from 1 to \( a \) do
12: \( run \text{ Init}(stat) \)
13: end for
14: \( CH_0 = \text{stat} || \text{stat} || \ldots || \text{stat}_a \)
15: /* Encoding */
16: for \( i \) from 1 to \( a \) do
17: \( TM_i = m_{ia} || m_{i+a} || \ldots || m_{i+(n-1)n'} \)
18: \( \{e_{i,1}, e_{i,2}, \ldots, e_{i,n'}\} = \text{Encode}(params, TM_i) \)
19: end for
20: for \( i \) from 1 to \( n'' \) do
21: \( CH_{i'} = e_{i,1} || e_{i,2} || \ldots || e_{i,n'} \)
22: end for
23: return Encoded chunks \( \{CH_0, \ldots, CH_{n''-1}\} \)

\( n'' \) chunks are generated by applying an encoding function for server corruption resistance. Here, we use an encoding function which takes \( i \)-th blocks of \( n' \) + 1 chunks \( \{CH_0, CH_1, \ldots, CH_a\} \) as an input and returns \( n'' \) blocks as an output. \( n'' \) blocks of the output are assigned to the \( i \)-th blocks of \( n'' \) chunks \( \{CH_{i'+1}, \ldots, CH_{i'+(n-1)}\} \). Then, Algorithm 1 returns \( \{CH_0, \ldots, CH_{n''-1}\} \) for given input values.

3.1.1 State Initialization

We want to note that the state initialization function is assumed to be executed by the cloud service provider instead of the owner of the data. We assume the above scenario for both the efficiency in client-side cost and the simplicity of explanation of the initial data encoding algorithm. However, if the owner of the data does not fully trust CSP, the client can compute the state information and send it to CSP. The main difference between two approaches is the entity that updates state information for each data update operation. The entity who generates the information will update it too. From now, we will describe proposed algorithms under the assumption that the cloud service provider generates and updates state information. We want to emphasize that the role can be easily shifted to the owner of the data.

We use state information to manage the status of data blocks stored in the same row in \( a \) storage servers for original data. For example, we store \( \alpha \)-bit \( stat_i \), \( \{m_1, m_2, \ldots, m_{(r+1)}\} \) as the \( i \)-th status information. Different from other servers, the server for status information should maintain a little bit longer data since it supports dynamic data management and reliable retrievability proof at the same time. For two functionalities, four values are required including \( s_i, n_i, r_i \), and \( au_i \). Among them, two values \( s_i \) and \( n_i \) are used for managing dynamic updates. Note that, along with dynamic updates, some data blocks could be deleted or inserted. In our construction, we cover such dynamic modifications by managing meaningful blocks in a row, which means that some meaningless blocks may be included in some rows. The role of \( (s_i, n_i) \) is to manage the status. From now, we explain the role of each value for our construction. The first \( 1 \)-bit information \( s_i \) indicates whether there is a meaningless block or not. If all blocks are valid, we set \( s_i = 1 \), otherwise \( s_i = 0 \). Using the second \( \nu \)-bits information \( n_i \), we count the number of invalid blocks. So the second value influences the number of storage servers for original data. We can support \( 2^\nu + 1 \) storage servers since \( \nu \)-bits information covers so far as to the case where only one block is valid among \( 2^\nu + 1 \) blocks. The role of \( (r_i, au_i) \) is to support retrievability proof. Though it is possible to support more useful functions such as public verifiability, in this section, we describe secret-key based simple construction which does not support the functionality due to the simplicity of explanation. In the next section, we will give a modification which can support the functionality. In secret-key based approach, the owner of the file randomly chooses a secret key \( sk \) with \( \alpha \) secret values \( a_1, \ldots, a_{\alpha} \in GF(2^\alpha) \) to generate authentication code. Then the authentication code for \( \alpha \) data blocks \( \{m_1, m_2, \ldots, m_{(r+1)}\} \) is computed as

\[
au_i = prng(sk, r_i) + a_1m_1 + \ldots + a_{\alpha}m_{(r+1)} \mod f(x)
\]

where \( prng \) is a pseudo-random number generator which guarantees collision resistance and \( f(x) \) is an irreducible polynomial defining \( GF(2^\alpha) \). In practice, we can use \( i = r_i \) instead of a randomly chosen value for the simplicity and efficiency of the initialization procedure. Since the goal of \( r_i \) in computing \( prng(sk, r_i) \) is to guarantee the collision resistance, the use of a sequential number is also acceptable. For the approach, the data owner should maintain the recently used counter.

When we need to use outer code to recover the original data from damaged data chunks stored in \( n \) distributed storage servers, two values \( r_i \) and \( au_i \) are not needed since they are useful only for retrievability proof. Hence, the following value can be used as \( d \)-bits input for outer code: \( s_i || n_i || 0^{d-1}(1+\nu) \) where \( 0^{d-1}(1+\nu) \) means \( d -(1+\nu) \) zeros in bit. If we use a sequential number for \( prng \), we can concatenate the counter at the end of \( n_i \) and fill zeros so that the size to be \( d \)-bits. As a result, we can guarantee that the size of data blocks in all storage servers for outer code is \( b \)-bits.

3.1.2 Encoding

We use an encoding function \( \text{Encode}() \) to resist the corruption of storage servers. Since we deal with dynamic data which is updated frequently, it is desirable to store original
data as it is. In general, encoding functions such as error-correcting codes that support data recovery transform the original data to encoded data so that the original data should be re-constructed from encoded data. Some error-correcting codes permit users to store the original data without transformation of the original form. In the literature, such schemes are called systematic codes. So, in our construction we use a systematic code, especially a systematic Reed-Solomon code will be used as the encoding function.

3.2 Update

Before discussing update procedures, we want to emphasize that we will focus on data blocks instead of dealing with all data blocks including the blocks in additional servers for data recovery. Since we can easily update additional data blocks for recovery by running the encoding function, we will explain the update method only using original data blocks and status information blocks for the simplicity of the explanation.

In all data update procedures, the owner of a file makes a query to CSP so that the service provider to control data blocks stored in storage servers according to the user’s request. To support a fully dynamic data update, we need to cover three queries, existing data modification query, existing data deletion query, and new data insertion query. Form now, we describe each query. Note that a client may make a request to update his data when he knows the full or a part of his data. Moreover, to utilize the stored data, status information is also required since not all data blocks are valid in our construction. So, we can assume that the client knows the current status of his data when he makes a query.

For each update query, the client makes a request packet $Q_{m}(m^*, i_1, i_2)$, $Q_{d}(i_1, i_2)$, or $Q_{i}(m^*, i_1, i_2)$ for modification, deletion, or insertion query. Note that each update query means the following status change. For $Q_{m}(m^*, i_1, i_2)$, a new message $m^*$ will be stored instead of $i_2$-th block in $i_1$-th storage server. For $Q_{d}(i_1, i_2)$, a new message $m^*$ will be inserted at the end of $i_2$-th block in $i_1$-th storage server. For given update queries, the CSP makes a number of queries to storage servers. Three types of queries made for storage servers are defined as follows. Let $Q_{m}(m, i)$, $Q_{d}(i)$, and $Q_{i}(m, i)$ be modification, deletion, and insertion queries made for storage servers. For given $Q_{m}(m, i)$, a storage server may modify $i$-th block by given $m$. For $Q_{d}(i)$, a storage server simply deletes $i$-th block. Finally, for given query $Q_{i}(m, i)$, a storage server inserts a new data block $m$ after $i$-th block.

From now, we will describe each update query based on the above defined query messages.

3.2.1 Modify Query

A modification query is easy to treat since the request packet given from the data owner is almost identical with the request packets will be sent to storage servers. To modify $i_2$-th block in $i_1$-th server, the data owner may compute the new state information $stat'_{i_2}$ for $i_2$-th blocks and generate an update query $Q_{m}(m^*, i_1, i_2)$. Then the client sends them to the CSP. For given query, the service provider generates $Q_{m}(m^*, i_2)$ and gives it to $i_1$-th storage server. Due to the modification of a block, the state of $i_2$-th blocks is also changed, and thus the service provider generates an update query $Q_{m}(stat'_{i_2}, i_2)$, and gives it to $Srvi_2$. Then $Srvi_0$ and $Srvi_1$ may update its own storage according to given update requests. $Srvi_0$ may stores $stat'_{i_2}$ instead of $stat_{i_2}$, and $Srvi_1$ may modify $i_2$-th block with given message $m^*$.

3.2.2 Delete Query

The goal of a deletion query is to remove a data block from stored data blocks. Differently from the simple purpose of the query, we have to carefully deal with deletion queries since possible scenarios to consider are not simple due to the existence of NULL data blocks. So, we will explain the delete query processing according to the possible scenarios. The data owner generates an update query $Q_{d}(i_1, i_2)$ and a new status information $stat'_{i_2}$ according to the status change in storage. Then, the service provider processes the query as in Algorithm 3. At first, CPS checks the status of $i_2$-th blocks by reading the first bit of $stat_{i_2}$, the current state information. If the most significant bit of $stat_{i_2}$ is 1, we can see that there is no invalid block in $i_2$-th blocks. In this case, CPS generates a set of query messages to shift $n' - i_1$ data blocks to other storage servers who are indexed by 1 smaller indexes. For example, the message block $m_{i_1+1,n'-(i_1-1)}$ which was stored in $Srvi_{i+1}$ will be stored in $Srvi_1$. Similarly, $m_{i_2+2,n'-(i_2-1)}$ will be stored in $Srvi_{i+1}$. Lastly, we will store NULL in the last storage server $Srvi_{n'}$. From step 2 to 8 of Algorithm 3, the first case will be covered. If $MSB(stat_{i_1}) = 0$ and $i_1 + MSB(stat_{i_2} < 1, v) = n'$, we can see that the position of the data block to be deleted is located at the end of valid blocks which means that the $i_2$-th block in $Srvi_{i+α}$’s storage is NULL for all $α = 1, 2, \ldots, n' - i_1$. So, it suffices to assign NULL to $i_2$-th block in $Srvi_{i_1}$’s storage except the case where $i_1 = 1$. Note that, all blocks are invalid for $i_1 = 1$. For the exceptional case, we execute step 10 to 15 in Algorithm 3 to remove $i_2$-th blocks in all storage servers since all of them are invalid. Then, it remains to explain the case where at least one block in $i_2$-th block of $Srvi_{i+α}$’s storage for $α ≥ 1$ is not null. Since valid blocks are stored in sequence, we shift valid blocks as in the first case. For detailed explanation about the procedure, see from step 19 to 27 in Algorithm 3.
3.2.3 Insert Query

For a query, the data owner makes a query message $Q(m^*, i_1, i_2)$ and sends it to the CSP with a set composed of $i_2$-th data blocks in $n'+1$ servers for state information and original data. To insert a new message, the CSP checks if there is a room for the message in $i_2$-th row or not. If $MSB(stat_{i_2}, 1) = 0$, then we can see that there are $MSB(stat_{i_2}, 1) = 1, v)$ invalid blocks assigned by NULL. In this case, the new message can be inserted in an $i_2$-th block of a storage server instead of existing NULL. If $i_1$ is the index of the last valid block among $n'$ $i_2$-th blocks, the new data block will be stored in $i_2$-th block of $Srv_{i_1+1}$. Note that the condition can be checked by the following $n' - MSB(stat_{i_2}, 1) = 1$. Otherwise, if the $i_2$-th block in $Srv_{i_1+1}$ is valid, we move $i_2$-th data block in $Srv_{i_1}$ to $Srv_{i_1+1}$ for all $\alpha = i_1 + 1, \ldots, n' - MSB(stat_{i_2}, 1), v)$. This procedure is described from step 2 to 14 in Algorithm 4. When all $i_2$-th blocks are valid, i.e., $MSB(stat_{i_2}, 1) = 1$, we don’t have room for storing the new message in any $i_2$-th block of storage servers. So, we set up one more blocks for all storage servers. If $i_1 = n'$, we simply insert $m^*$ at the end of $i_2$-th block of $Srv_1$ and insert NULL at the end of $i_2$-th block of $Srv_{n'}$ for $\alpha = 2, \ldots, n'$. The simple procedure is described from step 15 to 21 in Algorithm 4. If all $i_2$-th blocks are valid and $i_2$ is less than $n'$, we will divide $i_2$-th blocks into two groups,

\[
\begin{align*}
(m_1+n'(i_2-1), \ldots, m_{i_1+n'(i_2-1)}, m^*, \text{NULL}, \ldots, \text{NULL})_{n'-i_2-1} \\
\text{and (} m_{i_1+1+n'(i_2-1)}, \ldots, m_{n'i_2}, \text{NULL}, \ldots, \text{NULL})_{i_2}
\end{align*}
\]

and store the first group in $i_2$-th row and insert the later at the end of $i_2$-th row. The process is described from step 22 to 35 in Algorithm 4 for detail.

3.3 Decoding

To use entrusted data, the data owner may retrieve distributed chunks from $n'+1$ servers and reconstruct the data. In Algorithm 5, we describe the decoding procedure. Intuitively, we refer $i$-th state information to select valid blocks among $i$-th row of $n'$ storage servers for original data and reconstruct valid strings from the blocks. In the same way, we
The noise removal algorithm takes two indexes, \( i \) and \( j \), as input where \( CH_i[i,\ldots,j] \) is the \( k \)-th block of the chunk \( CH_i \).

3.4 Noise Removal

The noise removal algorithm takes two indexes, \( i_1 \) and \( i_2 \), of the first block and the last block of the interval to be updated. The algorithm also needs a set of data blocks to be updated. As seen in Algorithm 6, \( (CH_i[i_1,\ldots,i_2]) \) are required as input where \( CH_i[i_1,\ldots,i_2] = CH_i[i_1] \cdots CH_i[i_2] \) and \( CH_i[j] \) is the \( j \)-th block of the chunk \( CH_i \).

3.5 Retrievability Proof

We will give brief idea for our approach. Recall that, as defined in Sect. 3.1.1, the authentication code for \( n' \) data blocks \( \{m_1, m_2, \ldots, m_{n'}\} \) is computed as

\[
au = prng(sk, r) + a_1 m_1 + a_2 m_2 + \ldots + a_{n'} m_{n'} \mod f(x)
\]

where \( prng \) is a pseudo-random number generator which guarantees collision resistance, \( a_i \) are secret values kept by the data owner, and \( f(x) \) is an irreducible polynomial defining \( GF(2^n) \). Then, \( au, m_1, \ldots, m_{n'-1}, \) and \( m_{n'} \) are stored in \( Srv_0, Srv_1, \ldots, Srv_{n'-1}, \) and \( Srv_{n'} \), respectively. To check the integrity of the \( n' + 1 \) values, the data owner chooses a random challenge \( c \) and gives it to all servers. For each \( i \), \( Srv_i \) multiplies the stored value by the challenge and returns the result \( re_i \) to the data owner. Then, the data owner can verify the received values by testing the following condition:

\[
re_0 = c \cdot prng(sk, r) + \sum_{i=1}^{n'} a_i \cdot re_i.
\]

To prove the integrity of a file \( F \) maintained by the cloud service provider who stores the file to \( n \) storage servers, each party works as follows. The data owner chooses a random subset of \( I \subset \{1, \ldots, k\} \), and constructs a challenge set \( Chal = \{(i, c_i) | i \in I\} \) where \( r_i \) is a random value chosen from \( GF(2^n) \) for all \( i \). The data owner sends the challenge to the cloud service provider, then it forwards given challenge to storage servers, \( Srv_i \) for \( i = 0, \ldots, n' \). On receiving challenge, each storage server \( Srv_i \) first retrieves \( CH_i \) from his storage, computes his response as \( res_i = \sum_{j \in I} c_i CH_i[j] \), and sends it to CSP. Then the cloud service provider forwards \( Res = \{res_i | i \in I\} \) to the data owner. The integrity of \( F \) can be verified by the following condition:

\[
res_0 = \sum_{i=0}^{n'} c_i \cdot prng(sk, r_i) + \sum_{i=1}^{n'} a_i \cdot res_i.
\]

If the condition holds, the data owner can assured the integrity of the stored data \( F \).

4. Analysis

4.1 Security

Note that in the proposed scheme, data is separated into small fragments and scattered in multiple data servers. Therefore, the adversary who wants to obtain the original stored data has to attack or collude all of the data servers. Practically, increasing the number of servers or shedding into smaller fragments is sufficient to make an outside adversary hard to retrieve the original data. If we consider the higher security model, in which even the manager of data servers can be an adversary, then the inside adversary easily can collect all fragments stored in the data servers. For the case of the secrecy of the data is highly important, the user can choose that managing the data encoding procedure by himself and utilizing multiple storage providers so that one adversary collects all data fragments is not easy. Moreover, there are alternative solutions to guarantee the secrecy of the stored data, for example by inserting arbitrary-size random noise between data fragments.

For these reasons, we concentrate on preventing modification of stored data, which are summarized as following two theorems.

**Theorem 1.** The proposed scheme is secure against a poison attack.

With the poison attack, we mean that an adversary tries to change a specific part of the stored data without detection of the user. If an adversary succeeds in a poison attack,
then the user believes that the changed data is the original one. Note that changing a specific part of the stored data into other contents can be a more serious problem than data leakage. Changing the number of items or promised time due, etc., in the stored data can damage to real-life of the user. In the proposed scheme, this poison attack from outside or even inside attacker can not succeed with meaningful probability.

**Proof.** In the proposed scheme, the user checks the authentication code

\[ au = prng(sk, r) + a_1 m_1 + \cdots + a_{n'} m_{n'} \]

is correct while the data \((au, m_1, \ldots, m_{n'})\) is retrieved from the server. Therefore, the adversary who tries poison attack needs to find out another tuple \((au', m'_1, \ldots, m'_{n'})\) satisfying the equation

\[ au' = prng(sk, r) + a_1 m'_1 + \cdots + a_{n'} m'_{n'}, \]

where \(prng(sk, r)\) is a pseudo-random number and \(sk, r\) and \(a_i's\) are all secret to the adversary.

Without loss of generality, we define a poison attacker, \(POI\), as a polynomial time algorithm on the finite field \(GF(2^b)\). For randomly chosen secret variables \(\{a_i\} \in GF(2^b)\) and input values \(\{au, m_1, \ldots, m_{n'}\}\), \(POI\) outputs \(\{au', m'_1, \ldots, m'_{n'}\}\), where \(au' \in GF(2^b), m'_i \in GF(2^b)\). If the output of \(POI\) satisfies the equation

\[ au' = au + a_1 (m_1 - m'_1) + \cdots + a_{n'} (m_{n'} - m'_{n'}), \]

then \(POI\) wins the game. The advantage of the \(POI\) can be defined as

\[ \text{ADV}(POI) = |Pr[POI wins the game] - 1/|GF(2^b)||. \]

We claim that if there exists no polynomial time algorithm \(POI\) with non-negligible advantage, then the proposed scheme is secure against poison attack.

Note that \(sk\) is the secret key which is kept by the user and \(r\) is a randomly selected nonce whenever the user calculates the authentication code. Since the same value of \(r\) is never used for different authentication codes, it is assumed that no polynomial time algorithm can distinguish \(prng(sk, r)\) from a random string. It means that the adversary cannot construct a system of equations about \(a_i's\) even if the adversary can access the whole stored data. Therefore the possible strategy for an adversary is to choose random tuple \((au', m'_1, \ldots, m'_{n'}\).

Note that if the the equation

\[ au' = au + a_1 (m_1 - m'_1) + \cdots + a_{n'} (m_{n'} - m'_{n'}) \]

has unique solution in \(GF(2^b)\) for any given tuples \((au, m_1, \ldots, m_{n'})\) and \((m'_1, \ldots, m'_{n'})\). It implies that the probability of that randomly chosen tuple \((au', m'_1, \ldots, m'_{n'})\) satisfies the equation

\[ au' = au + a_1 (m_1 - m'_1) + \cdots + a_{n'} (m_{n'} - m'_{n'}), \]

is \(1/|GF(2^b)|\). Finally, if all \(a_i's\) are secret values and \(prng(sk, r)\) is never repeated, then the \(\text{ADV}(POI)\) is negligible. It concludes the proof.

\[ \square \]

**Theorem 2.** The proposed scheme mitigates deletion of stored data.

**Proof.** In this theorem, we consider another type of attack in which the adversary deletes stored data to disrupt the proper use of the storage service. Note that an attack of this type can be executable only by the outside attackers. Since any adversary cannot modify the stored data without detection, the inside attacker has no merit for deleting data which is maintained by himself. For the outside attackers, it is hard to delete all stored data. In the proposed scheme, the embedded error-correcting code guarantees that a user can recover the deleted data where the attacked storage is less than \(n''/2\), where \(n''\) is the system parameter.

\[ \square \]

### 4.2 Performance

The main merit of our system is that we need to modify limited blocks for each query, which means that only a few computational cost is required for each query. The technique in [8] also can support dynamic date update queries, but it requires \(O(n)\) operations for each update query, which means that our construction is very efficient in terms of data update since it requires \(O(1)\) operations. So, for the computational complexity, we will focus on the cost of encoding and decoding operations which are performed on the full data differently from other operations that treat only a few blocks. To analyze the efficiency of two main data processing operations, encoding and decoding, we compare their computational complexities with well-known cryptographic operation, the AES-CBC encryption. Note that, for comparison, we choose AES-CBC since it is one of well-known cryptographic function which is used for dealing large-scaled data. In the experiment, we repeat 1000 times of each operation for 100 MB data. As seen in Table 1, our technique guarantees sufficient efficiency.

In our technique, we have to determine the size of some parameters, including the number of servers, the size of a shred, and the size of a random noise. Our technique can support any parameters. So, we can choose a set of parameters for given application. For example, the number of servers will be determined by the expected number of corrupted server and the number of servers can be used. The size of a shred will be determined by the size of a unit data which is changed for each data update. Note that, the goal of

| Operation                  | Time Complexity |
|----------------------------|-----------------|
| AES-CBC Encryption         | 61.9 ms         |
| Encoding of Our Technique  | 21.9 ms         |
| Decoding of Our Technique  | 28.5 ms         |
our work is to give stronger security without using encryption schemes, and stronger security can be supported if more servers are used, smaller shreds are used, or larger random numbers are used.

5. Conclusion

In this paper, we proposed a new technique for using dynamic data in cloud storage systems. Specifically, our construction includes the following features. For dynamic data, we design a new data chunk generation strategy. Our chunk generating algorithm guarantee efficient data insertion, deletion, and modification due to the structure of data chunk. Note that existing chunking techniques cannot support data updates efficiently. For each data update, existing techniques require $O(n)$ operations where $n$ is the size of data. Due to the suitable data structure of chunks, our technique requires $O(1)$ operations for update queries.

References

[1] G. Ateniese, R. Di Pietro, L.V. Mancini, and G. Tsudik, “Scalable and efficient provable data possession,” Proc. 4th International Conference on Security and Privacy in Communication Networks, Article No 9, ACM, 2008.
[2] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, “DepSky: Dependable and secure storage in a cloud-of-clouds,” ACM Trans. Storage, vol.9, no.4, Article No 12, 2013.
[3] K.D. Bowers, A. Juels, and A. Oprea, “HAIL: A high-availability and integrity layer for cloud storage,” Proc. CCS ’09, pp.187–198, ACM, 2009.
[4] Y. Zhang and M. Blanton, “Efficient dynamic provable possession of remote data via update trees,” ACM Trans. Storage, vol.12, no.2, Article No 9, ACM, Feb. 2016.
[5] C.C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia, “Dynamic provable data possession,” ACM Trans. Inf. Syst. Secur., vol.17, No 4, Article No 15, ACM, April 2015.
[6] F. Wang, L. Xu, H. Wang, and Z. Chen, “Identity-based non-repudiable dynamic provable data possession in cloud storage,” Computers & Electrical Engineering, vol.69, pp.521–533, 2018.
[7] A. Juels and B.S. Kaliski, “Pors: Proofs of retrievability for large files,” Proc. CCS ’07, pp.584–597, ACM, 2007.
[8] Z. Ren, L. Wang, Q. Wang, and M. Xu, “Dynamic proofs of retrievability for coded cloud storage systems,” IEEE Trans. Serv. Comput., vol.11, no.4., pp.685–698, IEEE, Aug. 2018.
[9] H. Shacham and B. Waters, “Compact proofs of retrievability,” Proc. Advances in Cryptology - ASIACRYPT 2008, Lecture Notes in Computer Science, vol.5350, pp.90–107, Springer Berlin Heidelberg, 2008.
[10] Q. Zheng and S. Xu, “Fair and dynamic proofs of retrievability,” Proc. CODASPY ’11, pp.237–248, ACM, 2011.
[11] C. Gritti, R. Chen, W. Susilo, and T. Plantard, “Dynamic provable data possession protocols with public verifiability and data privacy,” Proc. ISPEC’ 17, Lecture Notes in Computer Science, vol.10701, pp.485–505, Springer, 2017.

Nam-Su Jho received the B.S. degree in Mathematics from Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1999, and his Ph.D. degree in Mathematics from Seoul National University, Korea, in 2007. Since 2007, he has been with the ETRI, Daejeon, Korea as a senior researcher. His research interests include cryptography and information theory.

Daesung Moon received the M.S. degree from the Department of Computer Engineering, Busan National University, Busan, South Korea, in 2001, and the Ph.D. degree in computer science from Korea University, Seoul, South Korea, in 2007. He joined the Electronics and Telecommunications Research Institute, in 2000, where he is currently a Principal Researcher. His research interests include network security, data mining, and biometrics.

Taek-Young Youn received his B.S., M.S., and Ph.D. from Korea University in 2003, 2005, and 2009, respectively. From 2010 to 2020, he has worked as a senior researcher at Electronics and Telecommunications Research Institute (ETRI), Korea. From 2016 to 2020, he served as an associate professor in University of Science and Technology (UST), Korea. From 2020, he serves as an assistant professor in Dankook University, Korea. His research interests include cryptography, information security, authentication, data privacy, and security issues in various communications.