Decentralized Custody Scheme with Game-Theoretic Security

Zhaohua Chen and Guang Yang

Conflux

Abstract. Custodian is a core financial service in which the custodian holds in safekeeping assets on behalf of the client. Although traditional custody service is typically endorsed by centralized authorities, decentralized custody scheme has become technically feasible since the emergence of digital assets, and furthermore it is badly needed by new applications such as blockchain and DeFi (Decentralized Finance).
In this work, we propose a framework of decentralized asset custody scheme that is able to support a large number of custodians and safely hold customer assets of multiple times value of the total security deposit. The proposed custody scheme distributes custodians and assets into many custodian groups via combinatorial designs and random sampling, where each group fully controls the assigned assets. Since every custodian group is small, the overhead cost is significantly reduced. The liveness is also improved because even a single alive group would be able to process transactions. The security of this custody scheme is guaranteed in the game-theoretic sense, such that any adversary corrupting a bounded fraction of custodians cannot move assets more than his own security deposit. We further analyze the security and performance of our constructions, and give explicit examples with concrete numbers and figures for a better understanding of our results.

Keywords: Decentralized asset custody · Mechanism design · Group assignment scheme · Game-theoretic security.

1 Introduction

Custody is a core financial service in which an institution, known as the custodian, holds in safekeeping assets such as stocks, bonds, precious metals and currency on behalf of the client. Custody service reduces the risk of clients losing their assets or having them stolen, and in many scenarios a third party custodian is required by regulation to avoid systematic risk. In general, security is the most important reason why people use custody service and place their assets for safekeeping in custodian institutions.

The security of traditional custody service is usually endorsed by the reputation of the custodian, together with the legal and regulatory system. Such centralized endorsement used to be the only viable option until the emergence of digital assets. A major advantage of digital assets over their physical counterparts is that they are intrinsically integrated with information technology such as Internet and modern cryptography. Along with the evolution of cryptographic authentication technology, e.g. multiple/threshold signature scheme and zero-knowledge proof systems, it becomes technically feasible to have in safekeeping assets held by multiple custodians, so that no single custodian is able to use the assets covertly.

From a systematic point of view, custody service provided by a federation of multiple independent custodian has better robustness and resistance against single point failure, and hence achieves a higher level of security. Such credit enhancement is especially important for safekeeping of cryptoassets on decentralized blockchains such as Bitcoin [21] and Ethereum [33], where the legal and regulatory system is absent or at least way behind the development of applications. For example, the largest cryptocurrency exchange at that time, Mt. Gox, announced that approximately 850,000 bitcoins were stolen and went bankrupt in 2014 [31]; and in the year 2019 alone, at least 7 cryptocurrency exchanges claimed being hacked and loss of cryptoassets totaled to around 1.39 billion dollars [2]. However, it is difficult for customers to distinguish that whether the claimed loss was caused by hacker attack or internal fraud and embezzlement, and hence raises the need for decentralized custody.

Decentralized custody finds applications in many scenarios related to blockchain and digital finance. A motivating example is the *cross-chain asset mapping* service which maps cryptoassets on one blockchain to tokens on another blockchain for inter-chain operability. For instance, the mapping from Bitcoin to Ethereum enables usage of tokens representing bitcoins within Ethereum ecosystem, and

in the meanwhile, the original bitcoins must be safeguarded so that the bitcoin tokens are guaranteed redeemable for real bitcoins in full. Nowadays the volume of cryptoassets invested into Ethereum DeFi applications has been increasing at an extraordinary pace and broke two billion dollars recently [1], among which a significant fraction (e.g. hBTC [15], imBTC [29], tBTC [19], wBTC [30], renBTC [25], etc.) is mapped from Bitcoin. Due to the reality that most of those DeFi applications and tokens remain in a gray area of regulation, decentralized cryptoasset custody (within or across institutions) turns out an attractive approach for better security and credit enhancement.

In this work, we propose a framework of decentralized asset custody scheme which is secure in the game-theoretic sense. More specifically, custodians and assets are distributed into multiple custodian groups, where each group consists of few custodians as its members and fully controls a small portion of in safekeeping assets. The authentication of each custodian group can be implemented with voting or threshold signature among group members. Under this framework transactions can be processed more efficiently within group members, since the computational and communicational cost is significantly reduced. The liveness and robustness is also improved since even a single alive custodian group is able to process some transactions.

The security of our proposed custody scheme is guaranteed in a game-theoretic sense: every custodian in this scheme must offer a fund as security deposit, and the system remains secure as long as an adversary cannot steal more asset than his deposit, i.e. comparing to launching an attack the adversary would be better off by just withdrawing the security deposit of custodian nodes under his control. Specifically, we prove that for adversary corrupting a bounded fraction of custodians, the proposed decentralized custody scheme is able to safeguard customer assets of multiple times value of the total security deposit.

1.1 Related Works

The prototype of decentralized custody scheme first appears in Bitcoin as multisignature (multisig) [6], where the authentication requires signatures from multiple private keys rather than a single signature from one key. For example, an M-of-N address requires signatures by M out of totally N predetermined private keys to move the money. This naïve scheme works well for small M and N but cannot scale out, because the computational and communicational cost of authenticating and validating each transaction grows linearly in M and N. Both efficiency and liveness of the scheme are compromised for large M and N, especially in the sleepy model proposed by Pass and Shi [22] where key holders do not always response in time. In practice, multisignature scheme is typically used at wallet level rather than as a public service, since the scheme becomes costly for large N and most Bitcoin wallets only support $N \leq 7$.

Multisignature schemes may be coupled with advanced digital signature techniques such as threshold signature [7, 13] or aggregate signature [4, 20, 27] to reduce the cost of verifying multi-signed signatures. However, the signing process still requires involvement of at least M parties and the cost of communication can hardly be significantly reduced. Furthermore we remark that these advanced signature techniques are also compatible with our scheme, as authentication mechanism for single custodian groups.

There are also custody solutions based on scripts or contract codes. These schemes usually only work within a single blockchain since it is difficult to get reliable information outside the consensus boundary, which essentially requires an oracle. Although there are techniques such as Atomic Swap [5] and Hashlock/Hash TimeLock Contracts (HTLC) [6] enabling code-based cross-chain interoperability, the use cases are limited and far from being able to provide general purpose custody service.

As for the cross-chain asset mapping service, existing solutions include following three types:

- Centralized: custody in a trusted central authority, with the endorsement fully from that authority, e.g. hBTC [15], wBTC [30] and imBTC [29];
- Consortium: custody in multisignature accounts controlled by an alliance of members, and endorsed by all alliance members, e.g. cBTC [9] (in its current version) and Polkadot [32];
- Decentralized (with deposit/collateral): custody provided by permissionless custodians, with security guaranteed by over-collateralized cryptoassets, e.g. tBTC [19] and renBTC [25].

Although the last type seems satisfiable in decentralization and security against single point failure and collusion, there are significant drawbacks as well: the first drawback is the inefficiency of over-collateralization, i.e. tBTC requires the custodian to provide collateral worth of 150% value of

customer's assets, and renBTC requires 300%; the second drawback is that the collateral is not the same type as the in safekeeping assets, and hence it may be insufficient to endorse safety of the custody service in market volatility.

Furthermore, we remark that tBTC and renBTC have security guaranteed in the game-theoretic sense that an adversary will not launch a non-profitable attack, and renBTC even partition custodians (i.e. Darknodes in its context) into groups for better efficiency and liveness. However, renBTC applies the trivial non-overlapping group partition and hence it has poor capital efficiency. More importantly, it cannot support homogeneous collateral as in safekeeping assets, since otherwise an adversary corrupting a single group will be able to control more assets than his own security deposit.

1.2 Our Contributions

In this work, we propose the framework of implementing decentralized asset custody scheme with overlapping group assignments, where such group assignment consists of many subsets of custodians as groups, and each group is controlled by few group members and holds a small faction of the total assets including both funds from customers and security deposit from custodians.

By distributing in safekeeping assets into many small custodian groups, every customer-requested transaction can be assigned to a specific custodian group and thereafter processed within the group members. Thus each transaction only requires attention and approval from custodians in that small group (rather than a majority of all participants in the scheme), which reduces the cost and latency of transaction processing. Furthermore this framework provides better liveness guarantee, since transaction can be processed even if only one single custodian group has sufficient many members available online, and in the mean while the vast majority of custodians in the system can stay offline.

Our custody scheme achieves security in the game-theoretic sense that for an adversary corrupting γ fraction of individual custodians and γ is not too large, the number of corrupted groups would be less than γ. Therefore, the adversary cannot jeopardize the system as long as the total security deposit of corrupted custodians outweighs the total assets held by corrupted groups, since any loss caused by such an adversary can be easily compensated with security deposits of misbehaved custodians.

The model of our decentralized custody scheme is formalized in Section 3, where we formally define security of a decentralized custody scheme. Furthermore we introduce the notion of efficiency factor η to measure the capital efficiency, which measures the capability of safely holding assets for exterior customers with a fixed total amount of security deposit.

We propose in Section 4 specific constructions of the underlying group assignments based on enumeration of all subsets of a fixed size and two kinds of combinatorial designs respectively. We also analyze the security and efficiency factors of custody schemes induced by these constructions. For example, we show it is possible to assign 20 custodians into $38,760$ groups such that as long as adversary corrupts $\gamma \leq 2/5$ fraction of all custodians, the whole custody scheme is capable of safekeeping assets worthy of $\eta > 20$ times of total security deposits, i.e. the scheme can safely hold over 100 units of value with total security deposit of 5 unit. Furthermore, based on the polynomial combinatorial design, the decentralized custody scheme is able to support 121 participants as custodians with $161,051$ groups such that the efficiency factor $\eta > 60$ against any adversary corrupting no more than $2/11$ of custodians. In general, the efficiency factor is better for larger number of participants and groups, except for the drawback that the number of groups grows polynomially in the number of participants and hence soon becomes infeasible to manage in practice.

To mitigate the problem of too many groups in one assignment, we apply a random sampling on the group assignment and prove that with high probability the newly sampled assignment has a comparable efficiency factor. More specifically, if the original custody scheme consists of n participants and has the efficiency factor η against adversary corrupting at most γ faction of all participants, then by sampling $O\left(n\eta \cdot \log(\gamma^{-1})/\gamma\right) \sim O\left(n\eta\right)$ many groups the newly generated custody scheme would have efficiency factor $\eta' \geq \sqrt{\eta+1} - 2$ against the same adversary with high probability. See Section 5 for the precise statement and more details.

Concretely, using the random sampling technique we are able to construct a custody scheme of 1000 participants and $323,825$ groups such that the efficiency factor $\eta > 10$ against adversary corrupting up to 388 participants.

In Section 6 we investigate the complexity of finding optimal corrupting solutions. We prove that given a group assignment, it is easy to find a solution no worse than average case but it turns out NP-hard to find an optimal corrupting strategy in general. However, this is still insufficient to complement the randomized sampling construction as a validation method.

2 Preliminaries

Hypergeometric distribution and tail bound. A hypergeometric distribution $\mathcal{H}(N, K, n)$ is a discrete probability distribution with the following probability mass function:

$$\Pr[\mathcal{H}(N, K, n) = k] = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Hypergeometric distribution draws the probability of getting k items with a feature in n draws without replacement, from totally N items, K of which with that feature. We are mostly interested in the tail bound of Hypergeometric distribution.

Lemma 1 ([8,17]). *Let $X \sim \mathcal{H}(N, K, n)$ and $p = K/N$, then we have the following:*

$$\Pr[X \le (p-t)n] \le \exp(-nD(p-t\|p)) \le \exp(-2nt^2), \quad 0 < t < p$$
$$\Pr[X \ge (p+t)n] \le \exp(-nD(p+t\|p)) \le \exp(-2nt^2), \quad 0 < t < 1-p$$

where $D(\cdot\|\cdot)$ stands for the Kullback-Leibler divergence with input real numbers naturally generalized to probability distributions, i.e. $D(a\|b) := a \cdot \ln(a/b) + (1-a) \cdot \ln\big((1-a)/(1-b)\big)$, for $a, b \in (0,1)$.

For $0 < a < b < c < 1$, the naturally generalized Kullback-Leibler divergence satisfies $D(a\|b) < D(a\|c)$, $D(c\|b) < D(c\|a)$, $D(b\|a) < D(c\|a)$, $D(b\|c) < D(a\|c)$.

Stirling's formula. Stirling's formula provides good estimation of factorial terms. Specifically, we will use the following version.

Lemma 2 ([26]). *For any integer $n \ge 1$, we have $\sqrt{2\pi} \cdot e^{-n} n^{n+\frac{1}{2}} < n! < e^{1-n} n^{n+\frac{1}{2}}$.*

Block designs. A block design is a particular combinatorial design consisting of a set and a family of subsets (called blocks) whose arrangements satisfy generalized concepts of balance and symmetry.

Definition 1 (Block design, from [28]). *Let n, k, λ and r be positive integers such that $n > k \ge r$. A block design $(S, \mathcal{A})$ is called an r-(n, k, λ)-design if $|S| = n$ and $\mathcal{A}$ is a family of k-element subsets of S (called blocks), such that every r-element subset of S is contained in exactly λ blocks in $\mathcal{A}$.*

In this work we mainly consider designs with $\lambda = 1$, and hence $\mathcal{A}$ cannot contain duplicate blocks.

Corollary 1. *In a r-(n, k, λ)-design, the number of blocks is $m = \lambda \binom{n}{r} \Big/ \binom{k}{r}$.*

Corollary 2. *Suppose $(S, \mathcal{A})$ is an r-(n, k, λ)-design. Then for any $1 \le t \le r$, $(S, \mathcal{A})$ is also a t-(n, k, λ_t)-design, for $\lambda_t = \lambda \cdot \binom{n-t}{r-t} / \binom{k-t}{r-t}$.*

Corollary 3. *Suppose $(S, \mathcal{A})$ is an r-(n, k, λ)-design. Then for any $Z \subseteq S$ with size $|Z| = t \le r$, $(S - Z, \{A - Z : Z \subseteq A \in \mathcal{A}\})$ is a $(r-t)$-$(n-t, k-t, \lambda)$-design.*

Furthermore, we remark that although non-trivial block designs exist universally by [28], only very few such designs are explicitly constructed.

3 Model

In this work we consider the decentralized custody scheme without relying on any trusted party. More specifically, we investigate the feasibility that n participants (a.k.a. n *nodes*) $S = \{1, 2, \cdots, n\}$ jointly provide the custody service, such that the security is guaranteed as long as the number of corrupted participants is bounded by a fraction of n, e.g. the adversary cannot corrupt more than $n/3$ nodes simultaneously. This assumption of existence of an honest majority is much milder than assuming a single party trusted by everyone, and hence it likely leads to better security guarantee in practice.

Furthermore, we consider the security of the custody scheme in a game-theoretic sense: the adversary may corrupt multiple nodes, but he will not launch an attack if the profit does not exceed the cost of launching such an attack. In other words, misbehavior of adversary nodes is not considered as breaking the security as long as the resultant damage is recoverable. To increase the cost of attack

and compensate potential loss caused by attacks, we require every custodian in our scheme to provide a security deposit, which will be confiscated and used for compensation in case of misbehavior. Thus, no rational adversary would ever launch an attack if the security deposit outweighs the revenue of a successful attack.

A trivial but useless solution. In the most trivial solution, the asset under custody can only be moved when approved by all custodians or at least a majority of them. However, getting such an approval becomes expensive and even infeasible in practice as n grows, especially when honest participants may go off-line (as in the sleepy model [22]), which renders the trivial scheme useless.

Although the above solution is not satisfiable, it does provide enlightening ideas for designing a better custody scheme. The threshold authorization scheme guarantees that the adversary cannot move any asset under custody if he does not control enough many nodes. In a more general view, this is a particular case of game-theoretically secure schemes where the adversary's security deposit outweighs his revenue of launching an attack when the adversary has bounded power. As long as this property is satisfied the custody scheme should be secure in the game-theoretic sense.

In particular, the following toy example shows feasibility of implementing our idea with multiple subsets of S as custodian groups, i.e. each subset of S only controls a fraction of the total asset under custody.

Example 1 (Toy example). Consider the case when $n = 5$, total asset under custody is 10, and total security deposit is $6 \times 5 = 30$. Let each of the $\binom{5}{3} = 10$ triads of S form a custodian group, and assign the asset under custody and the security deposit equally to every group, i.e. each custodian group controls 4 units. If the asset controlled by each custodian group can be moved with approval of 2 out of 3 members in that group, then an adversary controlling 2 nodes is able to corrupt exactly 3 custodian groups. However, by controlling 3 groups the adversary can only move $4 \times 3 = 12$ units, which is no more than the security deposit of adversary nodes, which is also 12. Thus the custody scheme for $n = 5$ is secure against adversaries controlling no more than two nodes.

In what follows, we will formalize the model of decentralized custody scheme with asset distributed among custodian groups.

Let $S = \{1, 2, \cdots, n\}$ be the set of all nodes, and let $\mathcal{A}$ denote the custodian group assignment, which is a family of m subsets of S. We emphasize that these subsets do not have to be disjoint. In fact, it is imperative to use overlapping subsets in any meaningful solution.

In this work we focus on the symmetric setting where every node provides the same amount of security deposit and every group in $\mathcal{A}$ is of the same size k. We let a universal parameter $\mu \in [1/2, 1)$ denote the authentication threshold for every single custodian group, i.e. with approval of *strictly above* μk group members the asset controlled by that group can be settled arbitrarily. In particular, our discussion mainly focuses on $\mu = 1/2$ and $\mu = 2/3$, corresponding to authentication with simple majority, and greater than 2/3 majority respectively. [1] For simplicity we let $r = \lceil \mu k + \epsilon \rceil$ denote the smallest integer greater than μk, and hence the authentication of every custodian group is essentially an r-of-k threshold signature scheme.

We represent the adversary power with $\gamma \in (0, 1)$, which refers to the fraction of corrupted nodes in S. Specifically, we let $s = \lfloor \gamma n \rfloor$ denote the number of corrupted nodes in S. For succinctness we slightly abuse the notation and assume that γn is always an integer, i.e. $s = \gamma n \in \mathbb{N}$. The adversary is allowed to adaptively select corrupted nodes and then get all information and full control of those nodes thereafter, as long as the number of corrupted node does not exceed s. In case a group in $\mathcal{A}$ contains $\geq r$ corrupted nodes, we say that group is corrupted. Furthermore, we remark that the adversary has reasonably bounded computing power, so that he cannot break cryptographic primitives such as digital signatures.

Given a custodian scheme with group assignment $\mathcal{A}$ and authentication threshold μ, together with γ for the bound of adversary power, we use the function $f(\gamma; \mathcal{A}, \mu)$ to denote the maximal number of groups that may be corrupted by an adversary controlling up to $s = \gamma n$ nodes (although it is indeed

[1] $\mu \geq 1/2$ is the necessary and sufficient condition for Byzantine agreement under synchrony, i.e. when all members are well-connected [18]. $\mu \geq 2/3$ is necessary and sufficient for Byzantine agreement under partial synchrony or asynchrony even with digital signatures [23]. We further remark that smaller μ implies less security but better liveness, e.g. when $\mu \to 1$ even a single corrupted member is able to block a custodian group. However, the discussion of liveness is beyond the scope of this work.

Table 1. The efficiency factor of the custodian assignment as in Example 1

Parameters \ Adversary power (γ)	$\gamma = 1/5$	$\gamma = 2/5$	$\gamma = 3/5$	$\gamma = 4/5$
Corrupted nodes (s)	1	2	3	4
Corrupted custodian groups ($f(\gamma; \mathcal{A}, \mu)$)	0	3	7	10
Efficiency factor (η)	∞	1/3	$-1/7$	$-1/5$

The authentication threshold is realized as $r = 2$ and $\mu = 1/2$ (in this example equivalent to have $\mu \in [1/3, 2/3)$). For $\gamma = 1/5$ and $\gamma = 2/5$, the scheme is secure with $\eta = \infty$ and $\eta = 1/3$ respectively. For $\gamma \geq 3/5$ the scheme is insecure and $\eta \leq 0$.

NP-hard to find the optimal attacking strategy in general, as discussed in Section 6). Formally,

$$f(\gamma; \mathcal{A}, \mu) := \max_{B \subseteq S : |B| = \gamma n} \left| \{ A \in \mathcal{A} \;\middle|\; |A \cap B| \geq \mu k \} \right| \tag{1}$$

From $f(\gamma; \mathcal{A}, \mu)$ we define the security of this specific custody scheme as follows:

Definition 2 (Security of Custody Schemes). *For a custody scheme with assignment $\mathcal{A}$ of m groups and parameters μ, γ as above, we say that the custody scheme is* γ-reliable *if*

$$f(\gamma; \mathcal{A}, \mu) \leq \gamma \cdot m.$$

Furthermore, the scheme is called secure against γ-adversary *(or simply* secure*) if it is γ'-reliable for every $\gamma' \in [0, \gamma]$.*

The above definition of security guarantees that by controlling γn nodes, the adversary is only able to corrupt $f(\gamma; \mathcal{A}, \mu)$ custodian groups. Therefore, by launching an attack the adversary is able to seize the funds of $f(\gamma; \mathcal{A}, \mu)$ custodian groups, at the cost of losing the security deposit of γn corrupted nodes. The custody scheme is secure as long as the asset under custody of $f(\gamma; \mathcal{A}, \mu)$ groups is bounded by the security deposit of γn nodes, which indeed requires $f(\gamma; \mathcal{A}, \mu) \leq \gamma m$ since the deposit of γn nodes can be as much as asset under custody of γm groups.

Note that the gap between $f(\gamma; \mathcal{A}, \mu)$ and γm corresponds to the ratio of external asset that can be securely held compared to security deposit, which should be as large as possible. For example, $f(\gamma; \mathcal{A}, \mu) = \gamma m$ implies that the custody scheme is only capable of securely holding the security deposit of its members but cannot provide custody service for any external users. To describe the ability of securely holding exterior assets, we introduce the notion of efficiency factor as in the following definition:

Definition 3 (Efficiency factor). *Given a custody scheme with $\mathcal{A}, m, \mu$ and γ defined as above, the* efficiency factor *of this scheme, denoted by η, is defined as:*

$$\eta := \frac{\gamma m - f(\gamma; \mathcal{A}, \mu)}{f(\gamma; \mathcal{A}, \mu)}.$$

It is immediate to verify that $\eta \geq 0$ for any custody scheme secure as in Definition 2. This factor η indeed equals to the maximal ratio of capable exterior assets to pledged assets that the underlying custody scheme is able to handle. For example, $\eta = 1$ means the system is secure when the total value of exterior assets is no more than total pledged security deposit.

Putting into our formal definition, the trivial solution has only one custodian group (i.e. $m = 1$, $k = n$), and $\eta = \infty$ for $\gamma \leq \mu$ and $\eta < 0$ for $\gamma > \mu$; the custody scheme in Example 1 has its efficiency factor η changing according to the adversary power γ as summarized in the following table.

From the formalization of our decentralized custody scheme, it is clear that the custodian group assignment $\mathcal{A}$ is the core of the whole custody scheme. In particular, for every specific group assignment $\mathcal{A}$ and fixed constant μ (say, $\mu \in \{1/2, 2/3\}$), the parameters n, m, k are already specified in $\mathcal{A}$, and the maximal number of corrupted groups and the efficiency factor η are functions solely depending on the adversary power γ.

Therefore, in the rest of this article we will focus on construction and analysis of custodian group assignments. More specifically, we want the group assignment to support a decentralized custody

scheme with the efficiency factor η as large as possible for adversary power γ bounded within a reasonable range (e.g. $\gamma < 1/3$ or so), and at the same time the group assignment $\mathcal{A}$ should have large n (e.g. over a thousand) and small m and k (e.g. ideally bounded by one hundred).

3.1 Limitation and Impossibilities

First we remark that the reliability and efficiency factor of an alignment $\mathcal{A}$ is not necessarily monotone. That is, there exists $\gamma_a > \gamma_b$ such that the custody scheme induced by some $\mathcal{A}$ is γ_a-reliable but not γ_b-reliable, as exhibited in the following example.

Example 2. Consider the case $S = \{1, 2, \cdots, 10\}$, $\mu = 2/3$, and the assignment scheme $\mathcal{A} = \{(1, 2, 3, 4, 5), (6, 7, 8, 9, 10)\}$. For $\gamma = 2/5$, the system is not γ-reliable since an adversary with 4 corrupted nodes may corrupt one group and get half of the total security deposit. However, when $\gamma = 3/5$, the system is γ-reliable since the adversary controls at most one group with more than half of the total deposit.

We also remark that in general a decentralized custody scheme based on group assignments cannot be γ-reliable for all γ unless $r = k$. Since when $r < k$, by corrupting $n - 1$ nodes the adversary is able to corrupt all custodian groups and hence get full control of the whole system.

4 Constructions of Group Assignments

In this section, we propose three types of group assignments and analyze the performance of resultant custody schemes. We also provide empirical analysis of these schemes with concrete numbers for a better understanding.

4.1 Type 1: Symmetric Design

Construction 1. Given n and k, let $\mathcal{A}_{all}$ be a family consisting of all size-k subsets of S as custodian groups, i.e. $\mathcal{A}_{all}$ is an assignment with $m = \binom{n}{k}$ groups where each group has k nodes. Then for every authentication threshold μ, a custody scheme can be constructed from $\mathcal{A}_{all}$.

Since all nodes are symmetric in $\mathcal{A}_{all}$, it immediately follows that the number of corrupted groups in the above custody scheme only depends on the number of corrupted nodes. Thus it suffices to consider the adversary corrupts any set of γn nodes, and the number of corrupted groups can be calculated as follows:

$$f(\gamma; \mathcal{A}, \mu) = \sum_{r \le t \le k} \binom{\gamma n}{t} \binom{n - \gamma n}{k - t} = \binom{n}{k} \cdot \sum_{t=r}^{k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}}. \tag{2}$$

By Definition 2 and recalling $m = \binom{n}{k}$, the custody scheme is γ-reliable if and only if:

$$\gamma \ge \frac{f(\gamma; \mathcal{A}, \mu)}{m} = \sum_{t=r}^{k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}},$$

and the efficiency factor η turns out:

$$\eta = \frac{\gamma m - f(\gamma; \mathcal{A}, \mu)}{f(\gamma; \mathcal{A}, \mu)} = \frac{\gamma \binom{n}{k}}{\sum_{t=r}^{k} \binom{\gamma n}{t}\binom{n-\gamma n}{k-t}} - 1.$$

The reliability of $\mathcal{A}_{all}$ against adversary corrupting exactly γn nodes naturally implies to security against adversary corrupting $\le \gamma n$ nodes for reasonable γ, as proved in the following lemma. For typical choice of $\mu \ge 1/2$ and $\gamma < 1 - k/n$, the translation from reliability to security holds for $\gamma < \mu - \frac{1}{2(k-1)}$.

Lemma 3. *Suppose $\mathcal{A}_{all}$ is a group assignment following Construction 1 with parameters n and k. If the custody scheme derived from $\mathcal{A}_{all}$ and any authentication threshold μ is γ-reliable and $\gamma \le \min\left\{\frac{\mu k - 1}{k-1} + \frac{1}{n}, 1 - \frac{k}{n}\right\}$, then it is secure against γ-adversary (i.e. γ'-reliable for every $\gamma' \le \gamma$).*

The proof of Lemma 3 is obtained by comparing η for adversary corrupting s nodes and $s-1$ nodes. See Appendix A.1 for the complete proof.

Based on Lemma 3, we prove that $\mathcal{A}_{all}$ is secure for γ close to $1/2$ and appropriately large k.

Theorem 1. *When $\mu \geq 1/2$, $k \geq 10$ and $n \geq 2k$, the custody scheme derived from $\mathcal{A}_{all}$ and μ is secure against γ_{all}-adversary, for γ_{all} defined as follows:*

$$\gamma_{all} := \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}.$$

Proof (Sketch). We first show that when $n \geq 2k$ and $k \geq 10$, the system is reliable for $\gamma \in [\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k}, \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}]$. In fact, recall that the system is secure as long as

$$\gamma \geq \sum_{\mu k < t \leq k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}},$$

and by the tail bound of hypergeometric distribution, it is sufficient with

$$\gamma \geq \exp(-kD(\mu\|\gamma)) \geq \sum_{\mu k < t \leq k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}}.$$

By the property of Kullbeck-Leibler divergence on two real numbers, we can obtain the above if we have

$$\gamma \geq (4\gamma(1-\gamma))^{k/2} \geq \exp(-kD(\mu\|\gamma)),$$

which establishes as long as $\gamma \in [\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k}, \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}]$, $\mu > 1/2$, $k \geq 10$ and $n \geq 2k$.

Lemma 3 shows that as long as the system is reliable w.r.t $\gamma < \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k} < \min\left\{\frac{\mu k - 1}{k-1} + \frac{1}{n}, 1 - \frac{k}{n}\right\}$, then the system is secure against γ-adversary. Therefore, we finish the proof. The complete proof of this theorem is deferred to Appendix A.2. □

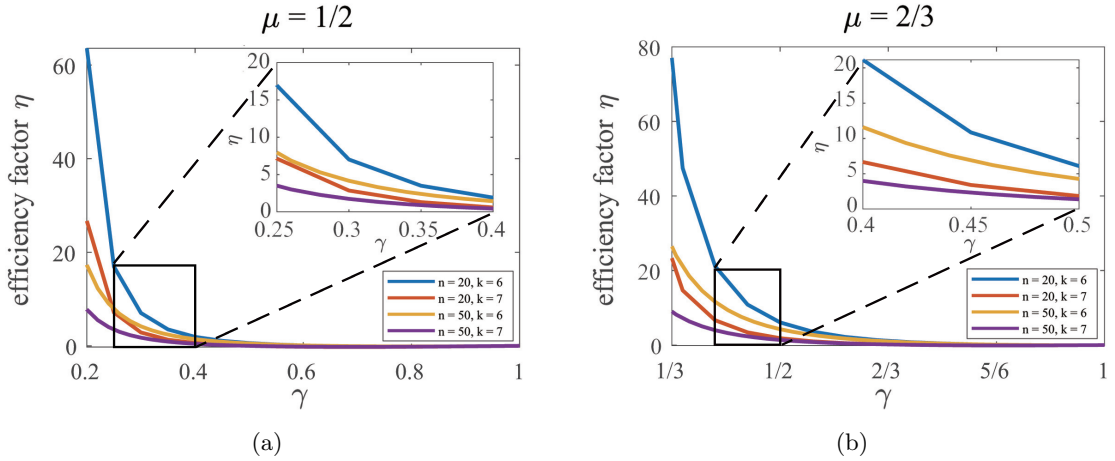

Fig. 1. The efficiency factor η versus the corrupted fraction γ for $\mathcal{A}_{all}$ as in Construction 1. In particular, $\eta < 0$ if the custody scheme is not secure for corresponding γ.

Fig. 1 depicts the relation between efficiency factor η and adversary corrupted fraction γ, for $n \in \{20, 50\}$, $k \in \{6, 7\}$, and $\mu \in \{1/2, 2/3\}$. From this figure we can see that with fixed n, k and μ, the efficiency factor η decreases as γ grows. Further, for combinations of reasonably large n and k, the efficiency factor η can be more than 10 while γ is roughly $1/3$. For instance, when $n = 20$, $k = 6$ and $\mu = 2/3$ as in Fig. 1(b), there is $m = \binom{20}{6} = 38,760$ and the efficiency factor $\eta = 21.1486$ against adversary with power $\gamma = 2/5$.

Fig. 2 depicts the behavior of the efficiency factor η versus the custodian group size k, for $n \in \{20, 50\}$, $\mu \in \{1/2, 2/3\}$ and $\gamma \in \{1/5, 1/4, 1/3\}$. The figure shows in general η increases with k in custody schemes induced by $\mathcal{A}_{all}$. The sawteeth appears on the curves because of the rounding of r and s, i.e. the authentication threshold and the number of corrupted nodes.

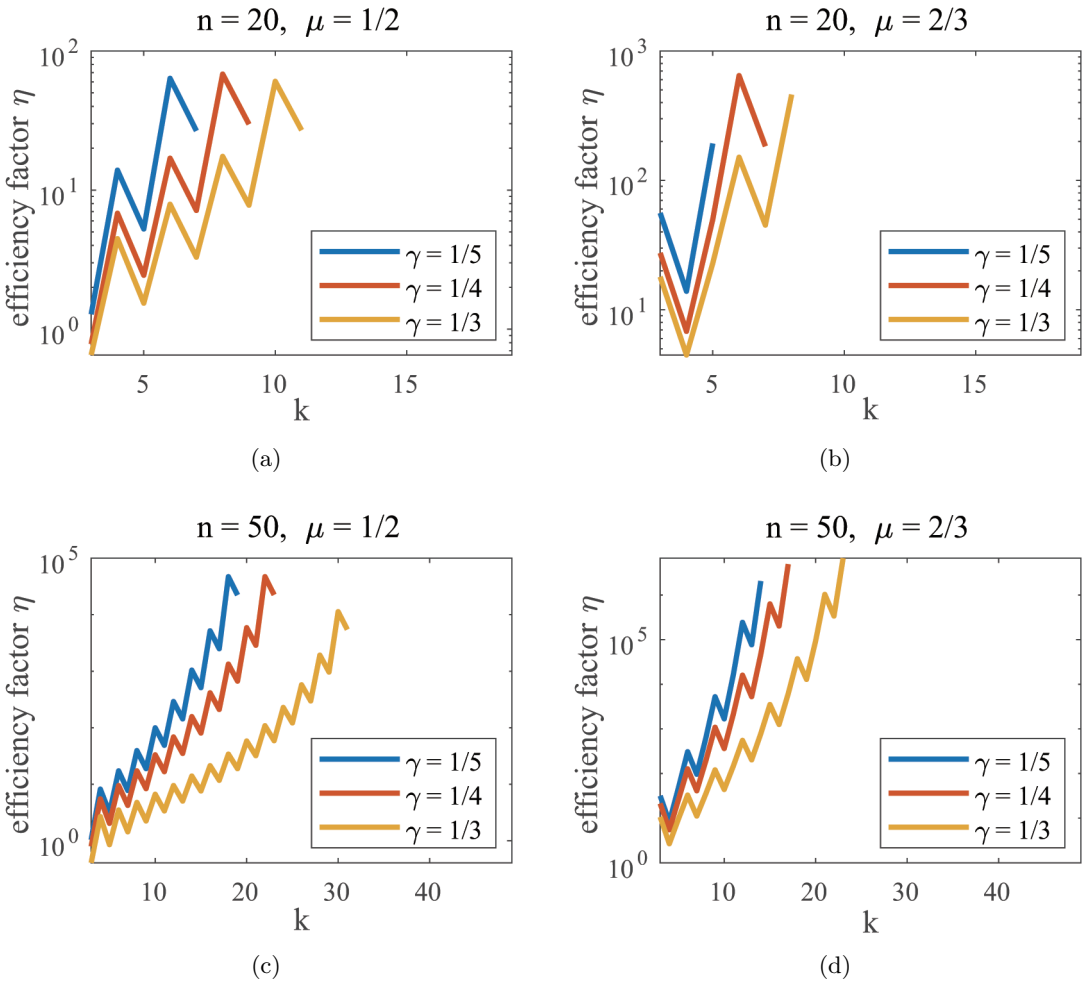

Fig. 2. The efficiency factor η versus the custodian group size k for $\mathcal{A}_{all}$ as in Construction 1. Blank points on the right side refer to $\eta = \infty$ when adversary cannot corrupt even a single group.

Finally we remark that the construction of $\mathcal{A}_{all}$ by itself is mainly a theoretical result. Because the size of such group assignment $m = \binom{n}{k}$ grows too fast and hence n and k must be severely bounded in practice, e.g. $n \sim 20$ and $k \sim 5$. Otherwise it would be too expensive or even impossible to manage the custody scheme consisting of m custodian groups. A solution to mitigate the above issues is by random sampling, as exhibited in Section 5.

4.2 Type 2: Block Design

In this section, we consider group assignment schemes induced by block designs (see Section 2 for necessary preliminaries). In fact, block designs naturally extend Construction 1, in the sense that $\mathcal{A}_{all}$ is a degenerated block design with $\lambda = 1$. In what follows, a "block" in the block design is also called a "group" in the group assignment scheme.

The following lemma shows the effectiveness of block designs:

Lemma 4. *For every r-(n,k,λ)-design and $\mu = r/k$, the induced custody scheme is γ-reliable with efficiency factor η lower bounded as follows:*

$$\eta \geq \frac{s}{n} \cdot \frac{\binom{n}{r}}{\binom{k}{r}} \Big/ \min\left\{ \binom{s}{r}, \frac{\binom{n}{r}}{\binom{k}{r}} \right\} - 1.$$

The above lemma holds since with $s = \gamma n$ corrupted nodes, the adversary controls at most $\lambda \cdot \binom{s}{r}$ custodian groups out of $m = \lambda \cdot \binom{n}{r}/\binom{k}{r}$ groups in total. Fig. 3 shows the lower bound of η obtained by Lemma 4 versus the adversary's power γ for different block designs with $\mu = (r-1)/k$.

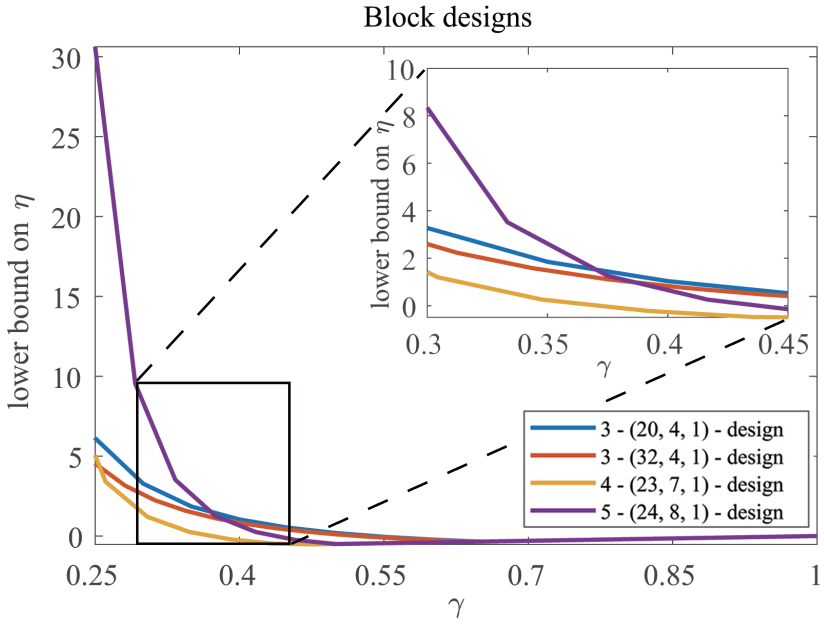

Fig. 3. The lower bound on η by Lemma 4, versus the corrupted fraction γ, when $\mu = (r-1)/k$. All four designs shown in this figure have explicit constructions [28]. $\eta < 0$ if the custody scheme is not secure for corresponding γ.

Resembling what we do in Section 4.1, we prove that block designs are always reliable with properly small γ.

Theorem 2. *Let $s = \gamma n$ be the number of corrupted nodes and $\mu = (r-1)/k$. Suppose we have $\mu \geq 1/2$, $r > \mu k \geq 2$ and $n \geq 3k - 3$. The custody scheme induced by an r-(n,k,λ)-design is γ-reliable as long as $\binom{s}{r} \leq \binom{n}{r} \Big/ \binom{k}{r}$ and $s \leq \frac{n}{k}\mu^{\frac{1}{r-1}} + r - 1$. Furthermore, the reliability is monotone in γ and hence it immediately translates to security against adversary corrupting up to γn nodes.*

Proof (Sketch). We consider the lower bound of the efficiency factor η given by Lemma 4. Under the condition that $\binom{s}{r} \leq \binom{n}{r} \Big/ \binom{k}{r}$, the system is reliable when

$$\frac{s}{n} \frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} \geq 1.$$

A key observation is that,

$$\frac{s}{n} \frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} = \prod_{t=1}^{r-1} \frac{(n-t)(r-t)}{(s-t)(k-t)} \cdot \frac{r}{k} > \prod_{t=1}^{r-1} \frac{(n-t)(r-t)}{(s-t)(k-t)} \cdot \mu.$$

With the above observation we conclude the following holds for any $1 \leq t \leq r-1$,

$$\frac{(n-t)(r-t)}{(s-t)(k-t)} \geq \mu^{-\frac{1}{r-1}}.$$

Therefore,

$$\frac{s}{n}\frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} > \prod_{t=1}^{r-1}\frac{(n-t)(r-t)}{(s-t)(k-t)}\cdot\mu \geq \left(\mu^{-\frac{1}{r-1}}\right)^{r-1}\cdot\mu = 1.$$

The full proof of the theorem is deferred to Appendix A.3. □

Theorem 2 guarantees the security of the custody scheme for proper γ. Although Lemma 4 does not provide an ideal lower bound estimation for large γ (see Fig. 3), we still manage to achieve some satisfying results. In particular, using the custody scheme induced from the 5-$(24, 8, 1)$-design with $m = 759$ custodian groups with $\mu = 1/2$, the efficiency factor $\eta \geq 30.6250$ for $\gamma = 1/4$. This significantly outperforms the symmetric design $\mathcal{A}_{all}$, which has $m = 38,760$ and $\eta = 16.9444$ when $n = 20$, $k = 6$, $\mu = 1/2$ and $\gamma = 1/4$ (see Fig. 1(a)).

4.3 Type 3: Polynomial Design

In this section, we propose another way to construct group assignments using polynomial-based combinatorial designs.

Construction 2. For a prime k and fixed positive integer $d < k$, let $S = \{(a, b) \mid a, b \in \mathbb{Z}/k\mathbb{Z}\}$ be the set of $n = k^2$ elements in $(\mathbb{Z}/k\mathbb{Z})^2$. The polynomial design of group assignment $\mathcal{A}_{poly}$ is a family of $m = k^d$ subsets of S defined as follows:

$$\mathcal{A}_{poly} := \{A_p \mid p \text{ is degree-}d \text{ monic polynomials over } \mathbb{Z}/k\mathbb{Z}\}, \tag{3}$$

where $A_p := \{(0, p(0)), (1, p(1)), \cdots, (k-1, p(k-1))\}$.

For every authentication threshold μ, a custody scheme can be constructed from $\mathcal{A}_{poly}$.

It is easy to verify that $\mathcal{A}_{poly}$ consists of m distinct groups, and the intersection of any two distinct groups in $\mathcal{A}_{poly}$ is bounded by d, following the Fundamental Theorem of Algebra.

$$\forall A_p, A_q \in \mathcal{A}_{poly} \wedge A_p \neq A_q \implies |A_p \cap A_q| \leq d \tag{4}$$

Hence, we lower bound the efficiency factor of custody schemes in Construction 2 as below.

Lemma 5. *The efficiency factor η for $\mathcal{A}_{poly}$ as in Construction 2 is lower bounded as follows:*

$$\eta \geq \gamma \cdot k^d \cdot \binom{r}{d} \Big/ \binom{s}{d} - 1.$$

where $s = \gamma n$ is the number of corrupted nodes, and $r > \mu k$ is the minimum number of nodes required to corrupt a group.

We assert that an adversary with s nodes is able to corrupt at most $\binom{s}{d}/\binom{r}{d}$ groups. By (4), every d tuple of corrupted nodes appears in at most one corrupted group. On the other hand, every corrupted group contains at least r corrupted nodes and hence $\geq \binom{r}{d}$ corrupted d tuples. Thus the above lower bound for η holds by Definition 2.

In the following theorem, we analyze the security of custody schemes in Construction 2.

Theorem 3. *Suppose $d < \min\{s/2, r\}$. Let $d = \nu k$, then the custody scheme derived from $\mathcal{A}_{poly}$ is secure against γ_{poly}-adversary for γ_{poly} defined as:*

$$\gamma_{poly} := \left(\frac{e^{2-d}}{2\sqrt{2}\pi}\sqrt{\frac{\mu}{\mu-\nu}}\left(\frac{\mu^\mu}{(\mu-\nu)^{\mu-\nu}}\right)^k\right)^{\frac{1}{d-1}}.$$

Proof (Sketch). By Lemma 5, the custody scheme remains reliable as long as

$$\gamma\binom{r}{d}k^d \Big/ \binom{s}{d} \geq 1.$$

The binomial term can be lower bounded using Stirling's formula, which turns out:

$$\gamma\binom{r}{d}k^d \Big/ \binom{s}{d} \geq \gamma^{1-d}\frac{e^{2-d}}{2\sqrt{2}\pi}\sqrt{\frac{\mu}{\mu-\nu}}\left(\frac{\mu^\mu}{(\mu-\nu)^{\mu-\nu}}\right)^k.$$

We prove the theorem by letting the right hand side of above formula equal to 1. The full proof with complete calculation is deferred to Appendix A.4. □

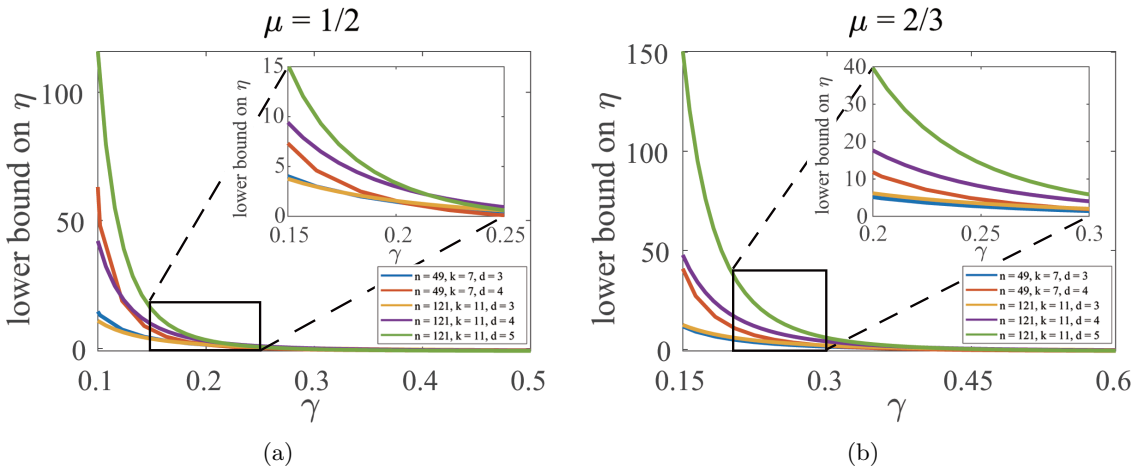

Fig. 4. The lower bound for the efficiency factor η (by Lemma 5) versus the corrupted fraction γ, for $\mu = 1/2$ and $\mu = 2/3$. Recall that $n = k^2$ in $\mathcal{A}_{poly}$ and $\eta < 0$ if the custody scheme is not secure for corresponding γ.

As an instance of Theorem 3, the custody scheme with $n = 49$, $k = 7$, $d = 3$ and $\mu = 1/2$ is secure against adversary with power $\gamma \leq 0.2486$, or equivalently the number of compromised nodes $s \leq 12$.

Fig. 4 depicts the relation between the lower bound of η by Lemma 5 and the adversary corruption factor γ, for $\mu \in \{1/2, 2/3\}$ and n, k, d as shown in the figure. It is easy to see that η increases as n, k and d become larger. For specific choices we get $\eta \geq 9.3637$ against adversary with $\gamma = 3/11$, when $\mu = 2/3$ and $\mathcal{A}_{poly}$ is parameterized by $n = 121$, $k = 11$ and $d = 5$. Furthermore, we remark that the efficiency factor η increases rapidly when γ decreases. For instance, $\eta \geq 61.2690$ when reducing γ from $3/11$ to $2/11$ in the above example. That is, the efficiency factor increased by 5 times at the cost of relaxing the tolerance of 33 corrupted nodes down to 22.

5 Compact Group Assignments via Random Sampling

We notice that in previous constructions a group assignment $\mathcal{A}$ may contain too many custodian groups which makes the induced custody scheme impossible to manage in practice. To mitigate this problem, we propose a randomized sampling technique to construct compact custody schemes with a small number of custodian groups sampled from $\mathcal{A}$ as representative.

Given a group assignment $\mathcal{A}$ consisting of m groups, as well as a sampling rate $\beta \in (0, 1)$, we uniformly sample a subset of βm elements from $\mathcal{A}$ as the new assignment $\mathcal{A}'$, and then construct custody scheme based on $\mathcal{A}'$. This sampling process does not change the authentication threshold μ. In what follows we analyze the efficiency of $\mathcal{A}'$ comparing to $\mathcal{A}$.

For a given corrupted fraction γ, let ξ be a function of γ defined as follows:

$$\xi := -\left(\gamma \log \gamma + (1-\gamma)\log(1-\gamma)\right) \tag{5}$$

The efficiency factor of custody scheme induced by $\mathcal{A}'$ is lower bounded as in the following theorem:

Theorem 4. *Let $\mathcal{A}$ and ξ be defined as above, and suppose the corrupted fraction γ satisfies $n\gamma(1-\gamma) \geq 1$ (which is trivial if $n > 4$ and $\gamma n \geq 2$). Let $\mathcal{A}'$ be the group assignment uniformly sampled from $\mathcal{A}$ with $\beta \in (0, 1)$, and let η' denote the efficiency factor of $\mathcal{A}'$. Then, for arbitrary $c \geq 0$, with probability at least $1 - \frac{e}{2\pi}\exp(-cn\xi)$ the following lower bound for η' holds against the same γ-adversary:*

$$\eta' \geq \frac{\gamma(\eta+1)\cdot\sqrt{\beta m}}{\gamma\cdot\sqrt{\beta m} + (\eta+1)\cdot\sqrt{(1+c)n\xi/2}} - 1.$$

Proof (Sketch). The theorem is proved in two steps.

First, we bound the probability that a specific attacking strategy corrupts more than $\beta f(\gamma; \mathcal{A}, \mu) + \sqrt{n\xi \cdot \beta m}$ in the new assignment $\mathcal{A}'$, where the probability is taken over the random sampling process

of $\mathcal{A}'$. In particular, for an adversary with a fixed corruption set, let $X := X(\mathcal{A}')$ be the random variable denoting the number of groups in $\mathcal{A}'$ corrupted by that adversary. Then, with the hypergeometric tail bound, we lower bound X as follows:

$$\Pr_{\mathcal{A}'}\left[X \geq \beta \cdot f(\gamma; \mathcal{A}, \mu) + \sqrt{\left(\frac{1+c}{2}\right) \cdot n\xi \cdot \beta m}\right] \leq \exp(-(1+c)n\xi).$$

Then, the following inequality holds by applying a union bound on all possible $\binom{n}{\gamma n}$ corrupting sets:

$$\Pr_{\mathcal{A}'}\left[f(\gamma; \mathcal{A}', \mu) \geq \beta \cdot f(\gamma; \mathcal{A}, \mu) + \sqrt{\left(\frac{1+c}{2}\right) \cdot n\xi \cdot \beta m}\right] \leq \frac{e}{2\pi}\exp(-cn\xi).$$

Therefore, with probability at least $1 - \frac{e}{2\pi}\exp(-cn\xi)$ there is $f(\gamma; \mathcal{A}', \mu) \leq \beta f(\gamma; \mathcal{A}, \mu) + \sqrt{n\xi \cdot \beta m}$, from which we conclude the following bound and finish the proof.

$$\eta' \geq \frac{\gamma(\eta+1) \cdot \sqrt{\beta m}}{\gamma \cdot \sqrt{\beta m} + (\eta+1) \cdot \sqrt{(1+c)n\xi/2}} - 1,$$

The complete proof is deferred to Appendix A.5. □

The following corollary follows Theorem 4 immediately.

Corollary 4. *Let $\mathcal{A}$ be a group assignment with efficiency factor η against adversary controlling s corrupted nodes. Let $\mathcal{A}'$ be the sampled group assignment from $\mathcal{A}$ with $(\eta+1)n\xi/\gamma^2$ groups. Then, with probability at least $1 - \frac{e}{2\pi}\exp(-n\xi)$, the custody scheme induced by $\mathcal{A}'$ has efficiency factor $\eta' \geq \sqrt{\eta+1} - 2$ against same adversary with s corrupted nodes.*

To illustrate the effect of the above sampling technique, we present the following two examples with concrete numbers.

Example 3 (Sampling from symmetric designs). Consider $\mathcal{A}_{all}$ with $n = 1,000$ and $k = 25$. Let $\mu = 2/3$ and $s = 428$ (or $\gamma = 0.428$). If we uniformly at random sample $88,695$ groups from $\mathcal{A}_{all}$ to form $\mathcal{A}'$, then by Theorem 4, the sampled custody scheme has efficiency factor $\eta' \geq 5$ with probability $1 - \frac{e}{2\pi} \approx 0.5674$ over the random sampling process. Further, if $s = 388$ (or $\gamma = 0.388$) and $m' = 323,825$ groups, then $\eta' \geq 10$ with probability $1 - \frac{e}{2\pi}$.

Example 4 (Sampling from polynomial designs). Consider $\mathcal{A}_{poly}$ with $n = 961$, $k = 31$ and $d = 12$. Let $\mu = 2/3$ and $s = 314$ (or $\gamma \approx 0.3267$). If we uniformly at random draw $138,767$ groups from $\mathcal{A}_{poly}$ to form $\mathcal{A}'$, then by Theorem 4 there is $\eta' \geq 5$ with probability $1 - \frac{e}{2\pi} \approx 0.5674$ over the random sampling process. Further, if $s = 285$ (or $\gamma \approx 0.2966$) and sampling $m' = 481,017$ groups, then $\eta' \geq 10$ with probability $1 - \frac{e}{2\pi}$.

6 Hardness Results

In this section, we consider the hardness issue of finding the best corrupting scheme. In general, with a common group assignment scheme $\mathcal{A}$, it is hard for the polynomial-time adversary to figure out the best attack strategy.

Theorem 5. *Take k and r as predetermined constants, then, with the group assignment scheme $\mathcal{A}$ which includes n nodes and the group number m be within a polynomial in n, and the adversary controlling $s := \gamma n = n^{\Omega(1)}$ nodes, we have the following results:*

- *When $1 \leq r < k$, it is NP-hard to find the optimal attacking strategy.*
- *When $r = k$, under the Small Set Expansion Conjecture [24], it is NP-hard to find the optimal attacking strategy.*

Proof (Sketch). To show the NP-hardness of the problem with $1 < r < k$, we consider the MAX s-VC problem; for the case of $r = k$, we consider the fixed cost minimum edge cover (FCMC) problem defined in [14]. For the details on the reduction step, please refer to Appendix A.6. □

Theorem 5 shows that generally, it is hard for the computationally bounded adversary to figure out the best attacking strategy. Nevertheless, there is a deterministic algorithm for the adversary to figure out a strategy which corrupts at least an average amount of groups. Let $\kappa := \Pr[\mathcal{H}(n, s, k) \geq r]$, recall that κm is the expected amount of groups the adversary can corrupt when the γn corrupted nodes are chosen uniformly at random. We have the following theorem:

Theorem 6. *Let κ be defined as above. Suppose computing each binomial coefficient takes time $O(1)$. Then there is a deterministic algorithm that gives an attacking strategy which corrupts at least κm groups, running in time $O(kmn)$.*

Proof (Sketch). Specifically, we construct an $O(kmn)$ deterministic algorithm (see Algorithm 1) and prove that the algorithm always return a corrupting strategy no beneath the average. Specifically, in each step, we judge whether or not to corrupt a node conditioning on previous choices. We prove that after each step before the algorithm ends, the conditional expectation of the number of corrupted groups is no less than average given previous decisions. Hence the algorithm always returns a solution no worse than average. For the construction of the algorithm and full proof, please see Appendix A.7.

7 Summary and Discussion

In this work we propose a framework of decentralized custody schemes based on overlapping group assignments. This custody scheme allows better efficiency as well as stronger liveness, and its security is guaranteed in a game-theoretic sense without resorting to any central authorities or trusted party. We believe such decentralized custody scheme will find applications in digital finance as well as DeFi on blockchains.

Future improvements of this work include explicit constructions of compact assignments with much less custodian groups, efficient approximation algorithms for estimating the actual efficiency factor of a given custody scheme in our framework, and more rigorous analysis of liveness guarantee as well as the tradeoff between liveness and security.

References

1. DeFi Pulse: Total Value Locked in DeFi, https://defipulse.com/
2. Data Innovation Privacy & Cybersecurity: VinDAX is the Seventh Cryptocurrency Exchange Hacked This Year: What Should Investors Be Considering? https://www.paulweiss.com/practices/litigation/data-innovation-privacy-cybersecurity/publications/vindax-is-the-seventh-cryptocurrency-exchange-hacked-this-year-what-should-investors-be-considering?id=30259
3. Ageev, A.A., Sviridenko, M.: Pipage rounding: A new method of constructing algorithms with proven performance guarantee. J. Comb. Optim. **8**(3), 307–328 (2004)
4. Bellare, M., Neven, G.: Multi-signatures in the plain public-key model and a general forking lemma. In: Proceedings of the 13th ACM conference on Computer and communications security. pp. 390–399 (2006)
5. BitcoinWiki: Atomic Swap. https://en.bitcoin.it/wiki/Atomic_swap
6. BitcoinWiki: Hashlock. https://en.bitcoin.it/wiki/Hashlock
7. Boneh, D., Gennaro, R., Goldfeder, S.: Using level-1 homomorphic encryption to improve threshold DSA signatures for bitcoin wallet security. In: Progress in Cryptology - LATINCRYPT 2017 - 5th International Conference on Cryptology and Information Security in Latin America, Havana, Cuba, September 20-22, 2017, Revised Selected Papers. pp. 352–377 (2017)
8. Chvátal, V.: The tail of the hypergeometric distribution. Discrete Mathematics **25**(3), 285–287 (1979)
9. Conflux: Conflux Shuttleflow: A Cross-Chain Asset Protocol. https://medium.com/conflux-network/conflux-shuttleflow-a-cross-chain-asset-protocol-15ad6b2a9539
10. Cornuejols, G., Fisher, M.L., Nemhauser, G.L.: Exceptional paper—location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Management science **23**(8), 789–810 (1977)
11. Feige, U.: A threshold of ln n for approximating set cover. J. ACM **45**(4), 634–652 (1998)
12. Feige, U., Langberg, M.: Approximation algorithms for maximization problems arising in graph partitioning. J. Algorithms **41**(2), 174–211 (2001)
13. Gagol, A., Kula, J., Straszak, D., Swietek, M.: Threshold ECDSA for decentralized asset custody. IACR Cryptol. ePrint Arch. **2020**, 498 (2020)
14. Gandhi, R., Kortsarz, G.: On set expansion problems and the small set expansion conjecture. Discrete Applied Mathematics **194**, 93–101 (2015)

15. HBTC.Finance: Hbtc whitepaper: Bridge between centralized and defi markets. https://www.hbtc.finance/static/pdf/whitepaper-en.pdf
16. Hochbaum, D.S.: Approximation Algorithms for NP-Hard Problems. PWS Publishing Co., USA (1996)
17. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (1963)
18. Katz, J., Koo, C.: On expected constant-round protocols for Byzantine agreement. J. Comput. Syst. Sci. **75**(2), 91–112 (2009)
19. Keep.Network: tBTC: A decentralized redeemable BTC-backed ERC-20 token. https://docs.keep.network/tbtc/index.pdf
20. Maxwell, G., Poelstra, A., Seurin, Y., Wuille, P.: Simple Schnorr multi-signatures with applications to Bitcoin. Designs, Codes and Cryptography **87**(9), 2139–2164 (2019)
21. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008)
22. Pass, R., Shi, E.: The sleepy model of consensus. In: Takagi, T., Peyrin, T. (eds.) Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part II. Lecture Notes in Computer Science, vol. 10625, pp. 380–409. Springer (2017)
23. Pease, M.C., Shostak, R.E., Lamport, L.: Reaching agreement in the presence of faults. J. ACM **27**(2), 228–234 (1980)
24. Raghavendra, P., Steurer, D., Tulsiani, M.: Reductions between expansion problems. In: Proceedings of the 27th Conference on Computational Complexity, CCC 2012, Porto, Portugal, June 26-29, 2012. pp. 64–73 (2012)
25. renProject: renProject - wiki. https://github.com/renproject/ren/wiki
26. Robbins, H.: A remark on Stirling's formula. The American mathematical monthly **62**(1), 26–29 (1955)
27. Schnorr, C.P.: Efficient signature generation by smart cards. Journal of cryptology **4**(3), 161–174 (1991)
28. Stinson, D.: Combinatorial designs: constructions and analysis. Springer Science & Business Media (2007)
29. Tokenlon: Tokenlon dex now on web. https://tokenlon.zendesk.com/hc/en-us/articles/360037584171-Tokenlon-DEX-now-on-Web
30. wBTC.Network: Wrapped tokens: A multi-institutional framework for tokenizing any asset. https://wbtc.network/assets/wrapped-tokens-whitepaper.pdf
31. Wikipedia: Mt. Gox. https://en.wikipedia.org/wiki/Mt._Gox
32. Wood, G.: Polkadot: Vision for a heterogeneous multi-chain framework. White Paper (2016)
33. Wood, G., et al.: Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper **151**(2014), 1–32 (2014)

Appendix A Proof of lemmas and theorems

A.1 Proof of Lemma 3

Proof. For simplicity, let $s := \gamma n$ be the number of nodes that the adversary corrupts, and

$$\eta_s = \frac{s}{n} \frac{\binom{n}{k}}{\sum_{t=r}^{k} \binom{s}{t}\binom{n-s}{k-t}} - 1$$

be the efficiency factor of the system in this case. We will compare η_s with η_{s-1}. Specifically, we compare every corresponding pair of terms in the sum. For any $r \le t \le k$, we have

$$\begin{aligned} s \cdot \frac{\binom{n}{k}}{\binom{s}{t}\binom{n-s}{k-t}} \Big/ (s-1) \cdot \frac{\binom{n}{k}}{\binom{s-1}{t}\binom{n-s+1}{k-t}} &= \frac{(n-s-k+t)!(s-t)!}{(n-s)!(s-1)!} \Big/ \frac{(n-s-k+t+1)!(s-t-1)!}{(n-s+1)!(s-2)!} \\ &= \frac{(s-t)(n-s+1)}{(n-s-k+t+1)(s-1)}, \end{aligned}$$

and

$$\begin{aligned} & \frac{(s-t)(n-s+1)}{(n-s-k+t+1)(s-1)} \le 1 \\ \iff & (s-t)(n-s+1) \le (n-s-k+t+1)(s-1) \\ \iff & s(k-1) \le (t-1)n + k - 1 \\ \impliedby & \ s \le \frac{\mu k - 1}{k-1} n + 1. \end{aligned}$$

Here, the second inequality is due to $s \le n-k$, while the fourth inequality is due to $t > \mu k$. The result implies that $\eta_s \le \eta_{s-1}$ always holds when $s \le \frac{\mu k-1}{k-1} n + 1$, which proves the lemma as $s = \gamma n$. □

A.2 Proof of Theorem 1

Proof. We will first prove that the system is reliable when $k \ge 10$ and $\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k} \le \gamma \le \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}$, and combine the result with Lemma 3 to prove the theorem. Note here that as $n \ge 2k$, there must exist some γ in $[\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k}, \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}]$ such that γn is an integer.

Recall that the system is γ-reliable as long as

$$\gamma \ge \sum_{\mu k < t \le k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}},$$

by the tail bound of hypergeometric distribution (see Section 2) and that $\gamma < \mu$, we have

$$\sum_{\mu k < t \le k} \frac{\binom{\gamma n}{t}\binom{n-\gamma n}{k-t}}{\binom{n}{k}} \le \exp(-kD(\mu\|\gamma)) \le \left(4\gamma(1-\gamma)\right)^{k/2}.$$

The last equality is due to

$$D(\mu\|\gamma) > D(\frac{1}{2}\|\gamma) = \frac{1}{2}\ln\frac{1}{2\gamma} + \frac{1}{2}\ln\frac{1}{2(1-\gamma)} = \frac{1}{2}\ln\frac{1}{4\gamma(1-\gamma)},$$

as $\gamma < 1/2 < \mu$. (See Section 2.)

When $\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k} \le \gamma \le \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}$ and $k \ge 10$,

$$\begin{aligned} \left(4\gamma(1-\gamma)\right)^{k/2} &\le \left(1 - 4\left(\frac{2\sqrt{k}+1}{2k}\right)^2\right)^{k/2} \\ &\le \exp\left(-\frac{(2\sqrt{k}+1)^2}{2k}\right) \\ &\le \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k} \le \gamma, \end{aligned}$$

The second inequality is due to $(1-1/x)^x \leq 1/e$ for any $x > 1$. This implies that the system is γ-reliable when $\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k} \leq \gamma \leq \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}$. Note that when $\mu > 1/2$, $n \geq 2k$ and $k \geq 10$,

$$\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k} \leq \frac{1}{2} - \frac{1}{2(k-1)} + \frac{1}{n} < \min\left(\frac{\mu k - 1}{k-1} + \frac{1}{n}, 1 - \frac{k}{n}\right),$$

with Lemma 3, the result is reached. □

A.3 Proof of Theorem 2

Proof. When $\binom{s}{r} \leq \binom{n}{r} \Big/ \binom{k}{r}$, by Lemma 4, we have the following lower bound on the efficiency factor of the system:

$$\eta \geq \frac{s}{n} \frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} - 1.$$

For the system to be reliable, it is sufficient if we have

$$\frac{s}{n} \frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} \geq 1.$$

Note that

$$\begin{aligned}
\frac{s}{n} \frac{\binom{n}{r}}{\binom{k}{r}\binom{s}{r}} &= \frac{s}{n} \frac{n!(s-r)!}{(n-r)!s!} \Big/ \binom{k}{r} \\
&= \frac{n-1}{s-1} \cdot \cdots \cdot \frac{n-r+1}{s-r+1} \cdot \frac{r}{k} \cdot \frac{r-1}{k-1} \cdot \cdots \cdot \frac{1}{k-r+1} \\
&= \prod_{t=1}^{r-1} \frac{(n-t)(r-t)}{(s-t)(k-t)} \cdot \frac{r}{k} \\
&> \prod_{t=1}^{r-1} \frac{(n-t)(r-t)}{(s-t)(k-t)} \cdot \mu.
\end{aligned}$$

which is satisfied if we have

$$\frac{(n-t)(r-t)}{(s-t)(k-t)} \geq \mu^{-\frac{1}{r-1}}, \quad \forall 1 \leq t \leq r-1.$$

Let $c := \mu^{-\frac{1}{r-1}} > 1$, the above condition is equivalent to

$$c(s-t)(k-t) - (n-t)(r-t) \leq 0, \quad \forall 1 \leq t \leq r-1,$$

and by the property of quadratic functions, we only need to work on the case of $t = 1$ and $t = r-1$.

When $t = 1$, the condition becomes

$$\frac{s-1}{n-1} \leq \frac{r-1}{c(k-1)},$$

note that $\frac{s-1}{n-1} < \frac{s}{n} = \gamma$. Hence it is sufficient with $\gamma \leq \frac{r-1}{c(k-1)}$. When $\mu \geq 1/2$, $r \geq 2$ and $n \geq 3k-3$, we have $c = \mu^{-\frac{1}{r-1}} < 3/2$. Therefore,

$$\begin{aligned}
s \leq \frac{n}{c \cdot k} + r - 1 &\Longrightarrow \gamma \leq \frac{1}{c \cdot k} + \frac{r-1}{n} \\
&\Longrightarrow \gamma \leq \frac{1}{c(k-1)} + \frac{2(r-2)}{n} \\
&\Longrightarrow \gamma \leq \frac{1}{c(k-1)} + \frac{2(r-2)}{3(k-1)} \\
&\Longrightarrow \gamma \leq \frac{1}{c(k-1)} + \frac{r-2}{c(k-1)} \\
&\Longrightarrow \gamma \leq \frac{r-1}{c(k-1)}.
\end{aligned}$$

When $t = r - 1$, the condition becomes

$$n - r + 1 \geq c(s - r + 1)(k - r + 1),$$

or

$$s \leq \frac{n - r + 1}{c(k - r + 1)} + r - 1.$$

which establishes when $s \leq \frac{n}{ck} + r - 1$, as $\frac{n}{k} \leq \frac{n-r+1}{k-r+1}$. □

A.4 Proof of Theorem 3

Proof. As per Lemma 5, it is sufficient if we have $\gamma \binom{r}{d} k^d / \binom{s}{d} \geq 1$.

As we have

$$\begin{aligned}
\gamma \binom{r}{d} k^d \Big/ \binom{s}{d} &> \gamma \binom{\mu k}{\nu k} k^{\nu k} \Big/ \binom{\gamma k^2}{\nu k} \\
&= \gamma \frac{(\mu k)!(\gamma k^2 - \nu k)! k^{\nu k}}{(\gamma k^2)!(\mu k - \nu k)!} \\
&\geq \gamma \frac{e^2}{2\pi} \frac{(\mu k)^{\mu k + \frac{1}{2}} (\gamma k^2 - \nu k)^{\gamma k^2 - \nu k + \frac{1}{2}} k^{\nu k}}{(\gamma k^2)^{\gamma k^2 + \frac{1}{2}} (\mu k - \nu k)^{\mu k - \nu k + \frac{1}{2}}} \\
&= \gamma \frac{e^2}{2\pi} \sqrt{\frac{\mu(\gamma k - \nu)}{\gamma(\mu - \nu)k}} \frac{(\mu k)^{\mu k} k^{\nu k}}{(\mu k - \nu k)^{\mu k - \nu k} (\gamma k^2)^{\nu k}} \left(1 + \frac{\nu k}{\gamma k^2 - \nu k}\right)^{\nu k - \gamma k^2} \\
&\geq \gamma^{1-d} \frac{e^{2-d}}{2\sqrt{2}\pi} \sqrt{\frac{\mu}{\mu - \nu}} \left(\frac{\mu^\mu}{(\mu - \nu)^{\mu - \nu}}\right)^k,
\end{aligned}$$

the theorem is proved by having the last formula no less than 1. Here, we use Stirling's formula (see Section 2) to estimate the factorial term in the third line. □

A.5 Proof of Theorem 4

Proof. To show the result, first consider a specific attacking strategy from the adversary who precisely corrupts γn nodes. We call the strategy $\mathcal{M}$. Denote $g(\mathcal{M})$ be the corrupted groups with strategy $\mathcal{M}$ under the group assignment scheme $\mathcal{A}$. Clearly, $g(\mathcal{M}) \leq f(\gamma; \mathcal{A}, \mu)$ by definition.

Now suppose we uniformly draw βm groups from $\mathcal{A}$ to obtain $\mathcal{A}'$, and let X be a random variable indicating the number of corrupted groups in the new scheme $\mathcal{A}'$. By hypergeometric tail bound (see Section 2), we have

$$\begin{aligned}
\Pr\left[X \geq \left(\frac{f(\gamma; \mathcal{A}, \mu)}{m} + \delta\right) \cdot \beta m\right] &= \Pr\left[X \geq \left[\frac{g(\mathcal{M})}{m} + \left(\frac{f(\gamma; \mathcal{A}, \mu) - g(\mathcal{M})}{m} + \delta\right)\right] \cdot \beta m\right] \\
&\leq \exp\left(-2\beta m \cdot \left(\frac{f(\gamma; \mathcal{A}, \mu) - g(\mathcal{M})}{m} + \delta\right)^2\right) \\
&\leq \exp(-2\beta m \cdot \delta^2).
\end{aligned}$$

Here, $\delta > 0$ is a parameter to be determined. The probability is over all possible choices of $\mathcal{A}'$. Let $f(\gamma; \mathcal{A}', \mu)$ denote the maximal number of groups that can be corrupted with γn compromised nodes in the new scheme $\mathcal{A}'$, by a union bound, we have

$$\Pr\left[f(\gamma; \mathcal{A}', \mu) \geq \beta f(\gamma; \mathcal{A}, \mu) + \delta \cdot \beta m\right] \leq \exp(-2\beta m \cdot \delta^2) \binom{n}{\gamma n}.$$

By Stirling's formula (see Section 2),

$$\begin{aligned}\binom{n}{\gamma n} &= \frac{n!}{(\gamma n)!(n-\gamma n)!} \\ &\leq \frac{e}{2\pi}\frac{n^{n+\frac{1}{2}}}{(\gamma n)^{\gamma n+\frac{1}{2}}(n-\gamma n)^{n-\gamma n+\frac{1}{2}}} \\ &= \frac{e}{2\pi}\sqrt{\frac{1}{n\gamma(1-\gamma)}}\cdot\left(\frac{1}{\gamma^{\gamma}(1-\gamma)^{1-\gamma}}\right)^{n} \\ &\leq \frac{e}{2\pi}\left(\frac{1}{\gamma^{\gamma}(1-\gamma)^{1-\gamma}}\right)^{n}.\end{aligned}$$

Hence,

$$\Pr[f(\gamma;\mathcal{A}',\mu) \geq \beta f(\gamma;\mathcal{A},\mu) + \delta\cdot\beta m] \leq \exp(-2\beta m\cdot\delta^2)\binom{n}{\gamma n} \leq \frac{e}{2\pi}\exp(-2\beta m\cdot\delta^2 + n\xi).$$

Let $\beta m\cdot\delta^2 = \frac{1+c}{2}\cdot n\xi$, then w.p. no less than $1-\frac{e}{2\pi}\exp(-cn\xi)$, we have $f(\gamma;\mathcal{A}',\mu) \leq \beta f(\gamma;\mathcal{A},\mu)+\delta\cdot\beta m$. Under such case, the efficiency factor η' of $\mathcal{A}'$ with malicious rate γ is lower bounded by

$$\begin{aligned}\eta' &= \gamma\frac{\beta m}{f(\gamma;\mathcal{A}',\mu)} - 1 \\ &\geq \gamma\frac{m}{f(\gamma;\mathcal{A},\mu)+\delta\cdot m} - 1 \\ &= \gamma\frac{m}{\frac{\gamma m}{\eta+1}+\delta\cdot m} - 1 \\ &= \gamma\frac{\eta+1}{\gamma+\delta\cdot(\eta+1)} - 1 \\ &= \frac{\gamma\sqrt{\beta m}(\eta+1)}{\gamma\sqrt{\beta m}+\sqrt{(1+c)n\xi/2}(\eta+1)} - 1.\end{aligned}$$

The third line is due to the definition of η:

$$\eta = \gamma\cdot\frac{m}{f(\gamma;\mathcal{A},\mu)} - 1.$$

□

A.6 Proof of Theorem 5

Proof. First of all, it is trivial that the given problem is harder than only to determine the number of corrupted groups under the best attacking strategy. We only consider the latter in this proof. Technically, we will work on the two cases, $r < k$ and $r = k$ respectively.

- $r < k$. We first deal with the case when $r = 1$. In fact, in this case, when $k = 2$, then finding the optimal attacking strategy is naturally equivalent to the MAX s-VC problem in common graphs. Here, the MAX s-VC problem is to compute the maximum number of edges that s vertices can cover given a graph of size n. This problem is known to be NP-hard.
  When $k > 2$, the problem is equivalent to the MAX s-VC problem in k-uniform hypergraphs. Here, a k-uniform hypergraph is a hypergraph in which each edge contains exactly k vertices. We reduce the MAX s-VC problem in common graphs to this problem. Specifically, consider a realization of the problem in the graph $G = (V, E)$ with m edges. We transfer G to a k-uniform hypergraph $G' = (V', E')$ by adding $(k-2)m$ vertices to V. Denote these vertices as $v_1^1, v_1^2, \cdots, v_1^m, \cdots, v_{k-2}^1, v_{k-2}^2, \cdots, v_{k-2}^m$. For each edge e_i in G, we add $v_1^i, v_2^i, \cdots, v_{k-2}^i$ to e_i and obtain an edge e_i' in E' with k vertices. Now consider the MAX s-VC solution in G'. Obviously, there is an optimum which contains no point in $V' - V$, as any point in $V' - V$ appears in only one edge, and substituting a selected point in $V' - V$ with an unselected point in the same edge that belongs to V will cause no loss on the objective function. If all vertices in that edge that

belongs to $V' - V$ are selected, we can alternatively pick any non-chosen vertices in $V' - V$, still with no loss on the goal. Therefore, the modified instance in the k-uniform hypergraph owns the same objective value with the original problem in common graphs. Note that the reduction step itself is polynomial as m is a polynomial of n. At the same time, $s = (n + (k-2)m))^{\Omega(1)}$ since k is a constant and m is within a polynomial of n. Hence, the MAX s-VC problem in k-uniform hypergraphs is NP-hard as well.
Here, we should mention that the MAX s-VC problem in k-uniform hypergraphs ($k \geq 2$) is indeed a special case of the MAX s-Cover problem, with each element existing in precisely k sets. The reduction is to deem each vertex as a set, including all edges incident to the vertex. There is a simple $(1 - e^{-1})$-approximation greedy algorithm for the MAX s-Cover problem [10, 16]. Furthermore, it is known that in general, for any $\epsilon > 0$, there is no $(1 - e^{-1} - \epsilon)$-approximation for this problem unless P = NP [11]. Nevertheless, concerning MAX s-VC in common graphs, there is a $(0.75 + \delta)$-approximation algorithm for some constant $\delta > 0$ [12]. Meanwhile, the MAX s-VC problem in k-uniform with $k \geq 3$ can also be approximated strictly within the ratio 0.75 [3]. When $k > r > 1$, the original problem is equivalent to the following problem:

Problem 1. For a k-uniform hypergraph $G = (V, E)$ where $|V| = n$, we say an edge is covered only if at least r vertices incident to the edge is selected. With s vertices, compute the maximal number of edges that can be covered.

we reduce the problem of finding a MAX $(s-r+1)$-VC problem in $(k-r+1)$-uniform hypergraphs to this problem. Again, consider an instance $G = (V, E)$ of the latter problem. We add $(r-1)$ vertices to V, as well as including them in each hyperedge in E to achieve $G' = (V', E')$, which is an instance in Problem 1. Consider the optimal solution in this problem. We claim that there always exists an optimum that concludes the appended $(r-1)$ vertices, as changing any chosen vertex in V to a point in $V' - V$ will never lessen the objective value. Under such optimum, an edge in G' is covered if and only if the corresponding edge in G is covered in the MAX $(s - r + 1)$-VC problem in $(k - r + 1)$-uniform hypergraph. Therefore, the two instances own the same objective value in their respective questions. Furthermore, such reduction is polynomial-bounded, which leads to the NP-hardness of Problem 1.

– $r = k$. When $k = 2$, the problem becomes

Problem 2. Given a common graph $G = (V, E)$ where $|V| = n$ and $s < n$, compute the maximal number of edges that s vertices can *completely* cover. Here, an edge is completely covered if and only if both endpoints are chosen.

Consider a optimal solution C with $|C| = s$, then the number of edges that are incident with any vertex in $V - C$ is minimized. Hence, Problem 2 is bidirectional-reducible to the following problem:

Problem 3. Given a common graph $G = (V, E)$ where $|V| = n$ and $s < n$, compute the minimal number of edges that $n - s$ vertices can cover.

Problem 3 is the uniform-weighted case of the fixed cost minimum edge cover (FCMC) problem defined in [14], which is proved to be NP-hard even to approximate with a factor $(2-\epsilon)$ under the Small Set Expansion Conjecture in the same paper. Hence, under such assumption, Problem 2 is hard to solve. For the case when $k > 2$, we can reduce Problem 2 with k-uniform hypergraphs to the corresponding problem with the similar reduction as what we do with Problem 1.

□

A.7 Proof of Theorem 6

Proof. We give the algorithm in Algorithm 1. Specifically, in step $1 \leq i \leq n$ (Line 2), given a temporary corrupted list T and honest list N, the adversary decides whether or not to corrupt node i. To figure this out, the adversary needs to compute the expected amount of corrupted groups X_i conditioning on nodes in $T \cup \{i\}$ already compromise and nodes in N are honest (Line 3). In detail, the adversary should compute the conditional probability on each group is corrupted and take a sum to derive X_i. If $X_i \geq \kappa m$, node i will be included in T (Line 5), else it will be given up by the adversary (Line 7). The algorithm ends whenever s nodes are already chosen (Line 9) or $n - s$ nodes

Algorithm 1 Corrupting at least an average amount of groups.

Input: The node set $S = \{1, 2, \cdots, n\}$, the group assignment scheme $\mathcal{A}$ with parameters m and k, least number of nodes to control a group r, number of corrupted nodes $s = \gamma n$.

Output: $T \subseteq S$ with $|T| = s$, such that when T is corrupted, the adversary can at least control κm groups, with κ defined as $\kappa := \Pr[\mathcal{H}(n, s, k) \geq r]$.

1: $T = \emptyset, N = \emptyset$
2: **for** $i = 1$ to n **do**
3: Compute the expected amount of compromised groups X_i conditioning on nodes in $T \cup \{i\}$ are malicious and nodes in N are honest
4: **if** $X_i \geq \kappa m$ **then**
5: $T = T \cup \{i\}$
6: **else**
7: $N = N \cup \{i\}$
8: **end if**
9: **if** $|T| \geq s$ **then**
10: **return** T
11: **end if**
12: **if** $|N| \geq n - s$ **then**
13: **return** $S - N$
14: **end if**
15: **end for**

are already given up (Line 12). Algorithm 1 runs in time $O(kmn)$ as there are at most n steps, and in each step, one needs to compute the conditional probability for each of m groups, and Computing each conditional probability gives the time complexity of $O(k)$.

Clearly, Algorithm 1 is sure to end up with a size-s subset T of $S = \{1, 2, \cdots, n\}$. To show that corrupting T will lead to at least κm groups controlled by the adversary, we need the following lemma.

Lemma 6. *Suppose Algorithm 1 does not end after step i. Let $Y_i (i \geq 0)$ be the expected amount of corrupted nodes conditioning on T is corrupted and N is honest after step i, then $Y_{i+1} \geq Y_i$. Specifically, we have $Y_0 = \kappa m$.*

Proof. To show this lemma, Let X be a random variable denoting the number of corrupted groups. Furthermore, let $a_i < s$ and $b_i < n - s$ be respectively the size of T_i and N_i after step i, $a_i + b_i = i$. Then we have the following equality:

$$Y_i = \mathbb{E}[X|T_i, N_i] = \frac{s - a_i}{n - i} \cdot \mathbb{E}[X|T_i \cup \{i+1\}, N_i] + \frac{n - s - b_i}{n - i} \cdot \mathbb{E}[X|T_i, N_i \cup \{i+1\}].$$

Here, $\mathbb{E}[X|T_i, N_i]$ is for the expectation of X conditioning on T_i malicious and N_i honest. Hence, at least one of the $\mathbb{E}[X|T_i \cup \{i+1\}, N_i]$ and $\mathbb{E}[X|T_i, N_i \cup \{i+1\}]$ is no less than Y_i. According to Line 4, $Y_{i+1} \geq Y_i$. □

With Lemma 6, note that $Y_0 = \kappa m$, then the theorem is proved. □