Migrated from research forum. Original author: sam-ng

Hey everyone,

Following the pro tip from @kydo, let's explore the principal-agent problem (PAP) and delve into this topic here for more discussions.

The PAP arises from conflicting interests and priorities when one party (the "agent") acts on behalf of another (the "principal"). A common query in this scenario is: why does the principal need to incentivize the agent? Why not act independently? This situation often occurs because capital holders (in this case restakers) seek liquidity and prefer to delegate tasks. The challenge is achieving a balance between capital efficiency/separation of capital and labor while managing the difficulties of the principal-agent problem.

Let's consider several mechanisms to mitigate this issue:

1) Reputation-Based Approach:

This scenario is akin to a repeated game. Validators, as part of their long-term business model, gain from maintaining a good reputation. A damaged reputation can introduce the end of their business. Assuming rational behavior, validators will act honestly if the value of their reputation outweighs the potential gains from corrupt practices. Note that it becomes even more complex when considering that an attack would require collusion among multiple operators, thereby increasing the cost significantly due to the cumulative reputational value at stake. Staking services with different business sectors could also be adversely affected e.g Coinbase.

2) Legal and Social Frameworks:

The observability of actions, such as double signing a block, is enhanced when operators are subject to regulations that make such misbehavior illegal. The transparency provided, along with slashing conditions, makes any misbehavior evident and reduces the likelihood of intentional/detectable cheating. While cheating might still occur, it's more probable to happen in non-observable ways to avoid liability (e.g censorship), the expected adherence to protocol rules in regulated environments establishes a foundation of trust. Even though the relationship between a delegator (staker) and a delegate (operator) might not be legally explicit, compliance with these rules forms an implicit agreement, enhancing trust based on the operators' jurisdiction.

3) Technological Safeguards: Anti-Slashing Mechanisms

This approach, demonstrated by the efforts of e.gCubist, involves the deployment of technological frameworks to mitigate the principal-agent problem. The core of this strategy is a sophisticated coding module designed to simulate and enforce slashing conditions. The purpose of this module is to ensure compliance by operators when issuing signatures, thereby preventing any violation of the protocol's rules.

A key aspect of this mechanism is its reliance on Trusted Execution Environments (TEE), such as Intel's SGX. TEEs provide an isolated execution space that aims to protect the code from external manipulation. This means that even if an agent intends to act dishonestly or manipulate the system, the operation within the TEE presents barriers to such actions.

4) Economic Approach: Incentivizing Compliance through Collateral

The economic approach, as used by e.g Rocketpool, offers a solution by aligning the economic interests of both the principal and the agent. In this model, the agent or operator is required to provide an amount of collateral to participate in the service operation. This collateral is a form of financial guarantee/bond that is at risk if the operator fails to comply with the agreed-upon rules or engages in any form of misbehavior.

We can think of it as mutual risk-sharing. Both the principal (staker) and the agent (operator) stand to lose financially if there is any failure in the AVS operation. This shared risk creates an incentive for the agent to act in the best interest of the principal. It's a form of economic alignment where the agent's potential for profit is directly tied to their performance and adherence to the rules.

Addressing the principal-agent problem requires a multifaceted approach. From leveraging the long-term value of reputation and enforcing legal/social frameworks, to utilizing tech safeguards and implementing

economic incentives, each method offers unique benefits and challenges. As the space continues to evolve, these strategies need to be explored/refined to enhance trust and efficiency. Eager to hear your thoughts/insights on these approaches and any other ideas that could further advance our understanding!

How do you think these mechanisms interplay with each other?

In your opinion, what are the potential downsides/risks associated with these approaches?

What are your theories on why there have been so few quorum attacks (if any) on Proof of Stake networks so far?