SE-ORAM: A Storage-Efficient Oblivious RAM for Privacy-Preserving Access to Cloud Storage

Qiumao Ma, Jinsheng Zhang, Wensheng Zhang, and Daji Qiao
Iowa State University, Ames, IA, USA 50011
Email:{qmma, alexzjs, wzhang, daji}@iastate.edu

***Abstract*—Oblivious RAM (ORAM) is a security-provable approach for protecting clients' access patterns to remote cloud storage. Recently, numerous ORAM constructions have been proposed to improve the communication efficiency of the ORAM model, but little attention has been paid to the storage efficiency. The state-of-the-art ORAM constructions have the storage overhead of $O(N)$ or $O(N \log N)$ blocks at the server, when N data blocks are hosted. To fill the blank, this paper proposes a storage-efficient ORAM (SE-ORAM) construction with configurable security parameter λ and zero storage overhead at the server. Extensive analysis has also been conducted and the results show that, SE-ORAM achieves the configured level of security, introduces zero storage overhead to the storage server (i.e., the storage server only storages N data blocks), and incurs $O(\log N)$ blocks storage overhead at the client, as long as $\lambda \geq 2$ and each node on the storage tree stores $4 \log N$ or more data blocks.**

Key words: Cloud System, Data Outsourcing, Oblivious RAM, Privacy Preservation, Access Pattern.

I. Introduction

A. Motivations

Cloud storage services such as Amazon S3 and Dropbox, have been popularly utilized by business and individual clients to host their data. Due to security and privacy concerns, the clients may encrypt their sensitive data before outsourcing them. Nevertheless, data encryption itself is insufficient for data security, because the secrecy of data can still be exposed if a client's access pattern to the data is revealed [1].

The oblivious RAM (ORAM) model [2], which continues shuffling data as the data are accessed, has been a well-known security-provable approach for access pattern protection. Many ORAM constructions [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18] have recently been proposed to make the ORAM model more feasible in practice.

Most of these research have focused on the communication efficiency improvement, but the storage efficiency has not received much attention. To host N data blocks, in general, the state-of-the-art ORAM constructions need the storage server to also store $O(N)$ or $O(N \log N)$ dummy data blocks; in particular, as the most communication-efficient ORAM constructions, Path-ORAM [13] and SCORAM [18] each requires the server to keep $5N$ dummy blocks, and a server running P-PIR [19] has to store $(2 \log N - 1)N$ dummy blocks. Though the unit price for storage is cheaper than that for communication, the significantly-enlarged demand for storage capacity due to the high storage overhead ratio could pose as a high monetary cost to the client, especially when the client needs to have a huge amount of data kept for a long period of time, and when the data need to be replicated in multiple copies for redundancy. Hence, reducing the storage overhead is also imperative.

B. Research Goal and Rationales

This study aims to design an ORAM construction with *zero storage overhead* at the server. The research is based on the following observations.

The security goal of an ORAM construction is to prevent the storage server from correctly inferring a client's private data access sequence from the client's storage location access sequence that the server can observe. Existing ORAM constructions target at perfect security; that is, the probability is at most $\frac{1}{N^n}$ for the server to correctly infer a sequence with n data accesses from any observed location access sequence, since N^n is the total number of sequences with n data accesses. To attain this goal,

the client's query and shuffling operations should be fully random and independent of each other.

Particularly, let us consider the tree-based ORAM [12]. When a data block is assigned to a path of the storage tree, the path is selected uniformly at random to make the query process appear fully random. During an eviction process, nodes are randomly selected to evict data, and each selected node is dictated to evict a data block to its left or right child with the equal probability. Due to the randomness, following *undesired situation* may happen: a node without any real data block evictable to its left (or right) child is selected to evict data to left (or right). To deal with such situation, dummy blocks are pre-introduced into the storage when the system is initialized; and it has been shown that, $O(N)$ or $O(N \log N)$ dummy blocks are needed to keep a low failure probability, i.e., the probability that a node has already used up dummy blocks when it is in the afore-described undesired situation.

To address the above issue without introducing storage overhead to the server, we design a new eviction algorithm based on the following intuitions:

- *Eviction with Non-uniform Probabilities.* When a node is selected to evict data to its children, it can use different probabilities for different children; i.e., a larger probability to evict data to its left child if more of its data blocks are evictable to left, and vice versa. This way, the chance could be significantly reduced for the afore-mentioned undesired situation to occur.
- *On-demand Introduction of Dummies.* Nevertheless, the undesired situation could still occur. To deal with it, a dummy block (evictable to both left and right) is inserted *on demand* to replace a real data block, which is moved to the client's cache. Note that, the storage server still stores the same number of data blocks, though some of the blocks become dummies.
- *Periodical Removal of Dummies.* As the system keeps running, more dummy blocks are inserted to the server and the client's cache may overflow. To address this issue, an extra query and eviction process is launched periodically to retrieve and discard a dummy from the server and evict a real data block from the client to the server.

Due to the non-uniform eviction probabilities used in the eviction algorithm, perfect security is not attained. To quantify the level of security that our new ORAM construction can achieve, we propose a more generic security definition, which quantify security level by a parameter λ: if an ORAM construction is secure with parameter λ, the probability is at most $(\frac{1}{N^n})^{1-\frac{1}{\lambda}}$ for the server to correctly infer a sequence with n data accesses from any storage location access sequence. That is, the advantage for the server to discover a client's access pattern is upper-bounded by $(\frac{1}{N^n})^{1-\frac{1}{\lambda}} - \frac{1}{N^n}$, which decreases as λ increases. We argue that, this notion of security can be useful in practice, particularly when a large number of data blocks are outsourced and/or protecting relatively long access patterns (i.e., n is large) is the major security goal. For example, when $N = 2^{40}$ and $n = 10$ (or $N = 2^{10}$ and $n = 40$), and $\lambda = 2$, the server's advantage is upper-bounded by 2^{-200}, which may be considered "negligibly small" in practice. Besides, the definition allows a client of our ORAM construction to configure her desired level of security, and manage the tradeoffs between security and performance.

C. Results

Based on the new eviction algorithm and the new definition of security, we formalize a generic SE-ORAM construction with parameter λ. Through rigorous security and cost analysis, we show that the construction is secure under the definition, and the number of introduced dummy blocks is no more than $x \log N$ with probability $1 - \frac{1}{N^{2x}}$, as long as $\lambda \geq 2$ and each node on the storage tree can store $4 \log N$ or more data blocks. We also instantiate a SE-ORAM construction by setting $\lambda = 2$, analyze its performance, and compare it with the state-of-art ORAM constructions. To summarize, this study makes the following contributions:

- We introduce a generic security definition for ORAM constructions. It allows a client to configure a desired security level and manage the tradeoffs between security and performance.
- We propose SE-ORAM, a generic storage-efficient ORAM construction with configurable security parameter λ. Rigorous analysis shows that, SE-ORAM achieves the configured level of security, introduces zero storage overhead to the storage server (i.e., the storage server only

storages N data blocks), and incurs $O(\log N)$ blocks storage overhead at the client, as long as $\lambda \geq 2$ and each node on the storage tree stores $4 \log N$ or more data blocks.

D. Discussions

The current SE-ORAM construction incurs a communication overhead of $O(\log^2 N \cdot B)$ bits for each data query launched by the client, where B is the size of a data block in the unit of bit. However, the construction can be enhanced by incorporating some existing techniques to achieve the same level of communication efficiency as the state-of-the-art constructions. For example, additive Homomorphic encryption-based PIR primitives [19] can be used to reduce the number of data blocks that should be transferred between the server and the client, as the way these primitives are used in P-PIR [19] to reduce the communication overhead of T-ORAM [12]. In addition, the technique of recursively exporting the index structure from the client to the server [12], [19], [13], which has been widely used in exiting ORAM constructions, can also be incorporated into the SE-ORAM construction to reduce the storage cost of the client. Due to space limit, we do not include such optimizations in this paper, in order to focus the presentation on the major goal of improving storage efficiency.

E. Organization

In the rest of the paper, Section II reviews the related works on ORAM. Section III presents the security definition. Section IV presents the basic design of SE-ORAM, which is followed by security analysis in Section V and overhead analysis and comparison in Section VI. Finally, Section VII concludes the paper.

II. Related Work

ORAM constructions can be roughly categoried into two classes, hash-based ORAMs and index-based ORAMs. This section reviews the performance of these constructions in terms of storage and communication costs.

A. Hash-based ORAMs

Hash-based Oblivious RAMs [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] organize the server storage as a hierarchy of layers. Each layer contains either a series of buckets [2], [9], [10], [11], or a pair of Cuckoo Hash tables with stash [8], [3], [4], [5], [6], [7]. In a bucket ORAM proposed in [2], the server needs to additionally store $(2 \log N - 1)N$ dummy blocks in order to host its client's N real data blocks; its communication cost is $O(\log^3 N)$ blocks per query, with a constant client-side storage. In a bucket ORAM proposed in [9], [10], [11], the server additionally stores at least N dummy blocks and cN bits ($0 < c < 1$) of Bloom Filters for each layer; its communication cost is $O(\log^2 N \log \log N)$ blocks per query, with a client-side storage of $O(\log^2 N)$ blocks. In a Cuckoo Hash ORAM [8], [3], [4], [5], [6], [7], the server stores at least $7N$ dummy data blocks; its communication cost is $O(\log^2 N)$ blocks per query with a constant client-side storage, or $O(\log N)$ blocks per query with a client-side storage of $O(N^c)$ blocks ($0 < c < 1$).

B. Index-based ORAMs

Index-based ORAMs [12], [13], [14], [15], [16], [17], [19] use index table for data lookup. They require the client to either store the index table locally, or outsource it to the server recursively in a way similar to storing their data, at the expense of increased communication cost. Representative index-based ORAMs include Partition ORAM [15], binary tree ORAM (T-ORAM) [12], Path ORAM [13], Gentry's ORAM [16], P-PIR [19] and SCORAM [18]. Partition ORAM organizes its server-side storage as a number of partitions, where each partition is a fully-functional Oblivious RAM. In Partition ORAM, the server side storage needs to store $2.2N$ dummy blocks, and incurs a communication cost of $O(\log N)$ blocks per query, with a client-side storage of $O(cN)$ blocks. The other index-based ORAMs organize their server-side storage as a tree, where each node is a bucket storing a certain number of data blocks. For T-ORAM and P-PIR, the server needs to store $(2 \log N - 1)N$ dummy blocks, and incurs a communication cost of $O(\log^2 N)$ blocks per query, with a constant client-side storage. Path ORAM and SCORAM each stores at least $5N$ dummy blocks at the server, and incurs a communication cost of $O(\log N \cdot B)$ blocks per query, with a client-side storage of $O(\log N) \cdot \omega(1)$ blocks, where $\omega(1)$ is a security parameter. At last,

Gentry's ORAM requires the server to store at least N dummy blocks, and it achieves a communication cost of $O(\log^2 N \log\log N)$ blocks per query, with a client-side storage of $O(\log^2 N)$ blocks.

III. Security Definition

Let $\lambda > 1$ be a security parameter. A client exports N equal-size data blocks to a remote storage server. Each data access from the client, which should be kept private, is one of the following two types: (i) read a data block D of unique ID i from the storage, denoted as a 3-tuple $(read, i, D)$; (ii) write a data block D of unique ID i to the storage, denoted as a 3-tuple $(write, i, D)$. To accomplish each data access, the client needs to access some storage location(s) at the remote storage server. Each location access, which can be observed by the server, is one of the following types: (i) retrieve (i.e., read) a data block D from a location l, denoted as a 3-tuple $(read, l, D)$; (ii) upload (i.e., write) a data block D to a location l, denoted as a 3-tuple $(write, l, D)$.

We assume the remote storage server is honest but curious; that is, it stores data and serves the client's location access requests honestly, but it may attempt to figure out the client's data access pattern hidden behind the location accesses. The network connection between the client and the server is assumed to be secure; in practice, this can be achieved using well-known techniques such as SSL [20].

We define the security of our proposed SE-ORAM(λ, N), which has security parameter λ and stores N real data blocks, as follows.

Definition *In SE-ORAM(λ, N), let $\vec{x}_n = \langle\ (op_1, i_1, D_1), (op_2, i_2, D_2), \cdots, (op_n, i_n, D_n)\ \rangle$ denote a private sequence of the client's n data accesses, where each op_i is either a $read$ or $write$ operation; let random variable $A(\vec{x}_n)$ denote the sequence of location accesses (observable by the server) that the client uses to accomplish data access sequence $\vec{x}_n$. Note that, there may exist multiple location access sequences that can accomplish $\vec{x}_n$, each with certain probability to be used by the client as $A(\vec{x}_n)$; hence, $A(\vec{x}_n)$ is a random variable.*

Let $\mathcal{X}_n$ denote the set of all possible sequences of the client's n data accesses, and $\mathcal{A}_n$ the set of all location access sequences that can accomplish at least one data access sequence in $\mathcal{X}_n$.

Let $Pr[\vec{T}_n|\vec{A}_n]$, where $\vec{A}_n \in \mathcal{A}_n$ and $\vec{T}_n \in \mathcal{X}_n$, denote the conditional probability of $A(\vec{T}_n) = \vec{A}_n$ given that $\vec{A}_n$ has been observed by the server.

SE-ORAM(λ, N) is said to be secure if $\forall \vec{A}_n \in \mathcal{A}_n$ and $\forall \vec{T}_n \in \mathcal{X}_n$:

$$(\frac{1}{N^n})^{1+\frac{1}{\lambda}} \leq Pr[\vec{T}_n|\vec{A}_n] \leq (\frac{1}{N^n})^{1-\frac{1}{\lambda}}. \qquad (1)$$

Note that, if the client's data access pattern is perfectly protected, $Pr[\vec{T}_n|\vec{A}_n] = \frac{1}{N^n}$; i.e., no matter what location access sequence (that can accomplish a certain sequence with n data accesses) has been observed, it is impossible for the server to infer the client's actual data access sequence hidden behind this observed pattern, because each of the N^n data access sequences has the same probability $\frac{1}{N^n}$ to be the one. According to the above definition, when $\lambda \to \infty$, $Pr[\vec{T}_n|\vec{A}_n] \to \frac{1}{N^n}$ indeed.

Generally speaking, if an SE-ORAM(λ, N) is secure, the advantage for the server to infer the client's actual data access sequence $\vec{T}_n$ from a location access sequence $\vec{A}_n$ that has been observed, i.e., $|Pr[\vec{T}_n|\vec{A}_n] - \frac{1}{N^n}|$, is upper-bounded by $(\frac{1}{N^n})^{1-\frac{1}{\lambda}} - \frac{1}{N^n}$; the larger is λ, the smaller is the bound. Hence, parameter λ quantifies the level of security that an SE-ORAM construction can attain.

IV. The SE-ORAM Construction

This section elaborates the SE-ORAM construction in terms of storage organization, data query and data eviction algorithms.

A. Storage Organization and Initialization

1) Server-side Storage: In the server, the storage is initially organized as a complete binary tree. Each node on the tree can store up to s data blocks, where s is a system parameter and an even number. To simplify presentation, we denote the height of tree as h and assume the total number of data blocks N as $N = s \cdot \sum_{l=0}^{h} 2^l = s(2^{h+1}-1)$. Hence, the number of level-h nodes is 2^h, which also is $\frac{N/s+1}{2} \approx \frac{N}{2s}$.

The content of each data block B_i is encrypted probabilistically with a symmetric cipher (e.g., AES) before the blocks are randomly distributed to the nodes on the tree. Specifically, denoting the plain-text content of a block B_i as D_i, we have $B_i = E(r|D_i)$, where r is a nonce and E is a symmetric encryption function.

In each node n, data blocks are randomly divided into two equal-size groups, called *left group* and *right group* and denoted as $G_L(n)$ and $G_R(n)$. Each block in the left group randomly picks a level-h node n' from the left branch of n, and the block is restricted to be evictable toward node n' only; hence, we call the ID of node n' as the *path ID* of the data block. Similarly, each block in the right group also randomly selects a level-h node from the right branch of n, whose ID becomes the block's path ID.

As the data query and eviction processes go on, the tree may become incomplete and some nodes may become non-full (i.e., containing less than s data blocks). Figure 1(a) shows an example of the server-side storage. Here, $h = 3$, two of the level-h nodes (i.e., $n_{3,1}$ and $n_{3,6}$) are absent, and one level-h node (i.e., $n_{3,2}$) is non-full. Also, the data blocks with path IDs of $n_{3,0}$, $n_{3,4}$ and $n_{3,7}$ cannot be completely contained in nodes between level 0 to level 3; hence, *supplementary nodes* have been introduced to provide additional storage, e.g., $n_{4,0}$ for $n_{3,0}$, $n_{4,4}$ and $n_{5,4}$ for $n_{3,4}$, and $n_{4,7}$ for $n_{3,7}$.

2) Client-side Storage: The client-side storage includes three parts: (i) an *index table* $\mathcal{I}$ maintaining the mapping between data block IDs and their path IDs (therefore it has N entries and each entry has h bits); (ii) a *data block cache* $\mathcal{C}$ used to cache data blocks; and (iii) a small *secret storage* storing the key for symmetric data encryption.

B. Data Query

When the client queries a data block of ID t (denoted as B_t), it first checks whether B_t is in $\mathcal{C}$; if so, the block is accessed and retained in $\mathcal{C}$. Otherwise, the client looks up the index table $\mathcal{I}$ to obtain B_t's path ID (i.e., the ID of a level-h node, denoted as n_t^h hereafter). Then, the client follows the steps below to obliviously retrieve B_t.

The client requests the server to return data blocks on the path from the root to the n_t^h. In response, the server first finds out all the nodes that should be returned to the client, based on the current topology of the tree: (i) *Case I* - if node n_t^h is currently on the tree and has no supplementary nodes, all the nodes along the path from the root to n_t^h should be returned. (ii) *Case II* - if n_t^h is currently on the tree and has supplementary nodes, all the nodes along the path from the root to n_t^h as well as all of n_t^h's supplementary nodes should be returned. (iii) *Case III* - if node n_t^h is absent, the server acts as follows. Let $n_t^{h_0}$ denote the node that is on the path from the root toward n_t^h (as if n_t^h were still there) and the furthest away from the root. Let $n_t^{h_1}$ denote the leaf node of the longest branch within the subtree rooted at $n_t^{h_0}$. Note that, the path from the root to $n_t^{h_1}$ is the longest path that has the largest overlap with the path from the root to n_t^h (as if n_t^h were still there). All the nodes along the path from the root to $n_t^{h_1}$ should be returned. Let us denote the nodes that should be returned as $n_t^0,\ n_t^1,\ \cdots,\ n_t^L$, where n_t^0 is the root and n_t^L is the leaf node. Among them, suppose node n_t^y on layer y contains B_t.

The server returns only the blocks in n_t^L in the first round. If B_t is among the blocks, the client keeps B_t locally, re-encrypts the rest of the blocks and uploads them back to the server; otherwise, one arbitrary block denoted as B_L is picked from the returned blocks, and the rest of the blocks are re-encrypted and uploaded back to the server.

Next, the server returns all the blocks in n_t^{L-1}. If B_t is among the blocks, the block is kept locally, and the rest of the blocks in n_t^{L-1} together with B_L are re-encrypted and uploaded back to the server. Otherwise, all the blocks in n_t^{L-1} are re-encrypted and uploaded to the server. This process continues until all the blocks on the selected path have been returned to the client, re-encryption and finally uploaded back to the server. Figure 1 shows two examples of data query.

C. Data Eviction

Data eviction should be conducted following the query process, to store the query target B_t back to the server obliviously.

A path (i.e., a level-h node) is selected uniformly at random for B_t, and then all the data blocks on the path are retrieved node-by-node. The eviction process should place B_t into a node on the selected path before the blocks are all re-encrypted and uploaded back to the server. The ID of the path becomes the new path ID of B_t and hence should be recorded in the client's index table $\mathcal{I}$. During the course of eviction, some other blocks may be moved; the movement should ensure that, *a data block stays in a node on the path specified by its*

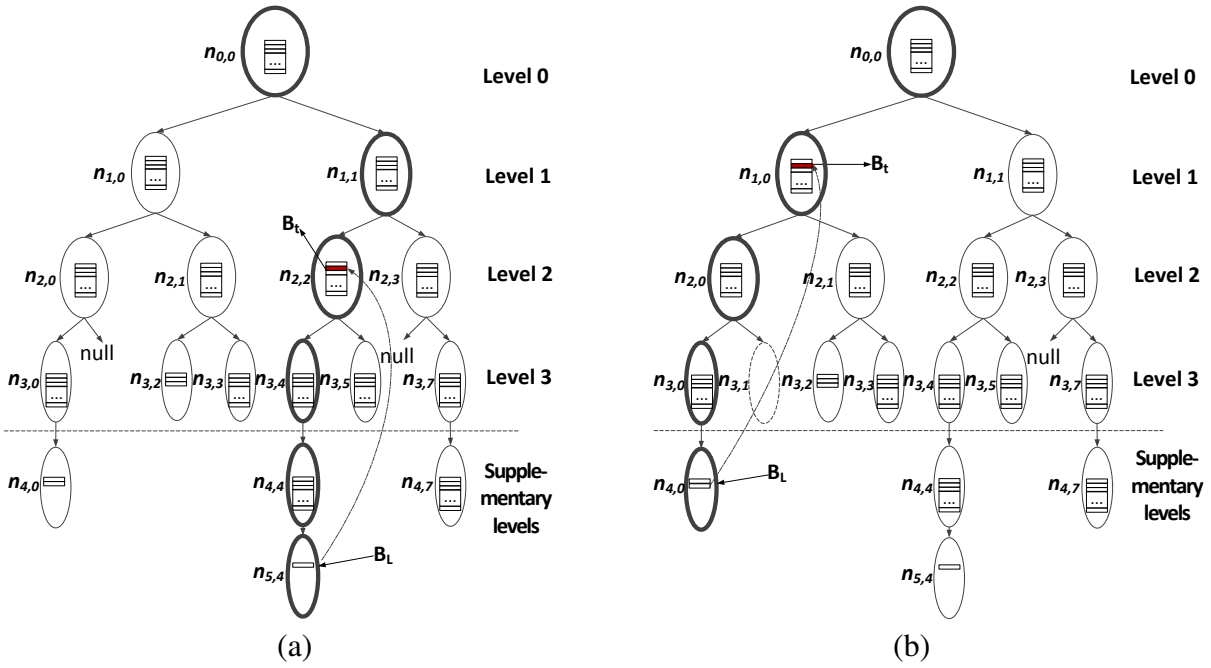

Figure 1. Query Examples. In (a), query target B_t is at node $n_{2,2}$ and has path ID $n_{3,4}$. Node $n_{3,4}$ exists on the tree and has two supplementary nodes. The client requests the server to retrieve nodes from the root to $n_{5,4}$ which is the further supplementary node of $n_{3,4}$. Then, B_L obliviously replaces B_t; finally, as node $n_{5,4}$ becomes empty after B_L has moved, the node is removed from the tree. In (b), query target B_t is at node $n_{1,0}$ and has path ID $n_{3,1}$. Node $n_{3,1}$ does not exist on the tree. The client requests the server to retrieve nodes on the path from the root to $n_{4,0}$, which is the longest path that has the largest overlap with the path from the root to $n_{3,1}$. Then, the client obliviously replaces B_t with B_L.

path ID or it stays in the local cache maintained by the client. The eviction steps are elaborated in the following, and an example containing evictions in four layers of the storage tree is given in Figure 3 in Appendix 1.

a) E1: Initial Step.: Let B_e denote the current block to evict (called the evicted block), and n_e the current node (called the evicting node) to accommodate B_e's eviction. Initially, $B_e = B_t$ and $n_e =$ root. All the data blocks in n_e are sent from the server to the client.

b) E2: Conditional Termination.: If n_e is non-full, the client writes B_e into n_e. Then, B_e is put into the left or right group of n_e (i.e., $G_L(n_e)$ or $G_R(n_e)$) according to its path ID; note that, if B_e is a dummy, it is randomly put into either $G_L(n_e)$ or $G_R(n_e)$.

Another condition for the process to terminate is when n_e is a level-h node. B_e should be written to the furthest supplementary node of n_e. If the supplementary node is full, an additional supplementary node is created to contain B_e.

For both cases, blocks in n_e (and its supplementary nodes if applicable) should be re-encrypted and uploaded back to the server.

c) E3: Selection of the Next Evicting Node.: Depending on the sizes of $G_L(n_e)$ and $G_R(n_e)$, the selection of the next evicting node (denoted as n'_e) works as follows:

If $|G_L(n_e)| > |G_R(n_e)|$, the left child of n_e is selected as n'_e with probability $1-p$ while the right child is selected as n'_e with probability p, where $p = \frac{1}{2^{1/\lambda}+1}$ and $1-p = \frac{2^{1/\lambda}}{2^{1/\lambda}+1}$.

If $|G_L(n_e)| = |G_R(n_e)|$, the left and right children of n_e have the same probability 0.5 to be selected as n'_e.

If $|G_L(n_e)| < |G_R(n_e)|$, the left child of n_e is selected to be n'_e with probability p while the right child is selected to be n'_e with probability $1-p$.

Note that, if n'_e does not exist on the tree, it should be created: 1) If n_e is a level-h node or a supplementary node, a supplementary node n'_e is created and linked to n_e; 2) otherwise, n'_e is created as a left or right child node of n_e accordingly.

d) E4: Selection of the Next Evicted Block.: There are a few different cases. Case I - If B_e is

a dummy block, it remains to be the next evicted block denoted as B'_e. Case II - If B_e is a real data block, and there is at least one block in $n_e \cup \{B_e\}$ (we also use n_e to denote the set of all data blocks in n_e, for simplicity) that is evictable to n'_e, one such block is selected to be B'_e and the selected block is replaced by B_e. Case III - If B_e is a real data block, no data blocks in $n_e \cup \{B_e\}$ are evictable to n'_e, but n_e contains dummy blocks, one dummy block is selected as B'_e and the selected block is replaced by B_e. Case IV - If B_e is a real data block, no data blocks in $n_e \cup \{B_e\}$ are evictable to n'_e, and n_e does not contain any dummy blocks, a new dummy block is created to be B'_e, while the original B_e is saved to the client's local cache. Finally, all current blocks in n_e are re-encrypted and uploaded back to the server; then, after $B_e \leftarrow B'_e$ and $n_e \leftarrow n'_e$ are performed, the process continues to Step E2.

D. Extra Query-Eviction Round

With the above eviction algorithm, the dummy blocks at the storage server and the cached blocks at the client may keep increasing as more data blocks are queried. To bound the number of these blocks and hence the storage overhead, we propose to periodically remove dummy blocks and dump cached data blocks as follows.

Every time after an eviction process is completed, with probability ρ, the following extra round of query and eviction is conducted: The client randomly selects a path. Depending on the selected path, this step proceeds with one of the following two cases. Case I - the selected path contains dummy blocks. In this case, one dummy block is retrieved from the selected path following the above data query algorithm. Then, one real data block is randomly picked from the client's cache, and evicted to the tree structure at the storage server following the above data eviction algorithm. Case II - the selected path does not contain any dummy blocks. In this case, one data block is randomly retrieved from the selected path following the above data query algorithm, and then evicted following the above data eviction algorithm.

V. Security Analysis

Recall that Section III defines the concepts of data access sequence and location access sequence, and introduces the notations of $\mathcal{X}_n$, $\mathcal{A}_n$, random variable $A(\vec{x}_n)$ for $\vec{x}_n \in \mathcal{X}_n$, and conditional probability $Pr[\vec{T}_n|\vec{A}_n]$ for $\vec{T}_n \in \mathcal{X}_n$ and $\vec{A}_n \in \mathcal{A}_n$. To facilitate the security analysis in this section, we further introduce the following notations:

For any $\vec{A}_n \in \mathcal{A}_n$, we expand it to $\vec{A}_n = q_1, e_1, \cdots, q_n, e_n$. Here, for each $i = 1, \cdots, n$, q_i denotes the path accessed during the i-th query process and e_i denotes the path accessed during the i-th eviction process.

Each e_i in the above is further expanded to $e_i = e_{i,1}, e_{i,2}, \cdots, e_{i,h_i}$. Here, $e_{i,j} \in \{0,\ 1\}$ for $j \in \{0,\ 1,\ \cdots,\ h_i\}$. $e_{i,1}$ represents whether the root node (i.e., the first evicting node in the i-th eviction process) evicts data to its left (if $e_{i,1} = 0$) or right child (if $e_{i,1} = 1$), and $e_{i,j}$ represents whether the $(j-1)$-th evicting node evicts data to its left (if $e_{i,j} = 0$) or right child (if $e_{i,j} = 1$).

Let $Pr[\vec{x}_n]$, where $\vec{x}_n \in \mathcal{X}_n$, denote the probability that $\vec{x}_n$ is the client's actual data access sequence.

Let $Pr[\vec{A}_n|\vec{x}_n]$, where $\vec{A}_n \in \mathcal{A}_n$ and $\vec{x}_n \in \mathcal{X}_n$, denote the conditional probability of $A(\vec{x}_n) = \vec{A}_n$ given that $\vec{x}_n$ is the client's actual data access sequence.

Let $Pr[q_i|\vec{x}_n; q_1, e_1, \cdots, q_{i-1}, e_{i-1}]$ denote the conditional probability of q_i being selected to access during the i-th query process given that the client's actual data access sequence is $\vec{x}_n$ and the location access sequence has been $q_1, e_1, \cdots, q_{i-1}, e_{i-1}$ before the i-th query is processed.

Let $Pr[e_{i,j}|\vec{x}_n; q_1, e_1, \cdots, q_i, e_{i,1}, \cdots, e_{i,j-1}]$ denote the conditional probability for the i-th evicting node to evict to left (if $e_{i,j}$ is 0) or right (if $e_{i,j}$ is 1), given that the client's actual data access sequence is $\vec{x}_n$ and the location access sequence has been $q_1, e_1, \cdots, q_i, e_{i,1}, \cdots, e_{i,j-1}$ before this evicting node is accessed.

Lemma 1: In SE-ORAM(λ, N), for $\forall\ \vec{x}_n \in \mathcal{X}_n$ and $\forall\ i \in \{1,\ 2,\ \cdots,\ n\}$,

$$Pr[q_i \mid \vec{x}_n;\ q_1,\ e_1,\ \cdots,\ q_{i-1},\ e_{i-1}] \ =\ \frac{2s}{N}.$$

Proof: Initially and after being queried, data blocks are all distributed to the paths uniformly at random. Hence, every path has the same probability to be selected for each query. The probability is $\frac{2s}{N}$ as the total number of paths is $\frac{N}{2s}$. ■

Lemma 2: In SE-ORAM(λ, N), for $\forall \vec{x}_n \in \mathcal{X}_n$, $\forall i \in \{1,2,\cdots,n\}$ and $\forall j \in \{1,2,\cdots,h_i\}$: $p \leq Pr[e_{i,j} \mid \vec{x}_n; q_1, e_1, \cdots, q_{i-1}, e_{i,1}, \cdots, e_{i,j-1}] \leq 1-p$.

Proof: During an eviction process, the probability for an evicting node to evict a data block to its left (or right) child is between p and $1-p$. The lemma is therefore proved. ■

Theorem 1: SE-ORAM(λ, N) is secure under Definition III. That is, for any $\vec{A}_n \in \mathcal{A}_n$ and $\vec{T}_n \in \mathcal{X}_n$,

$$\left(\frac{1}{N^n}\right)^{1+\frac{1}{\lambda}} \leq Pr[\vec{T}_n \mid \vec{A}_n] \leq \left(\frac{1}{N^n}\right)^{1-\frac{1}{\lambda}}. \quad (2)$$

Proof: In Appendix 2. ■

VI. Cost Analysis

A. Storage Overhead

In SE-ORAM, the storage server initially stores only real data blocks exported by the client. As the system keeps running, dummy blocks are introduced or removed, and the server needs to store some dummy blocks. However, when a dummy block is introduced, it always replaces a real data block which should be moved to the client's cache; when a dummy block is removed, it is always replaced with a real data block previously cached by the client. Hence, the storage consumption at the server keeps unchanged. In this sense, there is no storage overhead at the server. However, extra storage overhead has been introduced to the client, who needs to cache real data blocks that have been replaced by dummies.

In the following, we analyze the number of dummy blocks in the storage, which is equal to the number of real data blocks that should be cached by the client. We first introduce the notation of node state and its transitions. Then, we analyze the probability to introduce a new dummy block during every data eviction process. Finally, we show that, with appropriate setting of system parameters (i.e., $\lambda \geq 2$ and $s = 2c\lambda \log N$ for $c \geq 1$), the number of dummy data blocks is bounded by $x \log N$ with a probability greater than $1 - (\frac{1}{N})^{2x}$.

According to the eviction algorithm in SE-ORAM, a new dummy block may be introduced only when a data block is evicted from a node that is full and does not contain any dummy block. For any of such node, we use $(x, s-x)$ to represent its state, where x is the number of data blocks in the left group and $s-x$ is the number of data blocks in the right group. Thus, the state transition probabilities are as follows:

$$Pr[(x+1, s-x-1)|(x, s-x)] = \begin{cases} \frac{1-p}{2}, & \text{if } x < s-x, \\ \frac{1}{4}, & \text{if } x = s-x, \\ \frac{p}{2}, & \text{if } x > s-x, \end{cases} \quad (3)$$

$$Pr[(x-1, s-x+1)|(x, s-x)] = \begin{cases} \frac{p}{2}, & \text{if } x < s-x, \\ \frac{1}{4}, & \text{if } x = s-x, \\ \frac{1-p}{2}, & \text{if } x > s-x. \end{cases} \quad (4)$$

$Pr[(x-1, s-x+1)|(x, s-x)]$ can be computed similarly. We skip it due to space limit. Figure 2 shows the complete set of node state transitions. Based on the above analysis, we can get the following lemma.

Lemma 3: In SE-ORAM, any node on the storage tree has a probability less than $2^{-\frac{s}{2\lambda}}$ to stay in state $(0, s)$ (or state $(s, 0)$); that is, $Pr[(0, s)] = Pr[(s, 0)] < 2^{-\frac{s}{2\lambda}}$.

Lemma 4: In SE-ORAM, when $s = 2c\lambda \log N$, the probability for an eviction process to introduce a new dummy block is less than $\frac{\log N}{2N^c}$.

Theorem 2: In SE-ORAM, when $\rho \geq \frac{1}{2}$ and $s = 2c\lambda \log N$ for $c \geq 1$,

$$Pr[\text{number of dummy blocks } \leq x]$$

$$> 1 - (\frac{1}{2c\lambda N^{c-1}})^x \geq 1 - (\frac{1}{2\lambda})^x.$$

The proofs of the above lemmas and theorems are presented in Appendix 3, 4 and 5. As the number of data blocks cached by the client is the same as the number of dummy blocks stored at the server side, we have the following corollary based on Theorem 2.

Corollary 3: In SE-ORAM, when $\rho \geq \frac{1}{2}$ and $s = 4 \log N$, the number of data blocks cached at the client side is bounded by $x \log N$ with a probability of $1 - (\frac{1}{N})^{2x}$.

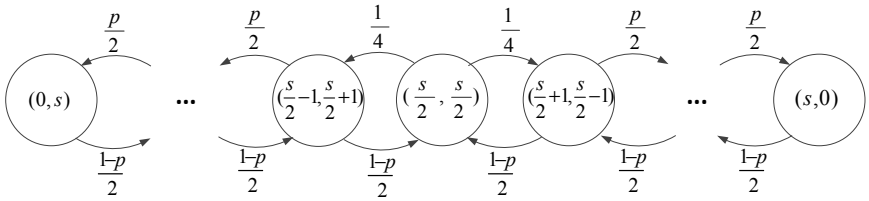

Figure 2. Node State and Transition

B. Communication Overhead

Each query or eviction process needs to access a series of nodes along a path from the root to a leaf. The communication and computational costs for query and eviction are therefore affected by the height of the storage tree.

Theorem 4: In SE-ORAM, the height of the storage tree is upper-bounded by $\log(N/s) + 3$ with a probability of at least $1-(\frac{1}{2})^s$. When $s = 2c\lambda \log N$, the probability is $1 - (\frac{1}{N})^{2c\lambda}$.

Proof: In Appendix 6. ■

For each query, all blocks on a path containing the target data block need to be downloaded and then uploaded; and in the following eviction process, all the data blocks on a randomly-selected path need to be downloaded and then uploaded. The number of nodes on each root-to-leaf path is $O(\log N)$ and each node stores $O(\log N)$ data blocks. Hence, the communication overhead is $O(log^2 N)$ data blocks per query.

C. Performance Comparison

We instantiate the generic SE-ORAM by setting $\lambda = 2$, $c = 1$ and thus $s = 4 \log N$, and compare the instantiated SE-ORAM with several state-of-the-art ORAMs including T-ORAM [12], G-ORAM [16], Path ORAM [13], SCORAM [18], and P-PIR [19], in terms of storage and communication overheads.

Table I compares SE-ORAM with state-of-the-art ORAM constructions in terms of the client and server storage overheads as well as the communication overhead per query. As we can see, SE-ORAM does not consume any extra storage in the server other than $N \cdot B$ bits for the N data blocks. On the contrary, the server storage overhead of each of the state-of-the-art ORAM constructions is $O(N \cdot B)$ or $(N \log N \cdot B)$ bits. Though the communication cost of SE-ORAM is on the same level as T-ORAM, as discussed in Section 1, it can be reduced to $O(\log N \cdot B)$ by adopting the additive Homomorphic encryption-based PIR primitives [19], similar to the way that P-PIR reduced the communication cost of T-ORAM from $O(\log^2 N \cdot B)$ to $O(\log N \cdot B)$.

VII. Conclusion

In this paper, we introduce a generic security definition for ORAM constructions, which allows a client to configure a desired security level and manage the tradeoffs between security and performance. We also propose SE-ORAM, a generic storage-efficient ORAM construction with configurable security parameter λ. The results of extensive analysis show that, SE-ORAM achieves the configured level of security, introduces zero storage overhead to the storage server (i.e., the storage server only storages N data blocks), and incurs $O(\log N)$ blocks storage overhead at the client, as long as $\lambda \geq 2$ and each node on the storage tree stores $4 \log N$ or more data blocks.

References

[1] M. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: ramification, attack and mitigation," in *Proc. NDSS*, 2012.

[2] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious RAMs," *Journal of the ACM*, vol. 43, no. 3, May 1996.

[3] M. T. Goodrich and M. Mitzenmacher, "Mapreduce parallel cuckoo hashing and oblivious RAM simulations," in *Proc. CoRR*, 2010.

[4] M. T. Goodrich, M. Mitzenmacher, O. Ohrimenko, and R. Tamassia, "Privacy-preserving group data access via stateless oblivious RAM simulation," in *Proc. SODA*, 2012.

[5] M. T. Goodrich and M. Mitzenmacher, "Privacy-preserving access of outsourced data via oblivious RAM simulation," in *Proc. ICALP*, 2011.

[6] M. T. Goodrich, M. Mitzenmacher, O. Ohrimenko, and R. Tamassia, "Oblivious RAM simulation with efficient worst-case access overhead," in *Proc. CCSW*, 2011.

[7] E. Kushilevitz, S. Lu, and R. Ostrovsky, "On the (in)security of hash-based oblivious RAM and a new balancing scheme," in *Proc. SODA*, 2012.

[8] B. Pinkas and T. Reinman, "Oblivious RAM revisited," in *Proc. CRYPTO*, 2010.

Table I

STORAGE AND COMMUNICATION OVERHEADS. N: NUMBER OF DATA BLOCKS; B: BLOCK SIZE IN BITS. SERVER STORAGE OVERHEAD IS DEFINED AS THE SERVER'S STORAGE CONSUMPTION OTHER THAN THE NB BITS FOR THE REAL DATA BLOCKS. CLIENT STORAGE OVERHEAD IS DEFINED AS THE CLIENT'S STORAGE CONSUMPTION OTHER THAN THE INDEX TABLE FOR THE EXPORTED DATA BLOCKS.

ORAM	Client Storage Overhead	Server Storage Overhead	Communication Overhead
T-ORAM [12]	$O(B)$	$O(N \log N \cdot B)$	$(\log^2 N \cdot B)$
G-ORAM [16]	$O(\log^2 N \cdot B)$	$O(N \cdot B)$	$(\frac{\log^2 N}{\log \log N} \cdot B)$
Path ORAM [13], SCORAM [18]	$O(\log N \cdot B)$	$O(N \cdot B)$	$O(\log N \cdot B)$
P-PIR [19]	$O(B)$	$O(N \log N \cdot B)$	$O(\log N \cdot B)$
SE-ORAM	$O(\log N \cdot B)$	0	$O(\log^2 N \cdot B)$

[9] P. Williams, R. Sion, and B. Carbunar, "Building castles out of mud: practical access pattern privacy and correctness on untrusted storage," in *Proc. CCS*, 2008.

[10] P. Williams, R. Sion, and A. Tomescu, "PrivateFS: a parallel oblivious file system," in *Proc. CCS*, 2012.

[11] P. Williams and R. Sion, "Single round access privacy on outsourced storage," in *Proc. CCS*, 2012.

[12] E. Shi, T.-H. H. Chan, E. Stefanov, and M. Li, "Oblivious RAM with $O((\log N)^3)$ worst-case cost," in *Proc. ASIACRYPT*, 2011.

[13] E. Stefanov, M. V. Dijk, E. Shi, C. Fletcher, L. Ren, X. Yu, and S. Devadas, "Path ORAM: an extremely simple oblivious RAM protocol," in *Proc. CCS*, 2013.

[14] E. Stefanov and E. Shi, "ObliviStore: high performance oblivious cloud storage," in *Proc. S&P*, 2013.

[15] E. Stefanov, E. Shi, and D. Song, "Towards practical oblivious RAM," in *Proc. NDSS*, 2011.

[16] C. Gentry, K. Goldman, S. Halevi, C. Julta, M. Raykova, and D. Wichs, "Optimizing ORAM and using it efficiently for secure computation," in *Proc. PETS*, 2013.

[17] E. Stefanov and E. Shi, "Multi-cloud oblivious storage," in *Proc. CCS*, 2013.

[18] X. S. Wang, Y. Huang, T.-H. H. Chan, A. Shelat, and E. Shi, "SCORAM: oblivious RAM for secure computation," in *Proc. CCS*, 2014.

[19] T. Mayberry, E.-O. Blass, and A. H. Chan, "Efficient private file retrieval by combining ORAM and PIR," in *Proc. NDSS*, 2014.

[20] A. Freier, P. Karlton, and P. Kocher, "The secure sockets layer (SSL) protocol version 3.0," in *RFC 6101*, 2011.

APPENDIX 1: EVICTION EXAMPLES

In Figure 3, an example involving evictions occurring on four layers of the storage tree is shown, to further illustrate the eviction process.

APPENDIX 2: PROOF OF THEOREM 1

Proof: Since

$$Pr[\vec{T}_n \mid \vec{A}_n] = \frac{Pr[\vec{A}_n|\vec{T}_n]Pr[\vec{T}_n]}{\sum_{\forall \vec{x}_n \in \mathcal{X}_n} Pr[\vec{A}_n|\vec{x}_n]Pr[\vec{x}_n]}, \quad (5)$$

we need to compute $Pr[\vec{T}_n]$, $Pr[\vec{A}_n|\vec{T}_n]$, $Pr[\vec{x}_n]$ and $Pr[\vec{A}_n|\vec{x}_n]$.

First, as the server has no a prior knowledge of the client's actual data access pattern, for $\vec{T}_n$ and any $\vec{x}_n \in \mathcal{X}_n$, it holds that

$$Pr[\vec{T}_n] = Pr[\vec{x}_n] = \frac{1}{N^n}. \quad (6)$$

Second, due to Lemmas 1 and 2 and $Pr[\vec{A}_n|\vec{x}_n]$ being equal to $\prod_{i=1}^{n} Pr[\ q_i \mid \vec{x}_n;\ q_1,\ e_1,\ \cdots,\ q_{i-1},\ e_{i-1}\ ] \cdot \prod_{i=1}^{n} \prod_{j=1}^{h_i} Pr[\ e_{i,j} \mid \vec{x}_n;\ q_1,\ e_1,\ \cdots,\ q_i,\ e_{i,1},\ \cdots,\ e_{i,j-1}]$, it follows that

$$\left(\frac{2s}{N}\right)^n \cdot p^{\sum_{i=1}^{n} h_i} \leq Pr[\vec{A}_n|\vec{x}_n] \leq \left(\frac{2s}{N}\right)^n \cdot (1-p)^{\sum_{i=1}^{n} h_i}. \quad (7)$$

Hence,

$$\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} h_i} \leq \frac{Pr[\vec{A}_n|\vec{T}_n]}{Pr[\vec{A}_n|\vec{x}_n]} \leq \left(\frac{1-p}{p}\right)^{\sum_{i=1}^{n} h_i}. \quad (8)$$

Since $h_i \leq \log(N/s) < \log N$,

$$\left(\frac{p}{1-p}\right)^{n \log N} \leq \frac{Pr[\vec{A}_n|\vec{T}_n]}{Pr[\vec{A}_n|\vec{x}_n]} \leq \left(\frac{1-p}{p}\right)^{n \log N}. \quad (9)$$

Based on Equations (5), (6) and (9), it holds that

$$\left(\frac{p}{2(1-p)}\right)^{n \log N} \leq Pr[\vec{T}_n|\vec{A}_n] \leq \left(\frac{1-p}{2p}\right)^{n \log N}. \quad (10)$$

As $p = \frac{1}{2^{1/\lambda}+1}$, Equation (10) becomes Equation (2), which completes the proof. ■

APPENDIX 3: PROOF OF LEMMA 3

Proof: In the Markov chain of node state transition shown in Figure 2, the steady state distribution has the following property:

$$\begin{aligned} \frac{1}{2} &> Pr[(\frac{s}{2}-1, \frac{s}{2}+1)] \quad (11) \\ &= Pr[(0,s)] \cdot (\frac{1-p}{p})^{\frac{s}{2}-1} \\ &= Pr[(0,s)] \cdot 2^{\frac{s-2}{2\lambda}}. \end{aligned}$$

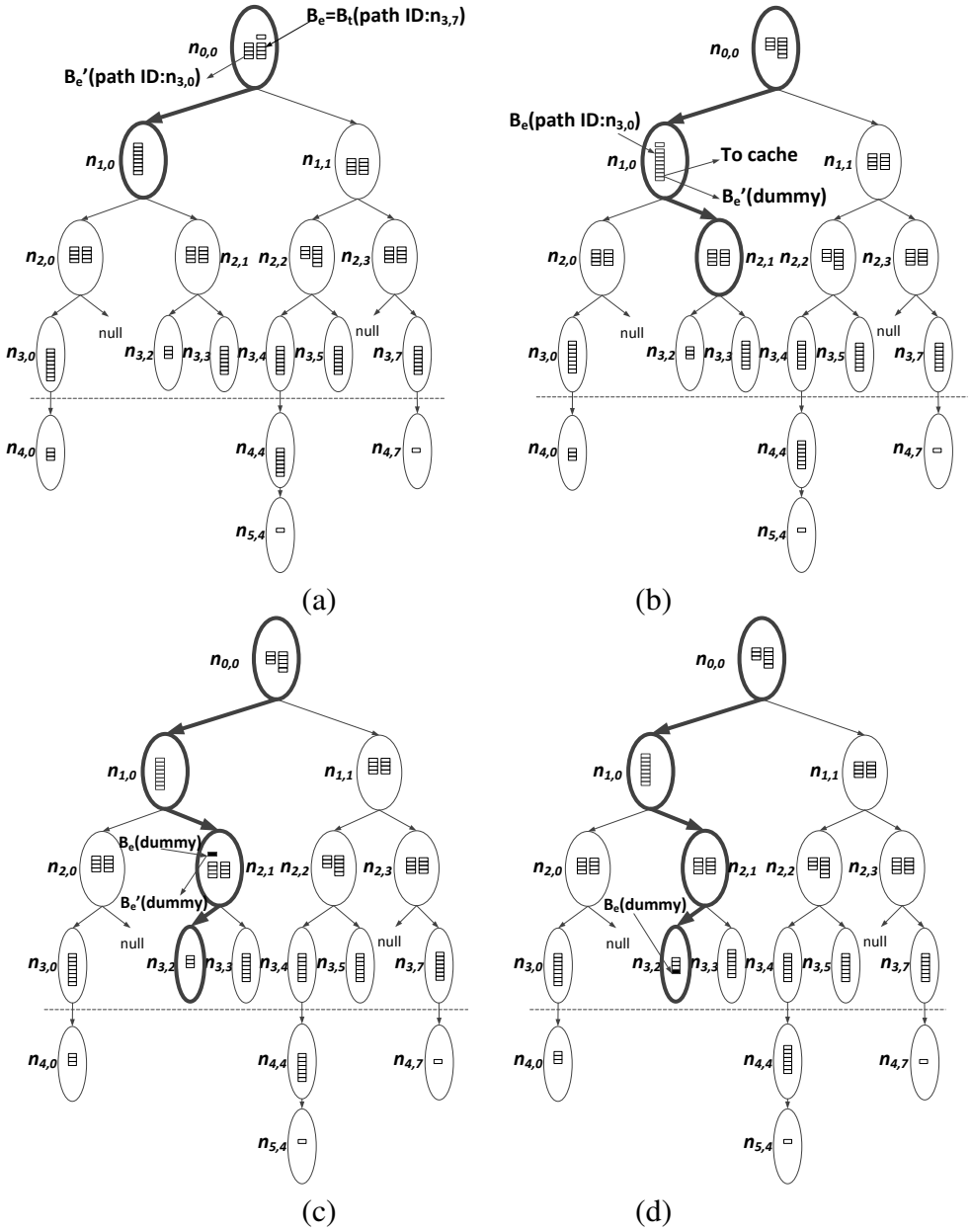

Figure 3. Data eviction example. (a) *Evicting data block from layer* 0 *to* 1: Evicted block B_e has path ID $n_{3,7}$ but the evicting node $n_{0,0}$ chooses to evict to left. Hence, B_e is written obliviously to $n_{0,0}$, while a block with path ID $n_{3,0}$ (and therefore evictable to left) is selected from $n_{0,0}$ to the new evicted block. This is Case II in Section 4.3 E4. (b) *Evicting data block from layer* 1 *to* 2: Evicting node $n_{1,0}$ chooses to evict to right. No block (including B_e and the blocks in $n_{1,0}$) is evictable to right. Hence, a dummy block is created to replace a randomly-selected real block in $n_{1,0}$, which is moved to the cache of the client. The dummy block then becomes the new evicted block. This is Case IV in Section 4.3 E4. (c) *Evicting data blocks from layer* 2 *to* 3: As the evicted block B_e is a dummy, it remains as the evicted block no matter whether the evicting node chooses to evict to left or right. This is Case I in Section 4.3 E4. (d) *Evicting data blocks from layer* 3 *to* 4: The evicting node is non-full. So the evicted block is written to it obliviously and the eviction process terminates, as explained in Section 4.3 E4.

Also because $\lambda > 1$, it follows that $Pr[(0,s)] < 2^{-\frac{s-2}{2\lambda}} \cdot \frac{1}{2} < 2^{-\frac{s}{2\lambda}}$. Similarly, it can be proved that $Pr[(s,0)] < 2^{-\frac{s}{2\lambda}}$. ∎

APPENDIX 4: PROOF OF LEMMA 4

Proof: According to Lemma 3, for any node, $Pr[(0,s)] = Pr[(s,0)] < 2^{-\frac{s}{2\lambda}}$, which is less than $\frac{1}{N^c}$ since $s = 2c\lambda \log N$.

During an eviction process, at most one dummy data block may be introduced. And the introduction occurs only if: there is at least one evicting node that is in state $(0,s)$ (or $(s,0)$), block evicted to this node is evictable only to right (or left), and the node chooses to evict to left (or right). For this to occur, the probability is at most

$$1 - \big(1 - (Pr[(0,s)] + Pr[(s,0)]) \cdot \frac{p}{2}\big)^{h+1},$$

which is less than $1 - (1 - \frac{1}{2N^c})^{\log N}$ since $p = \frac{1}{1+2^{1/\lambda}} < \frac{1}{2}$ and $h + 1 = \log(N/s) + 1 < \log N$. Expanding it, we obtain

$$1 - 1 + \log N \cdot \frac{1}{2N^c}$$

$$- \sum_{i=1}^{\infty} [\binom{\log N}{2i} (\frac{1}{2N^c})^{2i} - \binom{\log N}{2i+1} (\frac{1}{2N^c})^{2i+1}],$$

which is less than $\frac{\log N}{2N^c}$, since $\binom{\log N}{2i}(\frac{1}{2N^c})^{2i} > \binom{\log N}{2i+1}(\frac{1}{2N^c})^{2i+1}$ for every i. Hence, the Lemma is proved. ∎

APPENDIX 5: PROOF OF THEOREM 2

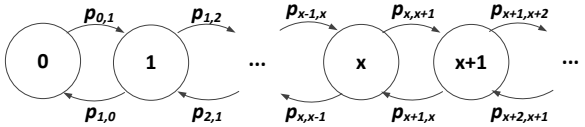

Figure 4. State Transition of Dummy Block Number N_d

Proof: Let N_d denote the number of dummy blocks in the storage server when an eviction process just finishes. Figure 4 depicts the Markov Chain of the transition of N_d value.

The transition from $N_d = x$ to $N_d = x+1$ occurs after one eviction process iff: during the eviction process and the immediately-preceding query process, (i) a new dummy block is introduced, and (ii) no dummy block is removed. The probability that (i) occurs is at most $\frac{\log N}{2N^c}$. Hence, $p_{x,x+1} = Pr[N_d = x+1 | N_d = x] \leq \frac{\log N}{2N^c}$.

The transition from $N_d = x+1$ to $N_d = x$ occurs after an eviction process iff: during the eviction process and the immediately-preceding query process, (i) no new dummy block is introduced, and (ii) an existing dummy block is removed. The probability for (i) to occur is at least $1 - \frac{\log N}{2N^c}$. The probability for (ii) to occur is $\frac{\rho}{1+\rho} \cdot \frac{2s}{N}$, where $\frac{\rho}{1+\rho}$ is the probability that this query-eviction round is an extra round, and $\frac{2s}{N}$ is the low bound of the probability that a path containing dummy block is queried and thus a dummy block can be removed. Also note that $1 - \frac{\log N}{2N^c} \geq \frac{3}{4}$ because $\log N \leq \frac{N}{2}$ for every $N \geq 4$; we do not consider $N < 4$ as it is trivial. Hence, $p_{x+1,x} = Pr[N_d = x | N_d = x+1] \geq \frac{3s\rho}{2(1+\rho)N}$.

Since $s = 2c\lambda \log N$ for $c \geq 1$, and $\rho \geq \frac{1}{2}$ and thus $(1+\rho)/\rho \leq 3$, $\frac{p_{x,x+1}}{p_{x+1,x}} < \frac{2(1+\rho)N \log N}{6s\rho N^c} \leq \frac{1}{2c\lambda N^{c-1}}$. So, $Pr[N_d > x] < (\frac{1}{2c\lambda N^{c-1}})^x \leq (\frac{1}{2\lambda})^x$; that is, $Pr[N_d \leq x] > 1 - (\frac{1}{2c\lambda N^{c-1}})^x \geq 1 - (\frac{1}{2\lambda})^x$. ∎

APPENDIX 6: PROOF THEOREM 4

Proof: In SE-ORAM, N data blocks are distributed to $\frac{N}{2s}$ paths uniformly at random. We first show that probability for a path to be assigned with more than $4s$ blocks is no greater than $\frac{1}{N^{4c}}$.

Assigning N blocks to $\frac{N}{2s}$ is a standard balls in bins game with N balls and $\frac{N}{2s}$ bins. The expected number of blocks assigned to each path is $2s$. According to Chernoff bound, the probability for any path to be assigned with more than $4s$ blocks is upper-bounded by $e^{-2s/3} \leq 2^{-s}$. That is, the probability is at least $1 - (\frac{1}{2})^s$ that every path is assigned with no more than $4s$ block.

A path has the longest length if all the blocks assigned to it have to be stored in its level-h node and supplementary nodes. In this extreme case, a path has a length of no longer than $\log(N/s) + 3$ if no more than $4s$ blocks are assigned to it.

So far, we have proved that the probability is at least $1 - (\frac{1}{2})^s$ that the height of the tree is no larger than $\log(N/s) + 3$. When $s = 2c\lambda \log N$, obviously the probability is $1 - (\frac{1}{N})^{2c\lambda}$. ∎