Decentralized Custody Scheme with Game-Theoretic Security

Zhaohua Chen[1,2] and Guang Yang[2]

[1] School of Electronics Engineering and Computer Science, Peking University
[2] Conflux
chenzhaohua@pku.edu.cn
guang.yang@conflux-chain.org

Abstract. Custody 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: Blockchain application · Decentralized asset custody · 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 blockchain and cryptocurrencies. Cryptocurrencies enjoy two major advantages over their physical counterparts: (1) they are intrinsically integrated with information technology such as Internet and modern cryptography, which technically enables multiple custodians to safeguard assets collectively; (2) with the underlying blockchain as a public ledger, the management of cryptocurrencies becomes transparent to everyone and hence any fraud behavior will be discovered immediately, which makes prosecution much easier.

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 [25] and Ethereum [37], 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 [35]; 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 [3]. However, it is difficult for customers to distinguish that whether the claimed loss was caused by hacker attack or internal fraud and embezzlement, and therefore 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 assets mapping* service (a.k.a. *cross-chain portable assets* [9,38]) 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 on the Bitcoin network. Nowadays the volume of cryptoassets invested into Ethereum DeFi applications has been increasing at an extraordinary pace and broke nine billion dollars recently [2], among which a significant fraction (e.g. hBTC [19], imBTC [33], tBTC [23], wBTC [34], renBTC [29], 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 cryptoassets custody turns out an attractive approach for better security and credit enhancement.

In this work, we propose a framework of decentralized asset custody scheme designed for cross-chain assets mapping (especially from blockchains with poor programmability, e.g. Bitcoin), with security 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 all assets under custody. The authentication of each custodian group requires consent of sufficiently many group members, which can be implemented with voting or threshold signature. Under this framework transactions can be processed more efficiently within the very few 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 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, which is kept together with the asset under custody and will be used to compensate any loss caused by misbehaving custodians. 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 propose several group assignment constructions for the decentralized custody scheme. Furthermore we prove that for an adversary who corrupts a limited fraction of custodians, our scheme can safeguard customer assets of multiple times value of the total security deposit. This approach significantly reduces the financing cost of a collateralized custody service.

1.1 Related Works

The prototype of decentralized custody scheme first appears in Bitcoin as multisignature (multisig) [7], 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 can hardly scale out, because the computational and communicational cost of authenticating and validating each transaction grows linearly in M. Both efficiency and liveness of the scheme are compromised for large M and N, especially in the sleepy model proposed by Pass and Shi [26] where key holders do not always response in time. In practice, multisignature scheme is typically used at the 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 [8, 16] or aggregate signature [5, 24, 31] 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 is not reduced. Furthermore we remark that these advanced signature techniques are also compatible with our scheme, when realized as authentication mechanism for every single custodian group.

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. One approach is to encapsulate light client verification inside smart contracts, e.g. BTCRelay [1] is an Ethereum contract capable of verifying Bitcoin block headers. However, such light client verification is only possible on blockchains with sufficiently high programability and throughput, i.e. BTCRelay only provides one-way communication from Bitcoin to Ethereum whereas a symmetric Ethereum light client on top of the Bitcoin ledger is completely impossible. Another line of techniques such as Atomic Swap [6] and Hashlock/Hash TimeLock Contracts (HTLC) [7] enables code-based cross-chain interoperability in a much simpler way, however the recipients and transferred value must be specified in advance, which severely limits the use cases and hence unable to implement general purpose custody service.

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

- Centralized: custody in a trusted central authority, with the endorsement fully from that authority, e.g. hBTC [19], wBTC [34] and imBTC [33];
- Consortium: custody in multisignature accounts controlled by an alliance of members, and endorsed by the reputation of alliance members, e.g. cBTC [12] (in its current version) and Polkadot [36];
- Decentralized (with deposit/collateral): custody provided by permissionless custodians, with security guaranteed by over-collateralized cryptoassets, e.g. tBTC [23] and renBTC [29] (in its future plan).

Although the last type seems satisfiable in decentralization and security against single point failure and collusion, significant drawbacks exist as well: the first drawback is the inefficiency caused by 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. We remark that [18] studies dynamically adjusting the deposit of custodians in the long run. However, it relies on an implicit assumption that the security of the system is irrelevant with the behavior of custodians under some certain assumption (e.g.

by introducing cryptographic methods like in Bitcoin), which is inapplicable in the game-theoretic setting we discuss in this work.

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 partitions 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 suffers from poor capital efficiency. More importantly, it cannot support homogeneous collateral as the assets under custody, since otherwise an adversary corrupting a single group would be able to take more than his own collateral.

1.2 Our Contributions

In this work, we propose the framework of decentralized asset custody service based on *overlapping* group assignments, and investigate performance of specific constructions under this framework. Furthermore, we introduce a random sampling technique capable of reducing the complexity of the group assignments significantly.

In our framework of decentralized asset custody: assets under custody are distributed into many (possibly overlapping) groups of custodians, where each group is fully controlled by its members and hence capable of processing transactions independently from other groups. Therefore, every customer-requested transaction only requires attention and approval from custodians in a specific group (rather than a majority of all participants in the scheme), which significantly reduces the cost and latency of transaction processing. Furthermore this framework provides better liveness guarantee, since one alive custodian group is able to process transactions while the vast majority of custodians in the system may stay offline.

The proposed decentralized custody scheme achieves security in the game-theoretic sense: for any bounded adversary corrupting a small fraction of individual custodians, the fraction of corrupted custodian groups is even smaller. Thus, the adversary cannot truly jeopardize the system, since the total security deposit of corrupted custodians outweighs the total assets held by corrupted groups, and hence funds stolen by such a bounded adversary can be easily compensated with his own security deposits.[3]

The model of our decentralized custody scheme is formalized in Section 3, where the notion of efficiency factor is introduced to measure the capital efficiency. The efficiency factor refers to the maximal capability of safely holding assets for exterior customers against a bounded adversary. The optimal efficiency is achieved when assets under custody are evenly distributed among all custodian groups. In case of an uneven distribution, the custody scheme gracefully degrades as if there are more assets under custody and hence a greater efficiency factor is required, i.e. the scheme is secure against a weaker adversary.

Specific constructions of the underlying group assignments are proposed and analyzed in Section 4. These constructions are based on two kinds of combinatorial designs and the enumeration of all subsets of a fixed size respectively. For example, we show that 20 custodians can be assigned to $38,760$ groups such that as long as the adversary corrupts $\gamma \leq 2/5$ fraction of all custodians, the custody scheme is capable of safekeeping assets worthy of $\eta > 20$ times of total collateral, i.e. the scheme can safely hold over 100 units of value with total security deposit worthy of 5 units.

[3] One may wonder whether such misbehavior is always easy to detect in practice. Indeed, the detection is quite straightforward when the custody service is implemented on a blockchain, where all instructions and actions are publicly examinable.

Furthermore, using the polynomial-based combinatorial design, the decentralized custody scheme assigns 121 custodians into 161,051 groups each containing 11 members, such that the efficiency factor $\eta > 60$ against any adversary corrupting no more than 22 individual custodians. In general, the efficiency factor is better for more custodians and groups, however complicated assignment with too many groups may be infeasible to manage in practice.

To mitigate the problem of too many groups in an assignment, we consider random sampling on the group assignment in Section 5, and prove that the sampling trick significantly reduces the size of group assignment without losing too much in the efficiency factor. More specifically, if there is a custody scheme consisting of n participants and its efficiency factor is η against an adversary corrupting at most γ faction of all custodians, then by randomly sampling $O\left(\eta n \cdot \log(\gamma^{-1})/\gamma\right) \sim O(\eta n)$ many groups, the newly induced custody scheme would have efficiency factor $\eta' \geq \sqrt{\eta+1}-2$ against the same adversary with high probability. Concretely, using the random sampling technique we are able to construct a custody scheme of 1,000 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.

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 (From [11, 21]). *Let $X \sim \mathcal{H}(N, K, n)$ and $p = K/N$, then we have the following:*

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

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 a good estimation of factorial terms. Specifically, we will use the following version.

Lemma 2 (From [30]). *For any integer $n \geq 1$, we have $\sqrt{2\pi} \cdot e^{-n} \cdot n^{n+\frac{1}{2}} < n! < e^{1-n} \cdot n^{n+\frac{1}{2}}$.*

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

Definition 1 (Block design, from [32], with notation revised). *Let n, k, λ and t be positive integers such that $n > k \geq t$. A block design $(T, \mathcal{B})$ is called an t-(n, k, λ)-design if T is a set with $|T| = n$ and $\mathcal{B}$ is a family of k-element subsets of T (called blocks), such that every t-element subset of T is contained in exactly λ blocks in $\mathcal{B}$.*

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

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

Corollary 2. *Suppose $(T, \mathcal{B})$ is an t-(n, k, λ)-design. Then for any $1 \leq w \leq t$, $(T, \mathcal{B})$ is also a w-(n, k, λ_w)-design, for $\lambda_w = \lambda \cdot \binom{n-w}{t-w} / \binom{k-w}{t-w}$.*

Corollary 3. *Suppose $(T, \mathcal{B})$ is an t-(n, k, λ)-design. Then for any $W \subseteq T$ with size $|W| = w \leq t$, $(T - W, \{B - W : W \subseteq B \in \mathcal{B}\})$ is a $(t - w)$-$(n - w, k - w, \lambda)$-design.*

Furthermore, we remark that although non-trivial block designs exist universally by [32], only very few such designs are explicitly constructed.

3 Model

Our goal is to implement the decentralized custody scheme without relying on any trusted party. More specifically, we investigate the feasibility that n custodians (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. no more than $n/3$ nodes are corrupted simultaneously. This assumption 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.

The decentralized custody scheme is based on overlapping group assignments. That is, the custodians are assigned to overlapping groups where each group is fully controlled by its members and holds a fraction of the total asset under custody, including both collateral from custodians and funds from customers. In what follows we assume that the in-safekeeping assets are evenly distributed to custodian groups, since an uneven distribution naturally reduces to the case of even distribution with a graceful degradation on security (the tolerable adversary power) and/or capital efficiency.

Furthermore, we consider the security of a custody scheme in a game-theoretic sense: the adversary may corrupt multiple nodes, but will not launch an attack if the profit does not exceed the cost of launching such an attack. To achieve such game-theoretic security, every custodian in our scheme must provide a security deposit, which will be confiscated and used for compensation in case of misbehavior. Thus, if misbehavior can be detected in time (by periodical examination or blockchain transparency), no rational adversary would ever launch an attack as long as his collateral outweighs the revenue of a successful attack. Here, we emphasize that instead of resorting to another level of collateral custody service, the deposit from custodians can be maintained as a part of the total assets under custody, together with assets from external customers.

We assume that attack in the decentralized custody scheme can be detected immediately. If the decentralized custody service is for cryptoassets and deployed on a blockchain (as we first consider it in mind), then all instructions from customers and transfers of assets are transparent to everyone, and hence any unauthorized transaction will be caught immediately. Alternatively, the detection may be implemented with periodically examination which ensures that misbehavior is discovered before the adversary is able to exit or change the set of corrupted nodes. In other scenarios, detecting corrupted behavior may be a non-trivial problem, but for the sake of this study we will leave it out to avoid another layer of complication.

The incentive of agents participating in this collateralized custody scheme is also indispensable for a full-fledged decentralized custody service. A reasonable rate of commission fee and/or inflation tax would be sufficient to compensate the cost of agents proving such custody service. For cross-blockchain portable assets, e.g. BTC-tokens mapped onto Ethereum, extra per-transaction fee is also an option. Overall we believe that the mechanism design to incentivize custodians is essentially another topic, which is beyond the scope of this work and should be left for future study.

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, as n grows getting such an approval becomes expensive and even infeasible in practice, especially when honest participants may go off-line (as in the sleepy model [26]), 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 specific 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 overlapping subsets of S as custodian groups, i.e. each subset of S only controls a fraction of the total assets under custody. Here S is the set of all custodians.

Example 1 (Toy example). Consider the case when $n = 5$, 10 units of asset are under custody in total, and the total security deposit is $6 \times 5 = 30$ units. 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 up to two nodes.

In what follows, we will formalize the model of decentralized custody scheme with asset evenly distributed among custodian groups. To start with, we introduce a formal definition on the custody scheme we consider in this work.

Definition 2 (Custody scheme). *A custody scheme $(S, \mathcal{A}, \mu)$ consists of the following three parts:*

- *$S = \{1, 2, \cdots, n\}$ denotes the set of all custodians (or simply nodes);*
- *$\mathcal{A}$ denotes a family of m subsets of S, such that each element in $\mathcal{A}$ (i.e. a subset of S) represents a custodian group under the given custody scheme;*

– $\mu \in [1/2, 1)$ *denotes a universal authentication threshold for all custodian groups, i.e. the asset controlled by that group can be settled arbitrarily with approval of* strictly above μk *group members.*

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 collateral and every group in $\mathcal{A}$ is of the same size k. In particular, our discussion of the authentication threshold μ mainly focuses on $\mu = 1/2$ and $\mu = 2/3$, corresponding to authentication with simple majority, and greater than 2/3 majority respectively. [4] 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 a natural number, 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 *at least* 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 custody scheme $(S, \mathcal{A}, \mu)$, together with γ for the bound of adversary power, we use the function $f(\gamma; S, \mathcal{A}, \mu)$ 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 NP-hard to find the optimal attacking strategy in general, as discussed in Section 6). Formally,

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

Recall that as all assets under custody are equally distributed to all custodian groups, the value of every corrupted group is equal to the adversary. Therefore, $f(\gamma; S, \mathcal{A}, \mu)$ represents the maximal gain of the adversary.

From $f(\gamma; S, \mathcal{A}, \mu)$, we define the efficiency factor of a custody scheme, which captures the ability of securely holding exterior assets.

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

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

where m *is the total number of custodian groups induced by* $\mathcal{A}$*.*

The efficiency factor η indeed equals to the maximal ratio of capable exterior assets to pledged assets that the underlying custody scheme is able to handle. Specifically, suppose that u units of

[4] $\mu \geq 1/2$ is the necessary and sufficient condition for Byzantine agreement under synchrony, i.e. when all members are well-connected [22]. $\mu \geq 2/3$ is necessary and sufficient for Byzantine agreement under partial synchrony or asynchrony even with digital signatures [27]. We further remark that smaller μ implies less security but better liveness, e.g. when $\mu \to 0$ even a single corrupted member is able to block a custodian group. However, the discussion of liveness is beyond the scope of this work.

asset are deposited in total, and v units of exterior asset are safeguarded. According to (1), by launching an attack the adversary is able to seize the funds of $f(\gamma; S, \mathcal{A}, \mu)$ custodian groups, which amounts to $(u+v) \cdot f(\gamma; S, \mathcal{A}, \mu)/m$ units of asset, at the cost of losing security deposits worthy of value $\gamma \cdot v$ units. Recall that in our model, collateral and exterior assets are homogeneous and kept together by the custodian groups, the custody scheme is secure as long as

$$\frac{f(\gamma; S, \mathcal{A}, \mu)}{m}(u+v) \leq \gamma \cdot u,$$

or equivalently, $\eta \geq v/u$ according to Definition 3.

As an example for the definition, $\eta = 1$ implies that the system is secure when the total value of exterior assets is no more than total pledged security deposit.

Notice that when the efficiency factor $\eta < 0$ for some γ, the custody scheme against that γ-adversary is always insecure, regardless of the amount of collateral. To capture such property, we further define the safety of a custody scheme based on the Definition 3.

Definition 4 (Safety of custody scheme). *For a custody scheme $(S, \mathcal{A}, \mu)$ and γ defined as above, we say that the custody scheme is* γ-reliable *if the efficiency factor η of the scheme is non-negative against γ-adversary, i.e.*

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

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

Putting into our formal definition, the trivial solution has only one custodian group (i.e. $m = 1$, $k = n$), with $\eta = \infty$ 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 Table 1.

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 ($f(\gamma; S, \mathcal{A}, \mu)$)	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 $\eta < 0$.

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 a fixed n, every specific group assignment $\mathcal{A}$ and fixed constant μ (say, $\mu \in \{1/2, 2/3\}$), as the parameters m and k are already specified in $\mathcal{A}$, the maximal number of corrupted groups and the efficiency factor η are functions solely depending on the adversary power γ.[5]

[5] We remark that the number of custodians n is not always extractable from the group assignment scheme $\mathcal{A}$, as in some cases(especially when we consider random sampling in Section 5), some custodians may belong to no group. This is also why we include the whole custodian set S in (1).

Therefore, in the rest of this article we will focus on construction and analysis of custodian group assignments. We want the group assignment $\mathcal{A}$ to support a decentralized custody scheme with efficiency factor η as large as possible against an appropriate γ-adversary (e.g. $\gamma < 1/3$ or so). At the same time, $\mathcal{A}$ should have small m (e.g. ideally within a few hundreds) and small k (e.g. ideally bounded by one hundreds), while S should imply rather a large k (e.g. ideally about a thousand). so that the induced custody scheme provides sufficient decentralization at a tolerable cost on nowadays blockchains such as Ethereum.

Remark 1 (Non-overlapping assignments do not work). Apart from the most trivial settings (e.g. $n = k$ such that one group containing all custodians, or γ is too small to corrupt even a single custodian group), it is imperative to using overlapping group assignments for any meaningful custody scheme. Consider the custody scheme $(S, \mathcal{A}, \mu)$such that groups in $\mathcal{A}$ do not overlap. Such a scheme is rather vulnerable, as the adversary only needs to partially corrupt a custodian group in order to control all assets kept by that group. Even if there is no external assets, a single custodian group holds at least the collateral from all its group members in the non-overlapping setting, which already exceeds the adversary's collateral. In case the assets are not distributed evenly, there must exist at least one group with more assets than the minimal cost of corrupting it, which will be the adversary's target.

Remark 2 (No perfect solution). 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 Assignment

In this section, we propose three types of group assignment 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 (Symmetric design). 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:

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

The efficiency factor η turns out:

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

When $\mu \geq \gamma^6$, according to the tail bound of hypergeometric distribution (see Section 2), we have

$$\eta = \frac{\gamma}{\sum_{t=r}^{k} \binom{\gamma n}{t}\binom{n-\gamma n}{k-t} \Big/ \binom{n}{k}} - 1 \geq \gamma \cdot e^{2(\gamma-\mu)^2 k} - 1,$$

which establishes a good lower bound on the efficiency factor of the symmetric design under appropriate γ.

The reliability of $\mathcal{A}_{all}$ against adversary corrupting exactly γn nodes naturally implies security against adversary corrupting $\leq \gamma n$ nodes for reasonable γ, as proved in the following lemma. For typical choice of $\mu \geq 1/2$ 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 $\gamma \leq \min\left\{\frac{\mu k-1}{k-1} + \frac{1}{n}, 1 - \frac{k}{n}\right\}$, then it is secure against γ-adversary (i.e. γ'-reliable for every $\gamma' \leq \gamma$).*

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 when $\mu > \gamma$, 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. □

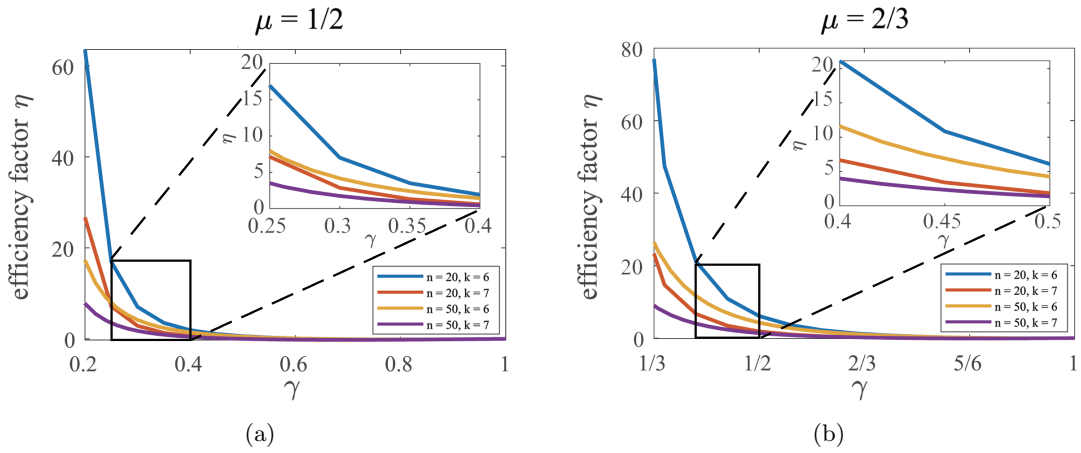


Fig. 1. The efficiency factor η versus the corrupted fraction γ for $\mathcal{A}_{all}$ as in Construction 1. In particular, $\eta < 0$ if the custody scheme is not secure for corresponding γ.

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 $\gamma = 2/5$. Here, we recall that in our context, all assets are evenly distributed to all custodian groups, or else the performance of the system will be worse than what the efficiency factor of the custody scheme implies.

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 that 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.

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 $k \sim 5$, in order to keep m reasonable. One solution to mitigate the above issues is by random sampling, as exhibited in Section 5.

4.2 Type 2: Block Design

Now 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 $t = k$ and $\lambda = 1$. In what follows, a "block" in the block design is also called a "group" in the group assignment scheme.

[6] We mention that in this work, when considering the reliability of a custody scheme, we tacitly approve that $\mu \geq \gamma$. For a better understanding, consider Example 1 with only one group consisting of all custodians. Under such group assignment, when $\gamma > \mu$, the scheme is surely γ-unreliable.

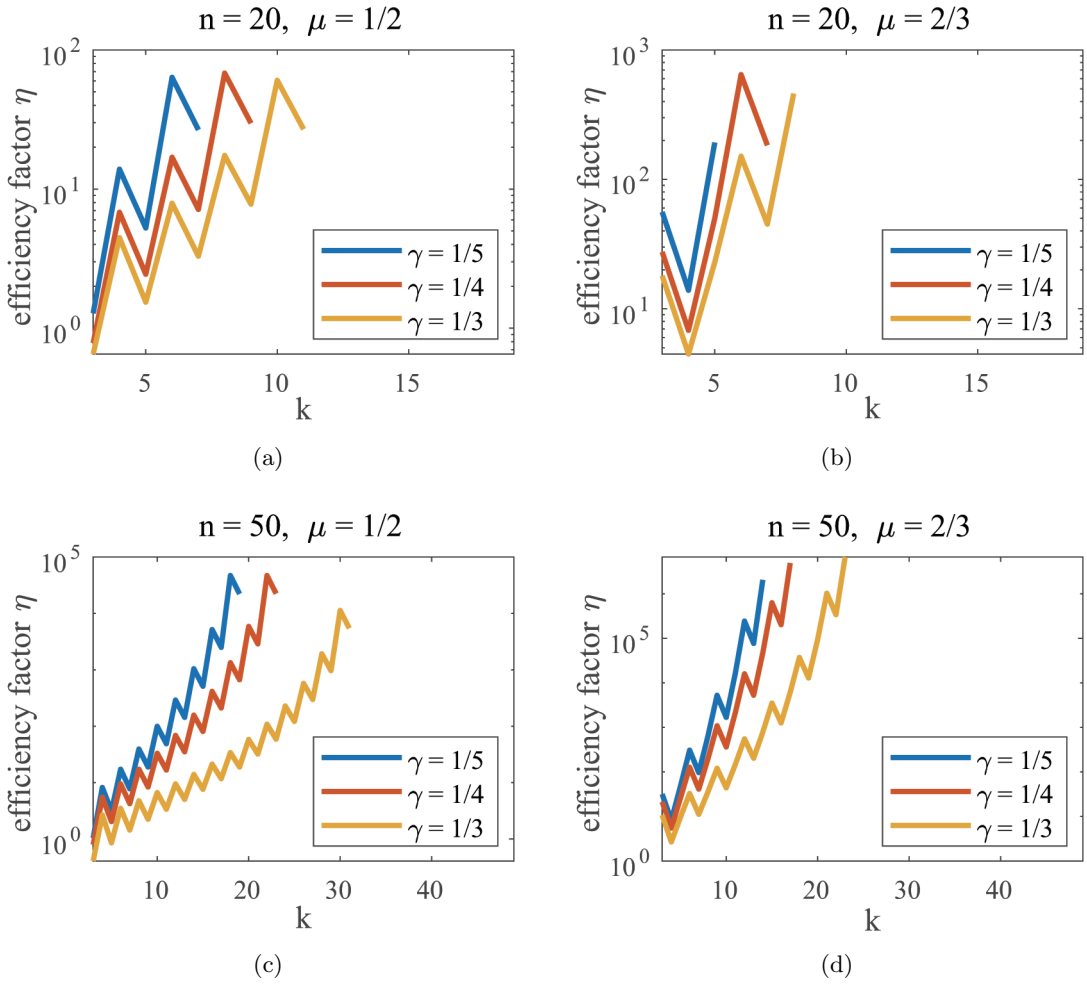


Fig. 2. The efficiency factor η versus the custodian group size k for $\mathcal{A}_{all}$ as in Construction 1. Blank points on the right side refer to $\eta = \infty$ when adversary cannot corrupt even a single custodian group.

The following theorem shows the effectiveness of block designs:

Theorem 2. *For every t-(n, k, λ)-design $(T, \mathcal{B})$, let the custody scheme be $(S, \mathcal{A}, \mu) = (T, \mathcal{B}, (t-1)/k)$ (which implies that $r = t$), and recalling that $s = \gamma n$ denotes the number of corrupted nodes, the induced custody scheme $(S, \mathcal{A}, \mu)$is γ-reliable with efficiency factor η lower bounded as follows:*

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

Proof. By corrupting s nodes, the adversary controls $f(\gamma; S, \mathcal{A}, \mu) \leq \lambda \cdot \binom{s}{r}$ custodian groups, and there are $m = \lambda \cdot \binom{n}{r} / \binom{k}{r}$ groups in total. Thus, the lower bound for η follows since $\eta = \frac{\gamma m - f(\gamma; S, \mathcal{A}, \mu)}{f(\gamma; S, \mathcal{A}, \mu)}$. □

Fig. 3 shows the lower bound of η obtained by Theorem 2 versus the adversary's power γ for different block designs with $\mu = (r-1)/k$.

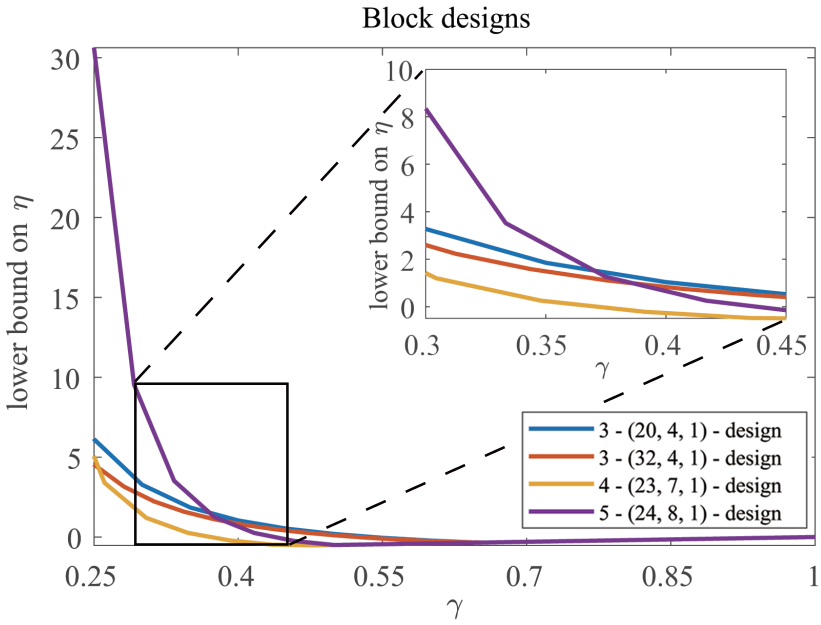


Fig. 3. The lower bound on η by Theorem 2, versus the corrupted fraction γ, when $\mu = (r-1)/k$. All four designs shown in this figure have explicit constructions [32]. $\eta < 0$ if the custody scheme is not secure for corresponding γ.

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

Proposition 1. *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 with μ is γ-reliable as long as $\gamma \leq \frac{1}{k} \cdot \mu^{\frac{1}{r-1}} + \frac{r-1}{n}$. 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 Theorem 2. Specifically, the system is reliable when

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

A key observation is that,

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

With the above observation we conclude the following inequality for $1\le t\le r-1$,

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

As a result,

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

The full proof of this proposition is deferred to Appendix A.3. □

Proposition 1 guarantees the security of the custody scheme for proper γ. Although Theorem 2 does not provide an ideal lower bound estimation for large γ (see Fig. 3), we still manage to achieve some satisfying results. For instance, using the custody scheme induced from the 5-$(24, 8, 1)$-design (see [10] for the construction) with $m = 759$ custodian groups and $\mu = 1/2$, the efficiency factor $\eta \ge 30.6250$ when $\gamma \le 1/4$. This significantly outperforms the symmetric design $\mathcal{A}_{all}$, where $m = 38,760$ and $\eta = 16.9444$ for $n = 20$, $k = 6$, $\mu = 1/2$ and $\gamma = 1/4$ (see Fig. 1(a)).

4.3 Type 3: Polynomial Design

The following construction of group assignments relies on polynomial-based combinatorial designs.

Construction 2 (Polynomial design). For a prime k and fixed positive integer $d < k$, suppose we have $|S| = n = k^2$ custodians. Let $T = \{(a, b) \mid a, b \in \mathbb{Z}/k\mathbb{Z}\}$ be the set of $n = k^2$ elements in $(\mathbb{Z}/k\mathbb{Z})^2$. Every custodian in S corresponds to a unique element in T via a one-to-one bijection (in what follows elements in T are used to represent correpsonding custodians in S for succinctness). The polynomial design $\mathcal{A}_{poly}$ is a family of $m = k^d$ subsets of S defined as follows:

$$\begin{aligned}\mathcal{A}_{poly} &:= \{A_p \mid p \text{ is a degree-}d \text{ monic polynomial over } \mathbb{Z}/k\mathbb{Z}\}\\ \text{where } \forall p,\ A_p &:= \{(0, p(0)), (1, p(1)), \cdots, (k-1, p(k-1))\}\end{aligned}$$

Then, for every specific authentication threshold μ, a custody scheme can be derived 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 by the Fundamental Theorem of Algebra, i.e.:

$$\forall A_p, A_q \in \mathcal{A}_{poly} \wedge A_p \ne A_q \implies |A_p \cap A_q| < d. \tag{3}$$

Hence, the efficiency factor η is lower bounded as below.

Theorem 3. *The efficiency factor η for $\mathcal{A}_{poly}$ as in Construction 2 is lower bounded as follows:*

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

Proof (sketch). We say a subset of S is *first-entry-unrepeated* if all nodes in the subset are with different first entry. Let $\mathcal{M}$ denote the set of corrupted nodes, then the number of size-d first-entry-unrepeated subsets of $\mathcal{M}$ is upper bounded by $\binom{k}{d}(s/k)^d$.

By (3), every size-d first-entry-unrepeated subset of $\mathcal{M}$ appears in at most one corrupted custodian group. On the other hand, every corrupted group contains at least r corrupted nodes with distinct first entry, and hence consumes $\geq \binom{r}{d}$ size-d first-entry-unrepeated subsets. Therefore, we have

$$f(\gamma; S, \mathcal{A}, \mu) \leq \frac{\binom{k}{d}}{\binom{r}{d}} \left(\frac{s}{k}\right)^d.$$

Recalling that $s = \gamma n = \gamma k^2$ and $m = k^d$, the proof finishes following the definition of η as in Definition 3. The complete proof of Theorem 3 is deferred to Appendix A.4. □

We further investigate the parameters of $\mathcal{A}_{poly}$:

- When $d \ll k$, according to Theorem 3, we have

$$\eta \geq \gamma^{1-d} \cdot \left(\frac{r}{k}\right)^d - 1 \geq \mu \cdot \left(\frac{\mu}{\gamma}\right)^{d-1} - 1,$$

  which induces that $\eta \geq 2^{d-1} \cdot \mu - 1$ for $\mu > 2\gamma$. Therefore, $\eta \geq 8.6$ holds for $d = 5$ and $\mu = 0.6$.
- When $r = k$ (i.e. $\mu \to 1$), the lower bound in Theorem 3 turns into the following form:

$$\eta \geq \gamma^{1-d} - 1.$$

  That is, for $\gamma = 1/3$ and $n = 121, k = 11, d = 3$, an assignment $\mathcal{A}_{poly}$ with $11^3 = 1,331$ groups has efficiency factor $\eta \geq 8$, where each custodian belongs to $11^2 = 121$ groups; for $n = 49, k = 7, d = 4$, an assignment $\mathcal{A}_{poly}$ with $7^4 = 2,401$ groups has efficiency factor $\eta \geq 26$, with each custodian appearing in $7^3 = 343$ groups.

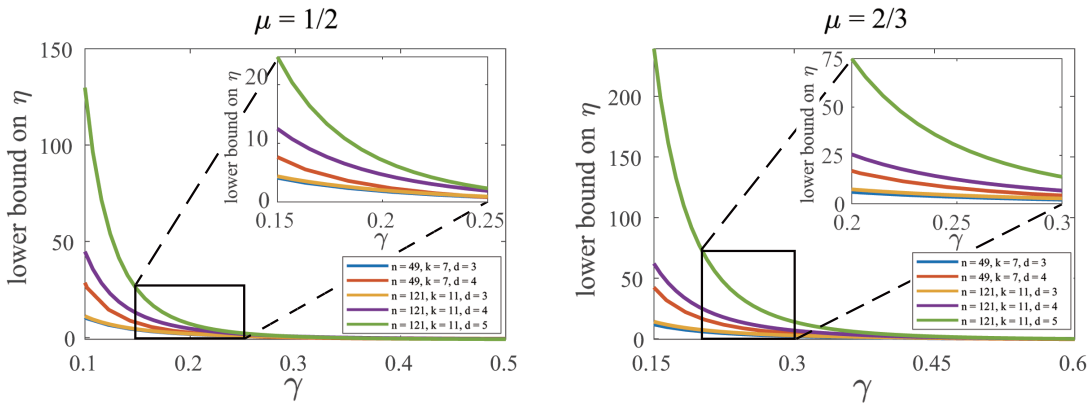
Furthermore, we remark that the efficiency factor $\eta \sim \gamma^{1-d}$ and every custodian belongs to k^{d-1} groups. Thus, the choice of d with good efficiency and low burden on every single custodian should be $2 \leq d \leq 4$. If γ is not too small (say, 0.1), then d should be 3 or 4.

Fig. 4 depicts the relation between the lower bound of η following Theorem 3 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 $\eta \geq 20.9094$ 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 as γ decreases since $\eta \sim \gamma^{1-d}$. For instance, the lower bound for η is improved to ≥ 109.9166 when γ is reduced from $3/11$ to $2/11$ in the above example.

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

Proposition 2. *Let $d = \nu k$, then the custody scheme derived from $\mathcal{A}_{poly}$ is secure against γ_{poly}-adversary for γ_{poly} defined as:*

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



(a) (b)

Fig. 4. The lower bound for the efficiency factor η (by Theorem 3) 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.

Proof (Sketch). By Theorem 3, the custody scheme remains reliable as long as

$$\gamma^{d-1} \leq \binom{r}{d} \Big/ \binom{k}{d} .$$

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

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

We prove the proposition by letting the right hand side of above formula no less than γ^{d-1}. The full proof with complete calculation is deferred to Appendix A.5. □

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:

$$\xi := -\left(\gamma \log \gamma + (1-\gamma) \log(1-\gamma)\right) . \tag{4}$$

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 sampling rate $\beta \in (0,1)$, 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 $\beta f(\gamma; S, \mathcal{A}, \mu) + \sqrt{n\xi\cdot\beta m}$ 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:

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

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

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

Therefore, with probability at least $1 - \frac{e}{2\pi}\exp(-cn\xi)$ there is $f(\gamma; \mathcal{A}', \mu) \leq \beta f(\gamma; S, \mathcal{A}, \mu) + \sqrt{n\xi\cdot\beta m}$, 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.6. □

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

Example 2 (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 sample $88,695$ groups uniformly at random from $\mathcal{A}_{all}$ to form $\mathcal{A}'$, then by Theorem 4 (with $c = 0$), 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 with probability $1 - \frac{e}{2\pi}$, $\eta' \geq 10$.

Example 3 (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 draw $138,767$ groups uniformly at random from $\mathcal{A}_{poly}$ to form $\mathcal{A}'$, then by Theorem 4 (with $c = 0$) 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 with probability $1 - \frac{e}{2\pi}$, $\eta' \geq 10$.

For $c = 1$, Theorem 4 transforms into an easy-to-digest version as in Corollary 4.

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.*

The effect of Corollary 4 is best illustrated in conjunction with the polynomial design construction in Section 4.3. Recall that with the polynomial-design assignment $\mathcal{A}_{poly}$, the efficiency factor η is roughly the order of γ^{-d} (especially when r and k are comparable), which prefers large d for better efficiency; on the other hand, the total number of custodian groups in the design is $k^d = n^{d/2}$, which would become infeasible to handle if d is too large. However, after fixing target γ and η, Corollary 4 achieves $\eta' \sim \sqrt{\eta} \sim (1/\gamma)^{d/2}$ with a polynomial-design assignment $\mathcal{A}_{poly}$ of $m' = O(n)$ groups, which allows to improve η' nearly for free, i.e. increasing d without blowing up the number of groups.

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 [28], 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 [17]. For the details on the reduction step, please refer to Appendix A.7. □

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.8. □

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 trade-off between liveness and security.

References

1. BTCRelay. http://btcrelay.org/
2. DeFi Pulse: Total Value Locked in DeFi. https://defipulse.com/
3. 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
4. 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)
5. 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)
6. BitcoinWiki: Atomic Swap. https://en.bitcoin.it/wiki/Atomic_swap
7. BitcoinWiki: Hashlock. https://en.bitcoin.it/wiki/Hashlock
8. 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)
9. Buterin, V.: Chain interoperability. Tech. rep., R3 (2016)
10. Carmichael, R.D.: Tactical configurations of rank two. American Journal of Mathematics **53**(1), 217–240 (1931)
11. Chvátal, V.: The tail of the hypergeometric distribution. Discrete Mathematics **25**(3), 285–287 (1979)
12. Conflux: Conflux Shuttleflow: A Cross-Chain Asset Protocol. https://medium.com/conflux-network/conflux-shuttleflow-a-cross-chain-asset-protocol-15ad6b2a9539
13. Cornuejols, G., Fisher, M.L., Nemhauser, G.L.: Exceptional paperlocation of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Management science **23**(8), 789–810 (1977)
14. Feige, U.: A threshold of ln n for approximating set cover. J. ACM **45**(4), 634–652 (1998)
15. Feige, U., Langberg, M.: Approximation algorithms for maximization problems arising in graph partitioning. J. Algorithms **41**(2), 174–211 (2001)
16. Gagol, A., Kula, J., Straszak, D., Swietek, M.: Threshold ECDSA for decentralized asset custody. IACR Cryptol. ePrint Arch. **2020**, 498 (2020)
17. Gandhi, R., Kortsarz, G.: On set expansion problems and the small set expansion conjecture. Discrete Applied Mathematics **194**, 93–101 (2015)

18. Harz, D., Gudgeon, L., Gervais, A., Knottenbelt, W.J.: Balance: Dynamic adjustment of cryptocurrency deposits. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019. pp. 1485–1502. ACM (2019)
19. HBTC.Finance: Hbtc whitepaper: Bridge between centralized and defi markets. https://www.hbtc.finance/static/pdf/whitepaper-en.pdf
20. Hochbaum, D.S.: Approximation Algorithms for NP-Hard Problems. PWS Publishing Co., USA (1996)
21. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (1963)
22. Katz, J., Koo, C.: On expected constant-round protocols for Byzantine agreement. J. Comput. Syst. Sci. **75**(2), 91–112 (2009)
23. Keep.Network: tBTC: A decentralized redeemable BTC-backed ERC-20 token. https://docs.keep.network/tbtc/index.pdf
24. 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)
25. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008)
26. 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)
27. Pease, M.C., Shostak, R.E., Lamport, L.: Reaching agreement in the presence of faults. J. ACM **27**(2), 228–234 (1980)
28. 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)
29. renProject: renProject - wiki. https://github.com/renproject/ren/wiki
30. Robbins, H.: A remark on Stirling's formula. The American mathematical monthly **62**(1), 26–29 (1955)
31. Schnorr, C.P.: Efficient signature generation by smart cards. Journal of cryptology **4**(3), 161–174 (1991)
32. Stinson, D.: Combinatorial designs: constructions and analysis. Springer Science & Business Media (2007)
33. Tokenlon: Tokenlon dex now on web. https://tokenlon.zendesk.com/hc/en-us/articles/360037584171-Tokenlon-DEX-now-on-Web
34. wBTC.Network: Wrapped tokens: A multi-institutional framework for tokenizing any asset. https://wbtc.network/assets/wrapped-tokens-whitepaper.pdf
35. Wikipedia: Mt. Gox. https://en.wikipedia.org/wiki/Mt._Gox
36. Wood, G.: Polkadot: Vision for a heterogeneous multi-chain framework. White Paper (2016)
37. Wood, G., et al.: Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper **151**(2014), 1–32 (2014)
38. Zamyatin, A., Al-Bassam, M., Zindros, D., Kokoris-Kogias, E., Moreno-Sanchez, P., Kiayias, A., Knottenbelt, W.J.: Sok: Communication across distributed ledgers. IACR Cryptol. ePrint Arch. **2019**, 1128 (2019)

Appendix A Proof of lemmas, theorems and propositions

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 $t > \mu k$.

Therefore,

$$\begin{aligned} \frac{(s-t)(n-s+1)}{(n-s-k+t+1)(s-1)} \le 1 &\iff s \cdot \frac{\binom{n}{k}}{\binom{s}{t}\binom{n-s}{k-t}} \le (s-1) \cdot \frac{\binom{n}{k}}{\binom{s-1}{t}\binom{n-s+1}{k-t}} \\ &\iff \frac{s}{n} \frac{\binom{n}{k}}{\sum_{t=r}^{k} \binom{s}{t}\binom{n-s}{k-t}} \le \frac{s-1}{n} \frac{\binom{n}{k}}{\sum_{t=r}^{k} \binom{s-1}{t}\binom{n-s+1}{k-t}} \\ &\iff \eta_s \le \eta_{s-1}, \end{aligned}$$

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

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

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 $\frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{k} \leq \gamma \leq \frac{1}{2} - \frac{1}{\sqrt{k}} - \frac{1}{2k}$ and $k \geq 10$,

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

The second inequality establishes as $(1-1/x)^x \leq 1/e$ 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 Proposition 1

Proof. By Theorem 2, 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. We say a subset of S is *first-entry-unrepeated* if all nodes in the subset are with different first entry. Now that the adversary corrupts s nodes in total. Let s_i be the number of corrupted nodes with first entry i. We have $s = s_0 + s_1 + \cdots s_{k-1}$. Therefore, the number of size-d first-entry-unrepeated subsets that is totally corrupted is

$$\sum_{0 \le i_1 < i_2 \cdots < i_d \le k-1} s_{i_1} s_{i_2} \cdots s_{i_d}.$$

To give an upper bound on the above formula, we extend the problem to the case where $s_0, s_1, \cdots, s_{k-1}$ are multiples of $1/k$. (The original problem is when $s_0, s_1, \cdots, s_{k-1}$ are all integers.) For any two indices u, v, suppose $s_v - s_u \ge 2/k$, note that

$$\begin{aligned}
&\sum_{0 \le i_1 < i_2 \cdots < i_d \le k-1} s_{i_1} s_{i_2} \cdots s_{i_d} \\
&= s_u s_v \sum_{\substack{0 \le i_1 < i_2 \cdots < i_{d-2} \le k-1 \\ \{i_1, i_2, \cdots, i_{d-2}\} \cap \{u,v\} = \emptyset}} s_{i_1} s_{i_2} \cdots s_{i_{d-2}} + (s_u + s_v) \sum_{\substack{0 \le i_1 < i_2 \cdots < i_{d-2} \le k-1 \\ \{i_1, i_2, \cdots, i_{d-1}\} \cap \{u,v\} = \emptyset}} s_{i_1} s_{i_2} \cdots s_{i_{d-1}} \\
&+ \sum_{\substack{0 \le i_1 < i_2 \cdots < i_d \le k-1 \\ \{i_1, i_2, \cdots, i_d\} \cap \{u,v\} = \emptyset}} s_{i_1} s_{i_2} \cdots s_{i_d}.
\end{aligned}$$

Hence, the sum is strictly increase by substituting s_u with $s'_u = s_u + 1/k$ and s_v with $s'_v = s_v - 1/k$. Therefore, the sum reaches a maximum with $s_0 = s_1 = \cdots = s_{k-1} = s/k$ in the generalized case, which implies that with $s_0, s_1, \cdots, s_{k-1}$ all integers, we have

$$\sum_{0 \le i_1 < i_2 \cdots < i_d \le k-1} s_{i_1} s_{i_2} \cdots s_{i_d} \le \binom{k}{d} \left(\frac{s}{k}\right)^d,$$

or that the number of size-d first-entry-unrepeated subsets that is totally controlled by the adversary is upper bounded by $\binom{k}{d}(s/k)^d$. Let $\mathcal{M}$ denote the set of corrupted nodes. By (3), every size-d first-entry-unrepeated subset of $\mathcal{M}$ appears in at most one corrupted group. On the other hand, every corrupted group contains at least r corrupted nodes all with different first entry, and hence $\ge \binom{r}{d}$ size-d first-entry-unrepeated subsets. Consequently, the number of corrupted groups $f(\gamma; S, \mathcal{A}, \mu)$ is upper bounded by $\frac{\binom{k}{d} s^d}{\binom{r}{d} k^d}$, and the efficiency factor η is lower bounded by

$$\eta \ge \frac{\gamma \binom{r}{d} k^{2d}}{\binom{k}{d} s^d} - 1 = \gamma^{1-d} \cdot \binom{r}{d} \Big/ \binom{k}{d} - 1,$$

as $s = \gamma n = \gamma k^2$. □

A.5 Proof of Proposition 2

Proof. As per Theorem 3, it is sufficient if we have

$$\gamma^{d-1} \le \binom{r}{d} \Big/ \binom{k}{d}.$$

.

As we have

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

the theorem is proved by requiring the last formula no less than γ^{d-1}. Here, we use Stirling's formula (see Section 2) to estimate the factorial term in the second line. Meanwhile, in the last-but-one line, we apply that $(1+1/x)^x < e$ for all $x > 0$. □

A.6 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, $g(\mathcal{M}) \leq f(\gamma; S, \mathcal{A}, \mu)$ 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

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

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

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

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,

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

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 $f(\gamma;\mathcal{A}',\mu) \leq \beta f(\gamma;S,\mathcal{A},\mu)+\delta\cdot\beta m$. Under such case, the efficiency factor η' of $\mathcal{A}'$ with malicious rate γ is lower bounded by

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

The third line is due to the definition of η:

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

□

A.7 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.
We should mention that the MAX s-VC problem in k-uniform hypergraphs ($k \geq 2$) is indeed a particular 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 [13, 20]. 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 [14]. Nevertheless, concerning MAX s-VC in common graphs, there is a $(0.75 + \delta)$-approximation algorithm for some constant $\delta > 0$ [15]. Meanwhile, the MAX s-VC problem in k-uniform hypergraphs with $k \geq 3$ can also be approximated strictly within the ratio 0.75 [4].
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 [17], 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.8 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 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 4. *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 4, note that $Y_0 = \kappa m$, then the theorem is proved. □

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 $\kappa := \Pr[\mathcal{H}(n, s, k) \geq r]$.

$T = \emptyset, N = \emptyset$
for $i = 1$ to n **do**
  Compute the expected amount of compromised groups X_i conditioning on nodes in $T \cup \{i\}$ are malicious and nodes in N are honest
  if $X_i \geq \kappa m$ **then**
    $T = T \cup \{i\}$
  else
    $N = N \cup \{i\}$
  end if
  if $|T| \geq s$ **then**
    return T
  end if
  if $|N| \geq n - s$ **then**
    return $S - N$
  end if
end for