Tian et al. EURASIP Journal on Information Security (2020) 2020:7
https://doi.org/10.1186/s13635-020-00110-1

EURASIP Journal on Information Security

RESEARCH Open Access

Low-cost data partitioning and encrypted backup scheme for defending against co-resident attacks

Junfeng Tian, Zilong Wang* and Zhen Li

Abstract

Aiming at preventing user data leakage and the damage that is caused by co-resident attacks in the cloud environment, a data partitioning and encryption backup (P&XE) scheme is proposed. After the data have been divided into blocks, the data are backed up using the XOR operation between the data. Then, the backup data are encrypted using a random string. Compared with the existing scheme, the proposed scheme resolves the conflict between data security and survivability via encrypted backup. At the same time, because the XOR-encrypted backup causes multiple data blocks to share the same backup data, the storage overhead of the user is reduced. In this paper, existing probabilistic models are used to compare the performances of an existing scheme and the P&XE scheme in terms of data security, data survivability and user storage overhead, and the overall performances of the two schemes in terms of these three aspects that are compared using control variables. Finally, the experimental results demonstrate the effectiveness of the P&XE scheme at improving user data security and survivability and reducing user storage overhead.

Keywords: Cloud computing, Co-resident attack, Data partition, Encrypted backup, Data theft, Data corruption

1 Introduction

Cloud computing provides users with various computing resources and storage resources in an on-demand and ubiquitous manner through the network, thereby substantially reducing users' computing and storage overhead [1–3]. Virtualization technology is an important part of cloud computing. To effectively utilize physical resources, a cloud service provider typically allocates multiple virtual machines of various tenants to the same physical machine, which is called the co-resident of the virtual machine [4]. Despite the logical isolation of a VM from its underlying hardware and from other VMs that are hosted on the same server, the co-resident architecture can be exploited by attackers, thereby exposing the cloud environment to a huge potential threat ([5–7]; Wei [8]; Wei [9]). For example, when an attacker co-resides with its target virtual machine, it can bypass the logical isolation to illegally access (steal) or destroy user data. In the relevant literature, the probabilities that data cannot be stolen and cannot be corrupted are called the data security and the survivability, respectively [10–12].

In recent years, research on resisting co-resident attacks has yielded fruitful results.

The most straightforward solution for resisting co-resident attacks is to eliminate the side channel [13]. There are also studies [14–17] that demonstrate the vulnerability of virtual machine monitors. Once an attacker controls a virtual machine monitor, all virtual machines that are running on the same physical machine will face significant security risks. Therefore, a mechanism that is based on removing virtual machine monitors for defending against such attacks was proposed in [18]. However, the above solution requires the modification or even redesign of the existing system architecture. Based on Intel cache allocation technology, a mitigation mechanism

* Correspondence: wangzilongx@163.com
Computer Science and Technology, School of Cyber Security and Computer, Hebei University, Baoding, China

was proposed in [19] for defending against co-resident attacks on the last-level cache in cloud servers in multi-core processors. Another mechanism for mitigating co-resident attacks by detecting abnormal behavior based on system components (CPU, cache, etc.) was proposed in [20, 21]. The HomeAlone defense mechanism, which was designed by [22], identifies malicious co-residents via the analysis of the side channel. Network flow watermarking technology was introduced in [23] to mitigate co-resident attacks by detecting malicious virtual machines in the same network. Based on LLC access collision, a covert channel communication method was proposed in [24] for virtual machine co-resident detection. A defensive mechanism, namely, virtual private cloud (VPC), was introduced in [25] for mitigating co-resident attacks in the Amazon Elastic cloud. Then, [26] further evaluated the performance of VPC technology. The virtual machine allocation strategy was first proposed in [27] for mitigating co-resident attacks by increasing the difficulty of the attacker co-residing with the target. Then, a new virtual machine placement strategy, namely, PSSF, was proposed in [28, 29] for increasing the security of virtual machines by prioritizing the physical machines that are used or in use by users to increase the difficulty of malicious users co-residing with their targets. In addition, a game-theory-based approach was used in [30] and ([31, 32];) to increase co-resident difficulties, thereby reducing the probability of co-resident attacks.

The prior works that address co-resident attacks have mostly focused on addressing side channels or VM allocations, which typically requires the modification of the existing cloud system architecture or assistance from the cloud service provider. In [33, 34], a new solution to the problem of co-resident attacks in the cloud environment was proposed. From the perspective of user's original requests based on the data partition technique, a user's information is divided into multiple separate data blocks. Each of these blocks is handled by a separate VM. For cases where data can be useful only in its integrity [35, 36], data partitioning has been used as an effective method for protecting sensitive information in the cloud. For example, the stripping method using data partitioning and image analysis was proposed in [36, 37] for protecting image data with sensitive information in the cloud. In [38], data partitioning techniques were used in conjunction with remote backup algorithms to enhance the security of data that are stored on cloud servers. Data partitioning techniques were first introduced in [33] for solving the problem of co-resident attacks. In addition, this article identifies the best data partitioning strategy (the optimal number of user VMs) for mitigating the effects of co-residence attacks.

Data partitioning technology can effectively improve the security of data: unless the attacker can access all the independent data blocks, the complete information cannot be obtained. However, data partitioning reduces the survivability of the data because any data block corruption will destroy the integrity of the information, thereby rendering the data unusable. To improve the survivability of data, users can create a replica of each block [39]; however, copying the data will increase the probability of the data being stolen. The trade-off between data security and data survivability under traditional information systems was studied in [35] without considering co-resident attacks. Then, [40] modeled the effects of co-resident attacks on the data partitioning and replication backup schemes. The partition and replica backup (P&R) scheme has been proposed for determining the

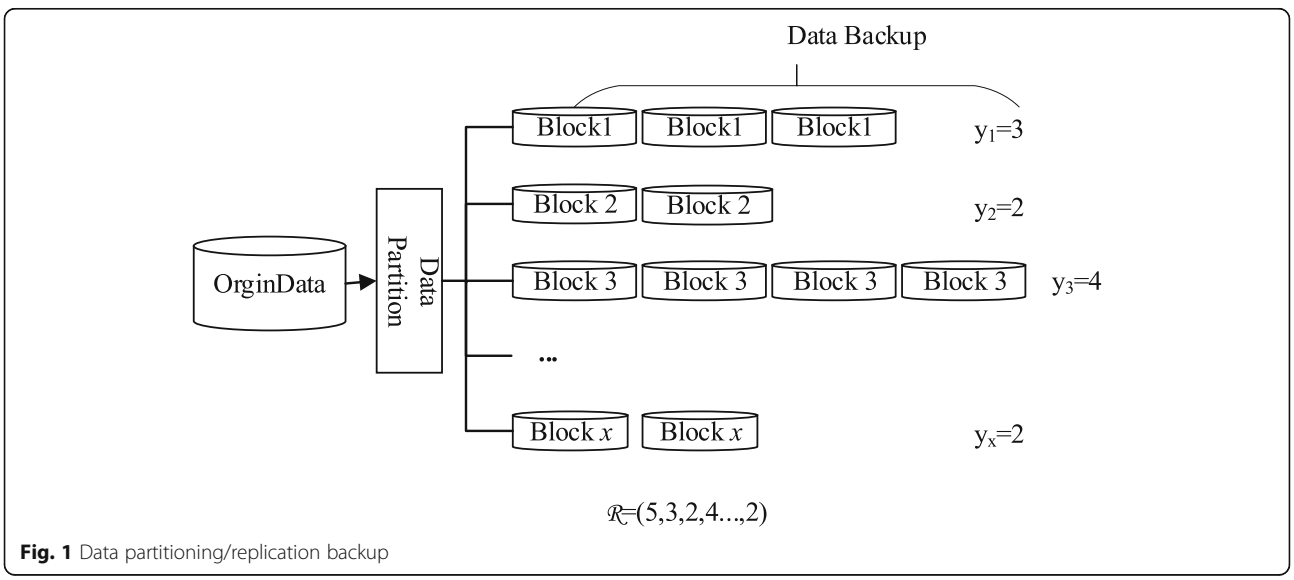

Fig. 1 Data partitioning/replication backup

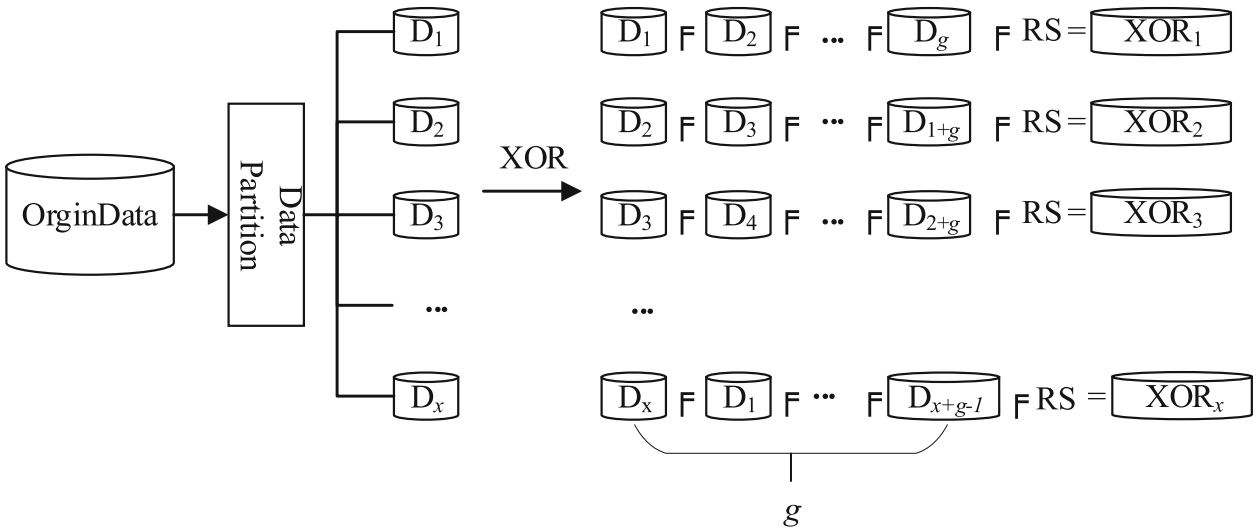

Fig. 2 P&XE scheme backup data generation

optimal partitioning and backup strategy against co-resident attacks and balances data security, data survivability, and user storage overhead. However, the P&R scheme imposes high storage overhead on users.

To reduce the user storage overhead and improve the data security and survivability, this paper proposes the partitioning and XOR-encrypted backup (P&XE) scheme. Via data partitioning and encrypted backup, the user's storage overhead is reduced, and the data security and data survivability are improved.

The remainder of the paper is organized as follows: Section 2 introduces the existing P&R scheme and attack model. Section 3 introduces the P&XE scheme. Section 4 presents the formulas for measuring data security and user storage overhead. The P&XE and P&R solutions are compared in terms of data security, data survivability, and user storage overhead in Section 5. Section 6 presents the conclusions of this work.

2 Existing scheme and attack models

2.1 Existing scheme

Users have sensitive information that must be protected. The attacker's actions may result in unauthorized access (stealing) of information and/or corrupted information, thereby rendering the information impossible for the user to use. To prevent information from being stolen

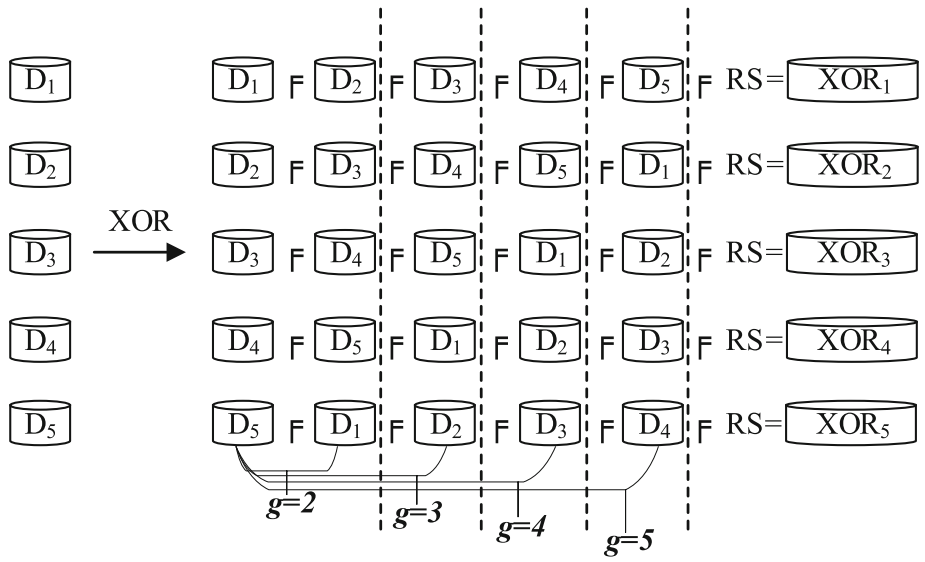

Fig. 3 XOR-encrypted backup data generation at $x = 5$

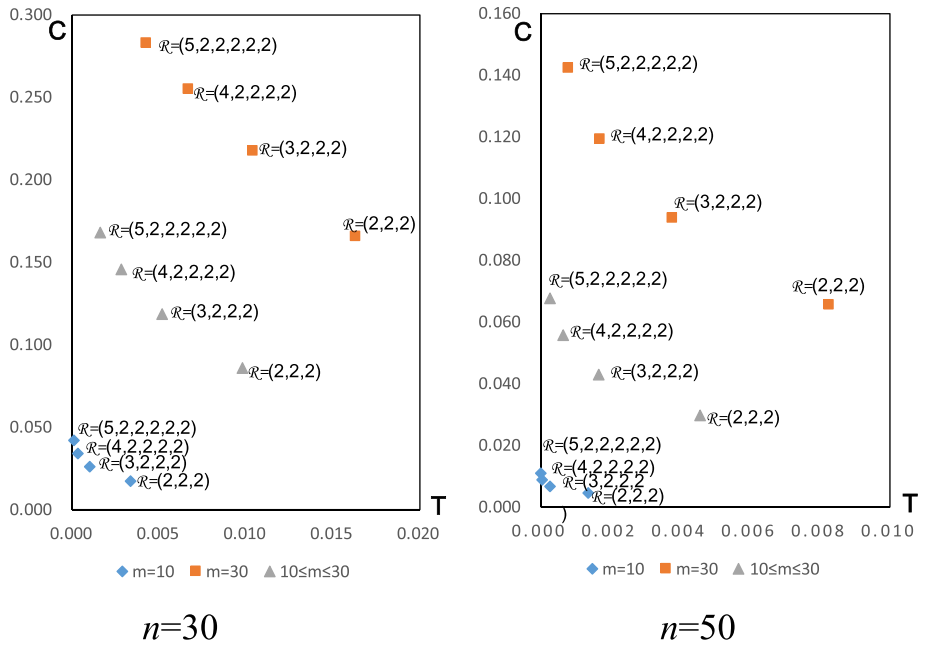

Fig. 4 Probabilities of data corruption and data theft under various values of *n*

(data security), the user divides it into x blocks of data (see Fig. 1), where $x > 1$ (the maximum number of blocks can be limited according to the scenario/demand). Unless the attacker has access to all x blocks of data, the data are safe. However, the data can only be used if its integrity is maintained. If any block of the information has been corrupted, the information integrity is lost and the user cannot use the information. To avoid this scenario, the user enhances the data survivability by creating y_i replicas of each data block i ($1 \le i \le 10$) (see Fig. 1). The data partitioning/replication scheme is denoted as $R = (x, y_1, \ldots, y_x)$, in which the user divides the data into x blocks. The number of replications of the i-th block of data is y_i.

To destroy the user's data, the attacker should destroy all y_i copies of any data block i. To steal information, an attacker must acquire at least one copy of any data block. Creating more blocks makes the information more difficult to steal but more vulnerable to corruption. Creating more copies for each block makes the data less susceptible to data corruption but makes the data more vulnerable to data theft. The optimal data blocking scheme should balance the security and survivability of the data.

Suppose there are n servers in the cloud computing system. After the user divides the data into separate blocks and creates multiple copies of these blocks (a total of k data blocks), the user sends k requests to the resource management system (RMS) to create a VM for each data block. The cloud resource management system (RMS) creates k users' user virtual machines (UVMs) and distributes these UVMs randomly to available physical servers. A server can obtain between 0 and k UVMs, and k UVMs can be distributed among between 1 and min (n, k) servers.

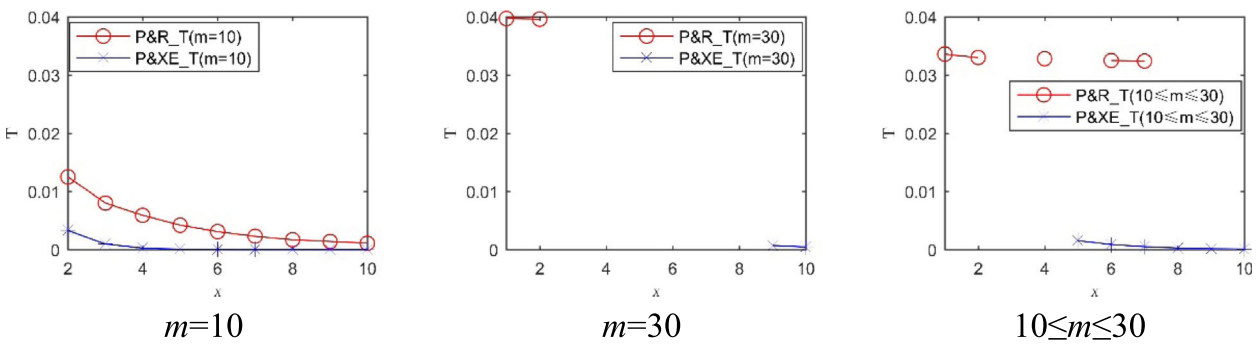

Fig. 5 Data security comparison under the same C^* and O^* constraints for $n = 30$

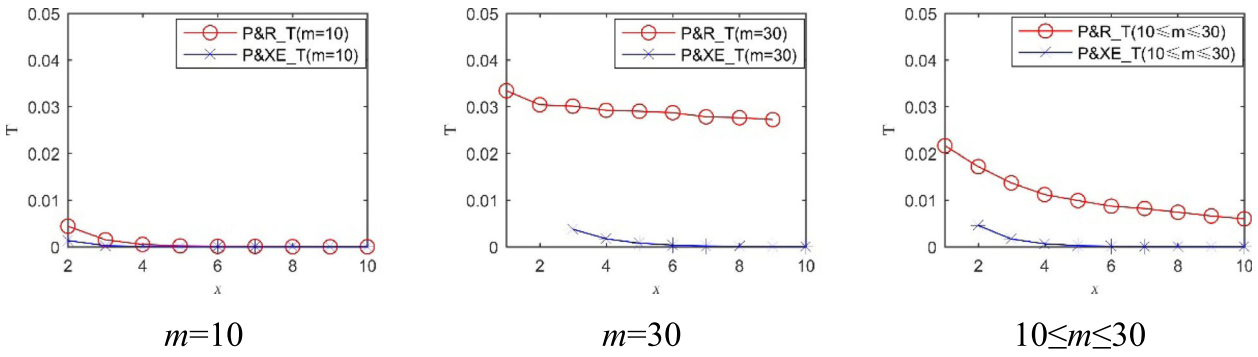

Fig. 6 Data security comparison under the same C^* and O^* constraints for $n = 50$

2.2 Attack Model

An attacker attempts to access a user's information to steal or destroy it. It is only possible to access the relevant data of the UVM if the attacker's virtual machine (AVM) is located on the same server as the UVM. To co-reside with the user's UVM, the attacker submits m requests to assign m AVMs to the same cloud system. The RMS creates an AVM for each request and randomly distributes it to n servers. If the AVM co-resides with UVMs on the same server, it can construct a side channel for each co-resident UVM and steal or destroy the data with a specified probability. Suppose the probability of an attacker stealing data is t, and the probability of corrupting data is c.

For convenience of discussion and without loss of generality, make the following assumptions:

1. The same data protection measures are used in all physical servers. The event that the attacker builds a side channel and steals or corrupts the data is the same for all servers where the AVM and the UVM are co-resident; hence, if the AVM successfully builds a side channel in one server, other AVMs that co-reside with the UVM can also successfully construct a side channel.
2. An attacker can steal data from all UVMs in the same server that are co-located with the AVM with probability t and damage the data with probability c.
3. The probabilities t and c do not depend on the numbers of UVMs and AVMs that co-reside in the same server.
4. The probabilities t and c are not necessarily equal. For example, if an attacker obtains encrypted data, the data cannot be decrypted and used; however, the data can be destroyed ($c > t$). Conversely, if the data are write-protected, stealing is easier than destroying ($t > c$).

To increase the difficulty of data theft, data partitioning technology is used to divide the data into multiple blocks, thereby improving the data security; however, this improvement also increases the probability of data corruption. To reduce the probability of data corruption, multiple copies are created of each data block to increase the difficulty of data corruption. Data partitioning and replication are in conflict between improving data security and the data survivability. Although increasing the numbers of blocks and data replications at the same time can improve the security and survivability of user data, it also imposes significant storage

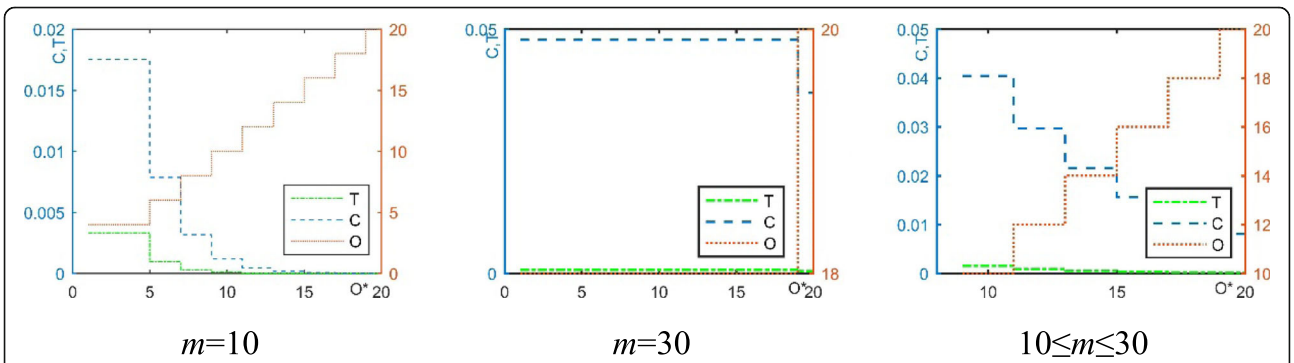

Fig. 7 Relationship among data security, data survivability, and user storage overhead under various O^* constraints for $n = 30$ and $C^* = 0.05$

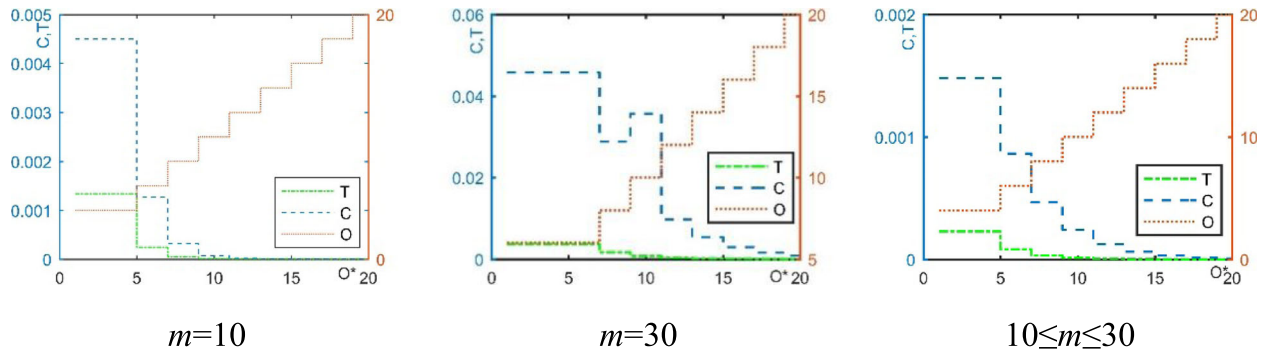

Fig. 8 Relationship among data security, data survivability, and user storage overhead under various O* constraints for n = 50 and C^* = 0.05

overhead on users. Via the P&XE scheme, this paper improves the security and survivability of data while reducing the user's storage overhead.

3 Partitioning and XOR-encrypted backup scheme

3.1 Backup data generation process

This section describes the process of generating backup data in the P&XE scheme (Fig. 2).

The P&XE scheme consists of two parts:

1. Original data partitioning. The user divides the data to be protected into x blocks via data partitioning technology and $D_{\text{origin}} = (D_1, D_2, \ldots, D_x)$.
2. XOR-encrypted backup of data blocks. The XOR-encrypted backup data are generated by XORing multiple blocks of data with a random string (RS) of the user. The number of data blocks that are used to generate the XOR-encrypted backup data is called the group size, which is denoted as g ($2 \le g \le x$).

The generation of the i-th XOR-encrypted backup data starts from data block D_i, which is XORed with the g-1 ($i < x + g$-1) block data behind it, and finally uses the RS to encrypt the backup data; the formula is as follows:

$$\begin{cases} XOR_i = D_i \oplus D_{i+1} \oplus \ldots \oplus D_{i+g-1} \oplus RS & i+g-1 \le x \\ XOR_i = D_i \oplus D_{i+1} \oplus \ldots \oplus D_{i+g-1-x} \oplus RS & i+g-1 > x \end{cases} \tag{1}$$

In the following, x = 5 is used as an example to illustrate the process of generating XOR-encrypted backup data.

According to Fig. 3, a change in the group size (g) only affects the number of times a data block appears in the operation of XOR-encrypted backup data and does not increase the number of XOR-encrypted backup data, namely, the number of UVMs that are used in the P&XE scheme depends only on the number of data blocks x; the number of UVMs that are used by the P&XE scheme is $2x$. This number is also why the P&XE scheme can maintain high security and low user storage overhead if the number of user data blocks is increased (see Section 5.4 for the analysis).

3.2 Data Recovery Process

Since the XOR operation satisfies the commutative law, namely, a ⊕ b = b ⊕ a, according to formula (1), there are g XOR-encrypted backup data that are related to D_i. When the i-th data block D_i is destroyed, one of the g XOR data is selected according to the formula for

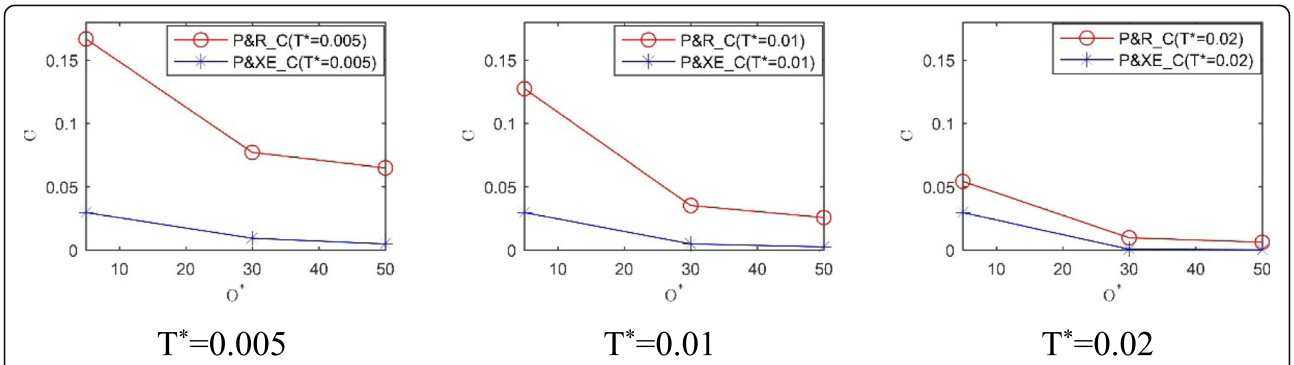

Fig. 9 Data survivability comparison under the same T^* and O^* constraints for n = 50 and 10 ≤ m ≤ 30

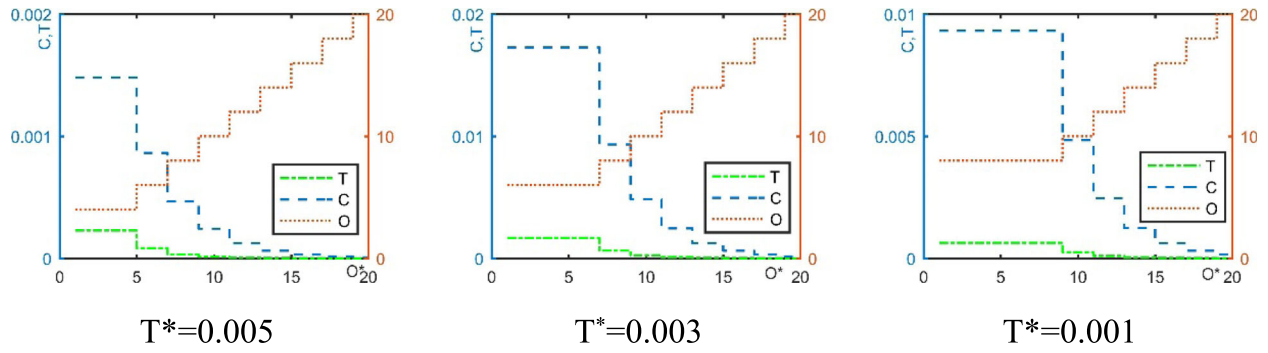

Fig. 10 Relationships among the data security, data survivability, and user storage overhead under various O^* and T^* constraints for $n = 50$ and $10 \le m \le 30$

generating the XOR-encrypted backup data. The XOR-encrypted backup data on both sides of the equation are converted to D_i, and the data D_i can be restored via the exclusive or operation. In the following, $x = 5$ is used as an example to demonstrate the data recovery process.

$$\begin{cases} XOR_1 = D_1 \oplus D_2 \oplus D_3 \oplus RS \\ XOR_2 = D_2 \oplus D_3 \oplus D_4 \oplus RS \\ XOR_3 = D_3 \oplus D_4 \oplus D_5 \oplus RS \\ XOR_4 = D_4 \oplus D_5 \oplus D_1 \oplus RS \\ XOR_5 = D_5 \oplus D_1 \oplus D_2 \oplus RS \end{cases}$$

If $g = 3$, all XOR-encrypted backup data are generated as above. Assume that data D_4 are damaged. According to the above formula, the XOR-encrypted backup data that are related to D_4 are XOR_2, XOR_3, and XOR_4 according to the properties of the exclusive or operation:

$$\begin{cases} D_4 = D_2 \oplus D_3 \oplus XOR_2 \oplus RS \\ D_4 = D_3 \oplus XOR_3 \oplus D_5 \oplus RS \\ D_4 = XOR_4 \oplus D_5 \oplus D_1 \oplus RS \end{cases}$$

D_4 can be obtained by swapping D_4 with XOR_2, XOR_3, or XOR_4 in the formulas and recovering D_4 by selecting one of the above equations for performing the XOR operation.

3.3 Theoretical analysis of data security and data survivability

This section analyzes the impact of the P&XE scheme on data security and data survivability and compares the impacts of P&R and P&XE on data security and data survivability.

According to the above, the security and survivability of user data are related to the number of blocks and the number of copies of the data, respectively. Consider the scheme R = (5,3,3,3,3,3) as an example, in which the data are divided into 5 blocks, each with 3 copies. In the P&R scheme [40], there are 3 copies of each block. When the attacker obtains data, at least one of the three blocks is obtained for each block, and the data can be successfully stolen. In the case of data corruption, the attacker simply destroys all copies of any data to successfully corrupt the data. For the P&XE scheme, R = (5,3,3,3,3,3) corresponds to $g = 2$ because for data corruption, and the attacker will destroy the original data and the two pieces

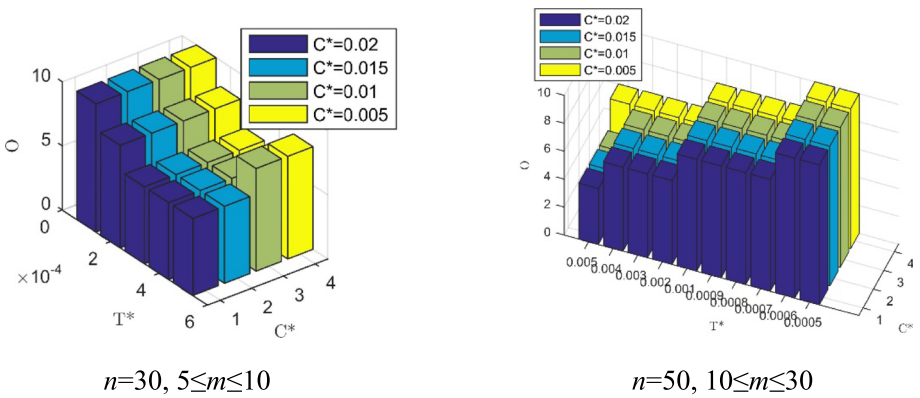

Fig. 11 User storage overhead under various values of T^* and C^*

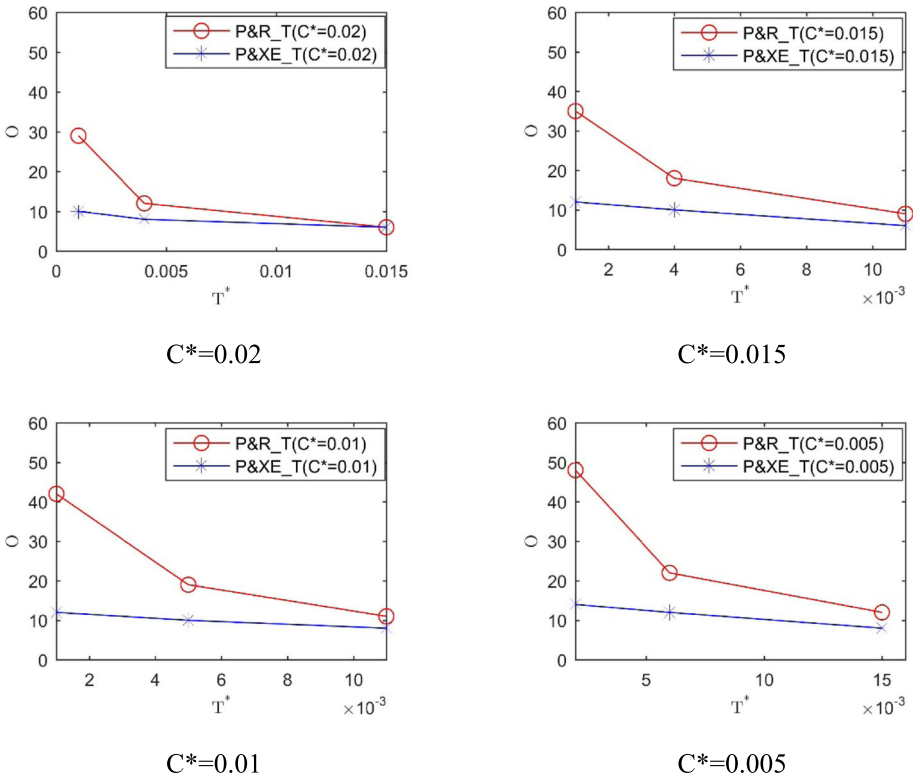

Fig. 12 User storage overhead comparison under the same T^* and C^* constraints for $n = 30$ and $5 \leq m \leq 10$

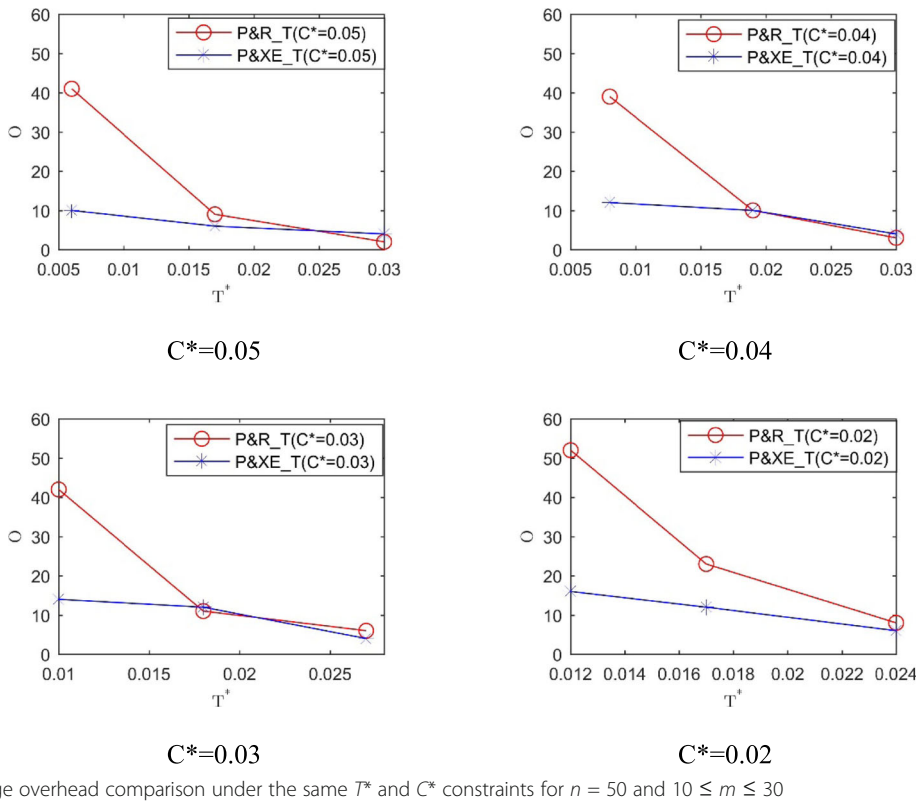

Fig. 13 User storage overhead comparison under the same T^* and C^* constraints for $n = 50$ and $10 \leq m \leq 30$

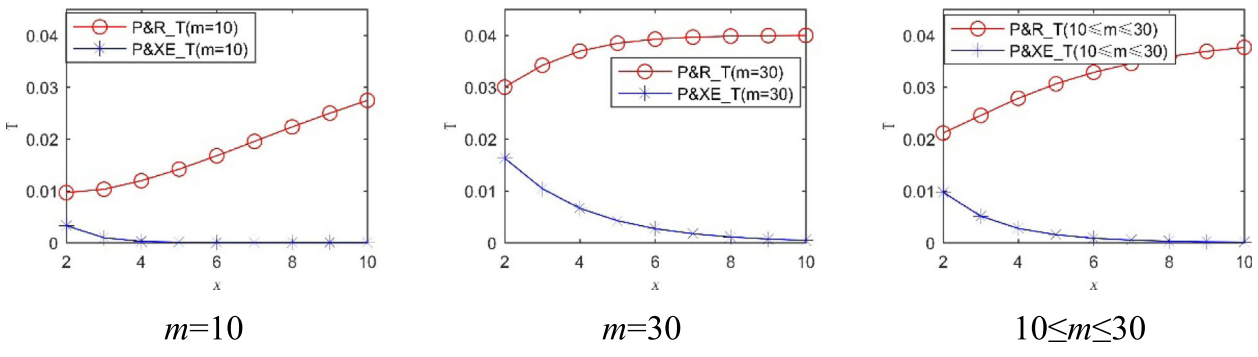

Fig. 14 Data security comparison under the same data survivability for n = 30

of XOR-encrypted backup data that are associated with the data. Therefore, the P&XE scheme has the same security as the P&R scheme for data corruption. However, for data theft, since the XOR-encrypted backup data are encrypted by the user random string (RS), the attacker cannot obtain other user data through the XOR-encrypted backup data; therefore, when addressing data theft, R = (5,1,1,1,1,1) and only when an attacker steals the original block can the data be stolen successfully. P&XE will outperform the P&R scheme in addressing data theft.

An attacker must steal the original data of each data block when stealing user data. If XOR-encrypted backup data are stolen, the attacker cannot use the data because it cannot crack the user's RS. It is not possible to obtain other data of the user from XOR-encrypted backup data. When the attacker destroys the data, not only the user's original data but also all XOR-encrypted backup data that are related to the original data must be destroyed. Therefore, the P&XE scheme can improve the survivability of user data without reducing the security of user data.

4 Probabilities of data Theft and data corruption

To measure the impacts of P&XE and P&R on data security, data survivability, and user storage overhead, the measurement formulas in [40] are used: T(R) is used for data security, W(R) for data survivability, and O(R) for user's storage overhead.

Consider the following scenario: there are n servers in the cloud environment, k UVMs, m AVMs, and data partitioning/replication scheme R = (x,y_1...,y_x). p (n,k,m) and w (n,k,m) are the probability that the attacker's AVM co-resides with all UVMs and the probability that the attacker's AVM co-resides with at least one UVM, respectively [40]. Then, if the number of AVMs is known, the probability of data being stolen is:

$$\mathrm{T}(\mathcal{R}, m) = t\left(\prod_{i=1}^{x} w(n, y_i, m)\right) \tag{2}$$

The probability of the data being corrupted is:

$$\mathrm{C}(\mathcal{R}, m) = c\left(1 - \prod_{i=1}^{x} 1 - p(n, y_i, m)\right) \tag{3}$$

When the value of m is uncertain, but the distribution form and range of m are known, μ (l) = Pr (m = l) and ($m_{min} \le l \le m_{max}$) and the probabilities of data theft and data corruption are:

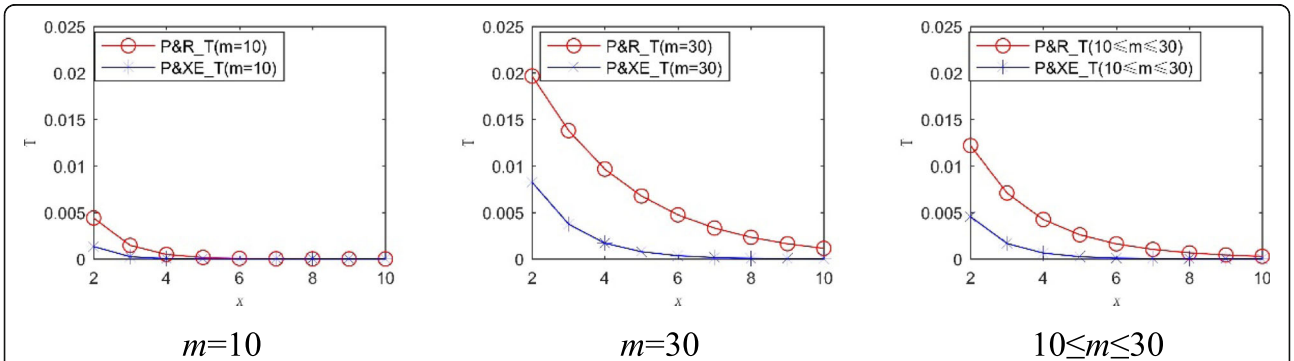

Fig. 15 Data security comparison under the same data survivability for n = 50

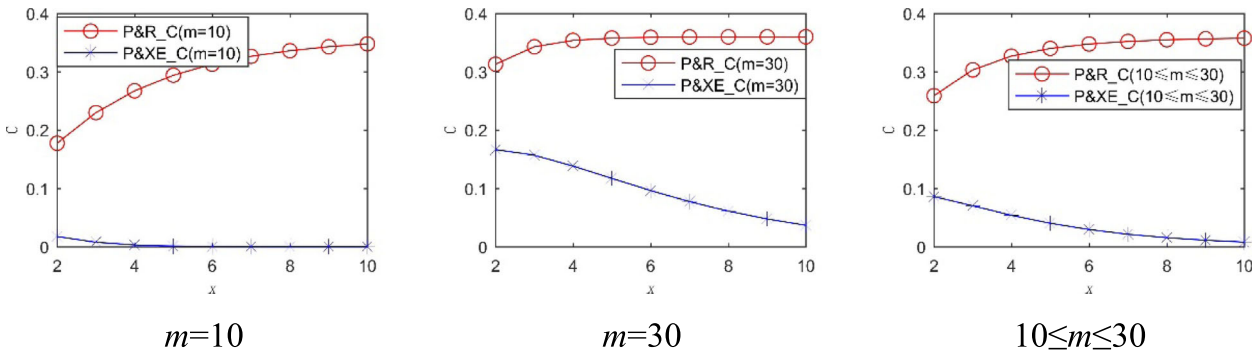

Fig. 16 Data survivability comparison under the same data security for $n = 30$

$$\mathrm{T}(\mathcal{R},\mu) = t\sum_{l=m_{\min}}^{m_{\max}} \mu(l)\prod_{i=1}^{x} w(n,y_i,l) \tag{4}$$

$$\mathrm{C}(\mathcal{R},\mu) = c\sum_{l=m_{\min}}^{m_{\max}} \mu(l)\left(1-\prod_{i=1}^{x} 1-p(n,y_i,l)\right) \tag{5}$$

Figure 4 shows the relationship between the data theft probability T (x-axis) and the data corruption probability C (y-axis) under various numbers of servers under the P&XE scheme, where $c = t = 1$; $n = 30$ or 50; and $m = 10$, $m = 30$, or $10 \le m \le 30$. According to Fig. 4, under the condition that the group size is the same, as the number of blocks increases, the probability of the user data being stolen is reduced, and the probability of the user data being damaged increases. At the same time, under the condition that the number of attackers' virtual machines is constant, increasing the number of physical machines can improve the security and survivability of user data.

The number of UVMs that are created by the user is $K(\mathcal{R}) = \sum_{i=1}^{x} y_i$. O_{vm} is the overhead associated with creating one VM, and the user's overhead that is associated with creating k UVMs is:

$$O(\mathcal{R}) = K(\mathcal{R})O_{vm} \tag{6}$$

5 Experimental comparison

In this section, the P&XE scheme and the P&R scheme are compared in terms of the probability of data theft (T), the probability of data corruption (C), and the user storage overhead (O). The comparison process considers the P&R scheme [40]. According to [40], set T^*, C^*, and O^* as the constraint of T, C, O separately as mentioned before. After defining the thresholds of two parameters, find the solution that optimizes the remaining parameters. Then, by controlling the variables, the overall performances of the two schemes on T, C, and O are compared. Finally, the feasibility of the P&XE scheme is evaluated in terms of the time cost of XOR (Since the P&XE scheme requires $x > 1$, there are no corresponding data in the experimental results for point $x = 1$).

5.1 Probability of data theft (T) comparison

Figure 5 compares the results with the optimal value of T under the P&R scheme with $n = 30$, $t = 0.2$, $C^* = 0.05$, and $c = 0.6$. The maximum number of blocks is $x_{\max} = 10$, and each block of data satisfies $y_{\max} = 10$. According to the figure, as m increases, using more UVMs and increasing the number of blocks can reduce the probability

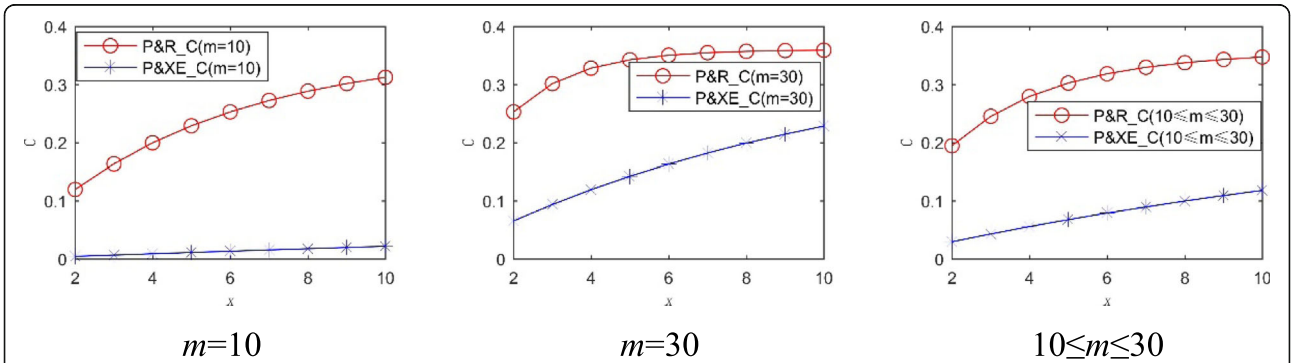

Fig. 17 Data survivability comparison under the same data security for $n = 50$

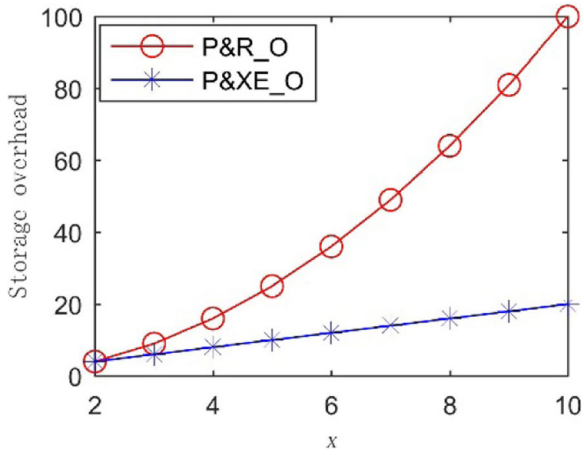

Fig. 18 Comparison of user storage overhead under the same data survivability

of user data being stolen because the increase in the number of blocks makes the attacker less likely to obtain complete data. Under the same number of AVMs, the more data blocks there are, the lower the probability that an AVM will co-reside with it and the lower the probability of data being stolen. If $m = 30$, the P&XE scheme identifies a scheme that satisfies $C < C^*$ when x is 9. As the number of data blocks increases, the probability of user data being stolen is reduced; hence, if x is 10, the scheme satisfies $C < C^*$.

Figure 6 compares the results with the optimal value of T for the P&R scheme with $n = 50$, $t = 0.2$, $C^* = 0.05$, and $c = 0.6$. Compared with Fig. 5, as the number of servers increases, the probability that the attacker's AVMs co-reside with the user's UVMs is reduced; hence, the probability of user data being stolen is reduced. At the same time, the probability of data being stolen decreases as the number of data blocks increases.

Table 1 Average data recovery times under various group sizes (unit:seconds)

Data					
Group size	x_i = 1 MB	x_i = 16 MB	x_i = 64 MB	x_i = 256 MB	x_i = 1 GB
2	0.026	0.24	1.04	4	14
3	0.039	0.36	1.56	6	21
4	0.052	0.48	2.08	8	28
5	0.065	0.6	2.6	10	35
6	0.078	0.72	3.12	12	42
7	0.091	0.84	3.64	14	49
8	0.104	0.96	4.16	16	56
9	0.117	1.08	4.68	18	63
10	0.13	1.2	5.2	20	70

According to Figs. 5 and 6, the P&XE scheme can effectively reduce the probability of user data being stolen because under the P&XE scheme, the probability of user data being stolen depends only on the number of blocks. Under the P&XE scheme, regardless of the group size, there is only one copy of each data block for the user. Therefore, only when a malicious user obtains all the original data of the user can the data be successfully stolen.

Figures 7 and 8 show the relationships among T, C, and O when $C^* = 0.05$, $t = 0.2$, $c = 0.6$, and $O_{vm} = 1$ in the P&XE scheme. According to Fig. 7, if $m = 30$, when the number of AVMs is large, increasing the number of blocks does not reduce the probability of data being stolen or the probability of data being corrupted. If the AVMs are distributed across all servers, any partitioning/backup strategy will fail. The probability of such an event occurring increases as n decreases or/and as m increases.

5.2 Probability of data corruption (*C*) comparison

Figure 9 compares the performances of P&XE and P&R on C under various T^* limits with $n = 50$, $t = 0.2$, and $c = 0.6$. The experimental results demonstrate that under the same T^* limit, the P&XE scheme makes the user data less likely to be destroyed and realizes higher security because the P&XE scheme can guarantee data security and data survivability at the same time. Due to the characteristics of the P&XE scheme, if the attacker corrupts the data, it must destroy the original data and all related XOR-encrypted backup data. However, since the XOR-encrypted backup data are encrypted by the user's random string, the attacker cannot decrypt the original data through XOR-encrypted backup data; therefore, when stealing data, the attacker must obtain all the original data. Hence, the P&XE scheme better protects the security of the user data.

Figure 10 shows the variations in T, C, and O at various values of T^* in the P&XE scheme when $t = 0.2$, $c = 0.6$, and $O_{vm} = 1$. With the relaxation of T^*, users can reduce the probability of data corruption by using more UVMs (increasing the number of data blocks or increasing the number of XOR-encrypted backups).

5.3 User storage cost (O) comparison

Figure 11 shows that in the P&XE scheme, under the same T^* constraint, as C^* decreases, users will use more UVMs to protect against data corruption. At this time, the increase of UVMs is due to the increase in the amount of XOR-encrypted backup data. Similarly, under the same conditions of C^*, as T^* decreases, users must also use more UVMs to prevent data theft. The increase in UVMs at this time is due to the increase in the number of blocks.

Figures 12 and 13 compare the user storage overhead between the P&XE and P&R schemes under various C^* and T^* constraints when t = 0.2, c = 0.6, and O_{vm} = 1. With the relaxation of T^*, users require fewer UVMs to satisfy the T^* requirements. In Fig. 13, when C^* = 0.05 and T^* = 0.03, the P&XE scheme uses more UVMs. This is because under the P&XE scheme, since the number of data blocks is at least 2, the number of generated XOR-encrypted backup data is 2, and the user's minimum overhead is 4. In contrast, in the P&R scheme, the data are not partitioned in this case, and only the replication backup is used; hence, the overhead is lower compared to the P&XE scheme.

To compare the overall performances of the P&XE and P&R schemes in terms of T, C, and O, in the following, the trends of T, C, and O under the two schemes and under the control of variables are compared.

5.4 Overall comparison

Figures 14 and 15 compare the security of data from two aspects: Fig. 14 shows the best performance R of C under the P&XE scheme (the case in which the group size is consistent with the number of blocks, namely, $x = g$) compared with T of the P&R scheme under the same strategy. According to Fig. 14, as the number of blocks increases, the probability of the P&R scheme data being stolen increases due to the increase in the number of blocks and in the number of copies of each block for the P&R scheme. As the probability of stealing any piece of data increases, the probability of an attacker obtaining the complete data increases. Under the P&XE scheme, the data security depends only on the number of blocks: the greater the number of blocks, the higher the security of the data.

Figure 15 shows R in the case in which the group size of the P&XE scheme is 2 and the change in T with the number of blocks. When the group size is determined (namely, for the P&R scheme, the number of copies of each piece of data is consistent), T of the P&R scheme decreases as the number of blocks increases. This is because the number of copies of each block of data is the same, the probability of obtaining a copy of any piece of data is the same, and the number of blocks to be acquired increases, thereby increasing the difficulty for attackers to obtain the full data. Therefore, the probability of an attacker stealing data is reduced.

Figure 16 compares C under the same scheme R of T (with the same T as the reference standard, namely, no backup after the data have been partitioned). Since there is no replication backup, there is only one block per data. As the number of blocks increases, the probability of an attacker destroying any block increases; therefore, as the number of blocks increases, the probability of data corruption under the P&R scheme increases. Under the P&XE scheme, XOR-encrypted backup data do not affect the probability of data being stolen. If n = 30 and C corresponds to the minimum ($x = g$) data, as the number of blocks increases, the XOR-encrypted backup data of each piece of data also increases; hence, the probability of data corruption decreases.

In Fig. 17, n = 50 and c is set to the maximum value (g = 2) for comparison. As the number of blocks increases, the probability of user data being corrupted under the P&XE scheme increases because the number of blocks increases; however, the number of XOR-encrypted backup data per block remains unchanged. The probability of the attacker destroying any block is unchanged, the number of data blocks is increased, and the possibility of destroying any block is increased; hence, the probability of user data being destroyed is increased.

Figure 18 compares the user storage overhead between the P&R scheme and the P&XE scheme in the same scenario R of C. The experiment selects the group size when $g = x$ (this is the case in which the data have the strongest survivability under the P&XE scheme, namely, C is minimal). In this case, the number of UVMs that are used by the P&R scheme is x^2, and the number of UVMs that are used by the P&XE scheme is $2x$. If $x > 2$, the overhead of the P&XE scheme is smaller than the overhead that is generated by the P&R scheme. According to the figure, as the number of blocks increases, the storage overhead of the P&R scheme increases sharply to realize the same data survivability, whereas that of the P&XE scheme increases relatively flatly.

5.5 Time overhead

According to Table 1, the data recovery time increases as the group size increases or/and as the data size increases, which accords with our expectations. If the number of data blocks is 10, the recovery time for 1 GB data is 70 s in the case of g = 10. This time is acceptable compared to the cost of purchasing more virtual machines for increased security.

6 Conclusions

As the most dangerous type of attack method in the cloud environment, co-resident attacks pose a substantial threat to user data. The P&XE scheme effectively reduces the storage overhead of users by increasing the security and survivability of user data through data partitioning and XOR backup. In the P&R scheme, increasing the survivability of data requires the maximization of the number of data blocks, which may reduce the data survivability. In contrast, increasing the survivability of data requires the maximization of the number of copies of each data block, which, in turn, reduces the data security. Maximizing the number of blocks and increasing the number of copies per block of data both increase the

user's storage overhead. The P&XE scheme compensates for the insufficiency of the P&R scheme for balancing data security and data survivability, thereby reducing the user's storage overhead. The experimental results demonstrate that the P&XE scheme reduces the user's overhead and improves the security and survivability of user data.

Abbreviations

P&R: Partition and replica backup; P&XE: Partitioning and XOR-encrypted backup; VPC: Virtual private cloud; VM: Virtual machine; RMS: Resource management system; UVM: User virtual machines; AVM: Attacker's virtual machine; RS: Random string

Acknowledgements

Not applicable.

Authors' contributions

ZLW carried out the data processing, design and implementation of the proposed algorithms, experiment setup, and results evaluation and drafted the manuscript. JFT contributed to the conception, experiment design, and evaluation of the proposed approach and results, and helped draft the manuscript. ZL provided oversight for data and experimentation and participated in the manuscript editing. The authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61802106).

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 24 September 2019 Accepted: 11 May 2020
Published online: 24 May 2020

References

1. J. Anselmi, D. Ardagna, M. Passacantando, Generalized Nash equilibria for SaaS/PaaS Clouds[J]. Eur. J. Oper. Res. **236**(1), 326–339 (2014)
2. T. Püschel, G. Schryen, D. Hristova, et al., Revenue management for cloud computing providers: decision models for service admission control under non-probabilistic uncertainty[J]. Eur. J. Oper. Res. **244**(2), 637–647 (2015)
3. Singh B , Dhawan S , Arora A , et al. A, View of cloud computing[J]. International Journal of Computers & Technology, 2013, 4(2b1):50-58.
4. R. Buyya, C.S. Yeo, S. Venugopal, et al., Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility[J]. Futur. Gener. Comput. Syst. **25**(6), 599–616 (2009)
5. M.M. Alani, Securing the cloud: Threats, attacks and mitigation techniques[J]. Journal of Advanced Computer Science & Technology **3**(2), 202 (2014)
6. Han Y. Defending against co-resident attacks in cloud computing[D]. , 2015.
7. G. Nalinipriya, P.J. Varalakshmi, K.G. Maheswari, et al., An extensive survey on co-resident attack in dynamic cloud computing environment[J]. Int. J. Appl. Eng. Res. **11**(5), 3019–3023 (2016)
8. W. Wang, Y. Shang, Y. He, Y. Li, J. Liu, BotMark: Automated botnet detection with hybrid analysis of flow-based and graph-based traffic behaviors. Inf. Sci. **511**, 284–296 (2020)
9. W. Wang, Y. Li, X. Wang, J. Liu, X. Zhang, Detecting Android Malicious Apps and Categorizing Benign Apps with Ensemble of Classifiers. Futur. Gener. Comput. Syst. **78**, 987–994 (2018)
10. M.M. Godfrey, M. Zulkernine, Preventing cache-based side-channel attacks in a cloud environment[J]. IEEE transactions on cloud computing **2**(4), 395–408 (2014)
11. Hlavacs H, Treutner T, Gelas J P, et al. Energy consumption side-channel attack at virtual machines in a cloud[C]//2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing. IEEE, 2011: 605-612.
12. Ristenpart T, Tromer E, Shacham H, et al. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds[C]// Proceedings of the 16th ACM conference on Computer and communications security. ACM, 2009: 199-212.
13. Varadarajan V, Ristenpart T, Swift M. Scheduler-based defenses against cross-VM side-channels[C]//23rd {USENIX} Security Symposium ({USENIX} Security 14). 2014: 687-702.
14. Barrowclough JP and Asif R. Securing cloud hypervisors: a survey of the threats, vulnerabilities, and countermeasures[J]. *Security and Communication Networks*, 2018, 20 pages, 2018. https://doi.org/10.1155/2018/1681908.
15. Liu L, Wang A, and Zang WY, *et al*. Empirical evaluation of the hypervisor scheduling on side channel attacks[C].2018 IEEE International Conference on Communications (ICC). Kansas City, MO, USA, 2018: 1-6. doi: 10.1109/ICC.2018.8422722
16. A. Nezarat, Y. Shams, A game theoretic-based distributed detection method for VM-to-hypervisor attacks in cloud environment[J]. *J. Supercomput.* **73**(2), 1–21 (2017). https://doi.org/10.1007/s11227-017-2025-7
17. Wang C, Ma S, and Zhang X, et al. A Hypervisor level provenance system to reconstruct attack story caused by kernel malware[C]. International Conference on Security and Privacy in Communication Systems. Niagara Falls, Canada, 2017:778-792.doi: https://doi.org/10.1007/978-3-319-78813-5_42
18. Szefer J, Keller E, and Lee R B, *et al*. Eliminating the hypervisor attack surface for a more secure cloud[C].Proceedings of the 18th ACM conference on Computer and communications security. Chicago, Illinois, USA, 2011: 401-412.doi: https://doi.org/10.1145/2046707.2046754
19. Liu F, Ge Q, Yarom Y, et al. Catalyst: defeating last-level cache side channel attacks in cloud computing[C]//2016 IEEE international symposium on high performance computer architecture (HPCA). IEEE, 2016: 406-418.
20. S. Sundareswaran, A.C. Squciarini, *Detecting malicious co-resident virtual machines indulging in load-based attacks[C]//International Conference on Information and Communications Security* (Springer, Cham, 2013), pp. 113–124
21. S. Yu, X. Gui, J. Lin, An approach with two-stage mode to detect cache-based side channel attacks[C]//The International Conference on Information Networking 2013 (ICOIN). IEEE, 186–191 (2013)
22. Y. Zhang, A. Juels, A. Oprea, et al., Homealone: Co-residency detection in the cloud via side-channel analysis[C]//2011 IEEE symposium on security and privacy. IEEE, 313–328 (2011)
23. A. Bates, B. Mood, J. Pletcher, et al., On detecting co-resident cloud instances using network flow watermarking techniques[J]. Int. J. Inf. Secur. **13**(2), 171–189 (2014)
24. INCI and Mehmet Sinan, et al. Seriously, get off my cloud! Cross-VM RSA Key Recovery in a Public Cloud[R]. IACR Cryptology ePrint Archive ia.cr/2015/898, 2015.
25. Xu Z, Wang H, Wu Z. A measurement study on co-residence threat inside the cloud[C]//24th {USENIX} Security Symposium ({USENIX} Security 15). 2015: 929-944.
26. Varadarajan V, Zhang Y, Ristenpart T, et al. A placement vulnerability study in multi-tenant public clouds[C]//24th {USENIX} Security Symposium ({USENIX} Security 15). 2015: 913-928.
27. Y. Azar, S. Kamara, I. Menache, et al., Co-location-resistant clouds[J]. CCSW **14**, 9–20 (2014)
28. Y. Han, J. Chan, T. Alpcan, et al., Using virtual machine allocation policies to defend against co-resident attacks in cloud computing[J]. IEEE Transactions on Dependable and Secure Computing **14**(1), 95–108 (2017)
29. Y. Han, J. Chan, T. Alpcan, et al., Virtual machine allocation policies against co-resident attacks in cloud computing[C]//2014 IEEE International Conference on Communications (ICC). IEEE, 786–792 (2014)
30. Zhang Y, Li M, and Bai K, et al. Incentive compatible moving target defense against VM-colocation attacks in clouds[C]. IFIP International Information Security Conference. Springer, Berlin, Heidelberg, 2012:388-399.doi: 10.1007/978-3-642-30436-1_32
31. Y. Han, T. Alpcan, J. Chan, et al., *Security games for virtual machine allocation in cloud computing[C]//International Conference on Decision and Game Theory for Security* (Springer, Cham, 2013), pp. 99–118
32. Y. Han, T. Alpcan, J. Chan, et al., A game theoretical approach to defend against co-resident attacks in cloud computing: preventing co-residence using semi-supervised learning[J]. IEEE Transactions on information Forensics and Security **11**(3), 556–570 (2016)
33. G. Levitin, L. Xing, Y. Dai, et al., Dynamic checkpointing policy in heterogeneous real-time standby systems[J]. IEEE Trans. Comput. **66**(8), 1449–1456 (2017)

34. L. Xing, G. Levitin, Balancing theft and corruption threats by data partition in cloud system with independent server protection[J]. Reliab. Eng. Syst. Saf. **167**, 248–254 (2017)
35. G. Levitin, K. Hausken, H.A. Taboada, et al., Data survivability vs. security in information systems[J]. Reliab. Eng. Syst. Saf. **100**, 19–27 (2012)
36. A. Soofi, M. Irfan Khan, F.-e. Amin, A Review on Data Security in Cloud Computing. Int. J. Comput. Appl. **94**, 975–8887 (2014). https://doi.org/10.5120/16338-5625
37. Leistikow R, Tavangarian D. Secure picture data partitioning for cloud computing services[C]//2013 27th International Conference on Advanced Information Networking and Applications Workshops. IEEE, 2013: 668-671.
38. Shaikh M, Achary A, Menon S, et al. Improving cloud data storage using data partitioning and data recovery using seed block algorithm[J]. International Journal of Latest Technology in Engineering, Management & Applied Science, 2015, 4(1).
39. A.N. Gullhav, J.F. Cordeau, L.M. Hvattum, et al., Adaptive large neighborhood search heuristics for multi-tier service deployment problems in clouds[J]. Eur. J. Oper. Res. **259**(3), 829–846 (2017)
40. G. Levitin, L. Xing, Y. Dai, Co-residence based data vulnerability vs. security in cloud computing system with random server assignment[J]. Eur. J. Oper. Res. **267**(2), 676–686 (2018)

Publisher's Note