Entre a Confiança e o Sigilo: Reputação Descentralizada com Garantia de Privacidade por Meio de Identidades Digitais Auto-Soberanas

Antonio Mateus de Sousa[1], **Allan Edgard S. Freitas**[2], **Leobino N. Sampaio**[1]

[1] Programa de Pós-Graduação em Ciência da Computação (PGCOMP)
Instituto de Computação – Universidade Federal da Bahia (UFBA)
Salvador – BA – Brasil

[2]Instituto Federal da Bahia (IFBA)

antonio.mateus@ufba.br, allan@ifba.edu.br, leobino@ufba.br

Abstract. *While authentication and authorization ensure identity and access control, they do not guarantee essential attributes such as service quality and reliability over time. In this context, reputation serves as a crucial complementary mechanism, reflecting factors like availability, data integrity, and compliance with SLAs. However, centralized approaches reduce system resilience and compromise user sovereignty. The proposed architecture adopts a decentralized model based on Decentralized Digital Identities (DDIs) and blockchain technologies, enabling verifiable and privacy-preserving proof of reputation. The PoC demonstrated that the system effectively penalizes malicious behavior, enhancing the security and robustness of the environment.*

Resumo. *Embora autenticação e autorização garantam identidade e controle de acesso, elas não asseguram atributos como qualidade e confiabilidade dos serviços ao longo do tempo. Nesse cenário, a reputação atua como mecanismo complementar essencial, refletindo aspectos como disponibilidade, integridade e conformidade com SLAs. No entanto, abordagens centralizadas reduzem a resiliência e comprometem a soberania dos usuários. A arquitetura proposta neste trabalho adota uma abordagem descentralizada, baseada em Identidades Digitais Descentralizadas (IDD) e tecnologias de blockchain, permitindo provar reputação de forma verificável e privada. A PoC demonstrou que o sistema penaliza eficazmente comportamentos maliciosos, aumentando a segurança e a robustez do ambiente.*

1. Introdução

Autenticação e autorização são mecanismos utilizados para assegurar a identidade e o controle de acesso aos sistemas; contudo, não são suficientes na garantia sobre a qualidade, confiabilidade ou consistência dos serviços providos por estes ao longo do tempo. A reputação, por outro lado, pode refletir aspectos como disponibilidade, tempo de resposta, integridade dos dados entregues, conformidade com acordos de nível de serviço (SLAs) ou mesmo resiliência diante de falhas e ataques [Liu et al. 2022, Nalini et al. 2023]. Seu uso pode ser aplicado tanto à avaliação de serviços prestados quanto à análise de fornecedores e consumidores, refletindo uma dinâmica intrinsecamente ligada à teoria dos jogos [Zhou et al. 2021, Arshad et al. 2022].

Como uma informação pública na Internet, a reputação pode ser alvo de ataques ou ser por si um ponto de falha do sistema. A centralidade das informações pessoais em grandes plataformas, os chamados silos de identidade, como Meta, Microsoft e Google, pode apresentar sérios desafios à resiliência de sistemas computacionais [Mazzocca et al. 2025, Ernstberger et al. 2023]. Ao contrário da abordagem centralizada, a literatura apresenta iniciativas de reputação descentralizada, em que a confiança é construída de forma distribuída [Arshad et al. 2022]. Contudo, persistem desafios relacionados à privacidade, como ataques de *impersonation* e envenenamento de reputação.

Diante deste contexto, o presente artigo apresenta **ShadowRep**, uma arquitetura para sistemas de reputação descentralizados, baseada em Identidade Digital Descentralizada (IDD) [Hoang et al. 2024], que permite o controle de credenciais por meio de tecnologias como blockchain. **ShadowRep** combina esses mecanismos com técnicas de *Privacy-Preserving Computation* [Kerschbaum 2012], como provas de conhecimento zero (ZKPs), oferecendo dois modelos: reputação anônima, onde agentes calculam reputação local sem exposição direta, e reputação às cegas, onde nem mesmo o agente conhece sua própria reputação, aumentando a robustez contra ataques Sybil.

2. Trabalhos Relacionados

Diversos trabalhos abordam reputação e privacidade em contextos descentralizados. Em [Feraudo et al. 2024], é proposto um modelo para redes veiculares que integra identidade descentralizada à segurança, embora sem detalhar a implementação dos identificadores descentralizados (DIDs), focando na detecção de informações falsas. Já [Solomon et al. 2023] apresenta o smartFHE, uma estrutura pioneira que aplica criptografia homomórfica completa (FHE) a contratos inteligentes, permitindo que usuários leves realizem computações privadas em blockchains, embora sem tratar diretamente da reputação, valida o uso de FHE nesse contexto. Por fim, [Liu et al. 2019] propõe um sistema de reputação anônima para redes industriais IoT (IIoT), utilizando blockchain de camada 2 para garantir anonimato verificável nos *feedbacks*, protegendo os consumidores contra retaliações, mas permitindo a recuperação da identidade em casos de má conduta.

A proposta neste trabalho apresenta sistemas de reputação baseados em identidades descentralizadas que resguardam a privacidade da reputação. Ou seja, um agente pode ter sua reputação utilizada por serviços sem necessariamente revelar o valor real. Para tal, combinamos conceitos criptográficos como provas de conhecimento zero (do inglês, *Zero Knowledger Proofs* - ZKPs) e criptografia completamente homomórfica (do inglês, *Fully Homomorphic Encryption*- FHE), para privacidade e compartilhamento seguro, respectivamente.

3. Arquitetura ShadowRep: Reputação descentralizada com garantia de Privacidade

Esta seção apresenta **ShadowRep**, arquitetura para sistemas de reputação puramente ou semi-descentralizados. A **ShadowRep** adota métodos e ferramentas criptográficas robustas, garantindo segurança e confiabilidade nas interações, com a compreensão de que sistemas de reputação atuais apresentam dados sensíveis que podem ter a privacidade explorada. Para isso, a arquitetura se baseia no uso de Identidade Digital Descentralizada (IDD), adotando um modelo no qual indivíduos controlam suas credenciais digitais,

geralmente gerenciadas por carteiras e suportadas por tecnologias como *blockchain* e assegurando a privacidade e o controle das informações pessoais por parte de seus legítimos titulares [Allen 2018].

Ao atrelar a reputação, ou múltiplas reputações, aos identificadores descentralizados (do inglês, *Decentralized Identifiers* — DIDs), a **ShadowRep** garante unicidade e autenticidade, uma vez que o DID é gerado a partir de um par de chaves criptográficas assimétricas, sendo a chave pública registrada em uma infraestrutura pública, como uma *blockchain*. A estrutura dos DIDs prevê que sua resolução seja possível por meio de métodos padronizados, permitindo acesso ao documento DID contendo a chave pública e outras metainformações essenciais [Sporny et al. 2022].

A arquitetura proposta apresenta duas abordagens: (i) um sistema de reputação anônima verificável, via *Zero Knowledge Proofs* e criptografia homomórfica; e (ii) um sistema de "reputação às cegas", no qual ambos, proprietário e demais nós, desconhecem sua própria reputação ao mesmo tempo em que ainda é possível utilizá-la. A partir de tais contribuições, a **ShadowRep** endereça a falta de privacidade da reputação associada quando verificável publicamente através de DIDs, uma vez que reputações públicas podem servir de motivação para ataques direcionados, i.e., em que agentes maliciosos buscam difamar identidades específicas ou mesmo subtrair suas credenciais [Allen 2018]. Outra questão a ser destacada é que um agente malicioso pode construir confiança suficiente na rede para atacá-la em seguida.

3.1. Arquitetura proposta

A Figura 1 apresenta a arquitetura de cinco camadas proposta. Como é possível observar, a arquitetura está organizada em cinco camadas com funções relacionadas à infraestrutura pública de confiança, atestações e credenciais, provas e verificação, gestão da reputação e, por fim, aplicações. Cada uma das referidas camadas engloba um conjunto diversificado de paradigmas, conceitos e tecnologias, que juntos provêem o serviço de reputação proposto.

A arquitetura proposta organiza-se em cinco camadas funcionais para reputação segura e privada: A Camada de Infraestrutura (1) implementa um oráculo baseado em *blockchain* (Ethereum/Hyperledger) que registra *hashes* de provas criptográficas de forma imutável. A Camada de Credenciais (2) emprega *Verifiable Credentials* (VCs) vinculadas a DIDs, habilitando portabilidade, revogação e métricas multivariadas como tempo de serviço e especializações.

Para preservação de privacidade, a Camada de Provas Criptográficas (3) implementa ZKPs através de circuitos aritméticos, compatíveis com verificadores *on-chain* e *off-chain*. A Camada de Gestão (4) provê APIs programáveis para operações reputacionais, enquanto a Camada de Aplicação (5) oferece interfaces de integração com dApps. O fluxo operacional segue a sequência: registro de identidades $\rightarrow$ emissão de VCs $\rightarrow$ geração de ZKPs $\rightarrow$ aplicação de políticas $\rightarrow$ integração via APIs, formando um ciclo completo de gestão reputacional descentralizada.

3.2. Reputação anônima verificável baseado em computação confidencial

A arquitetura **ShadowRep** suporta um sistema de reputação anônima e descentralizada, onde agentes assumem papéis transitórios (desafiante/desafiado) em uma rede dinâmica e

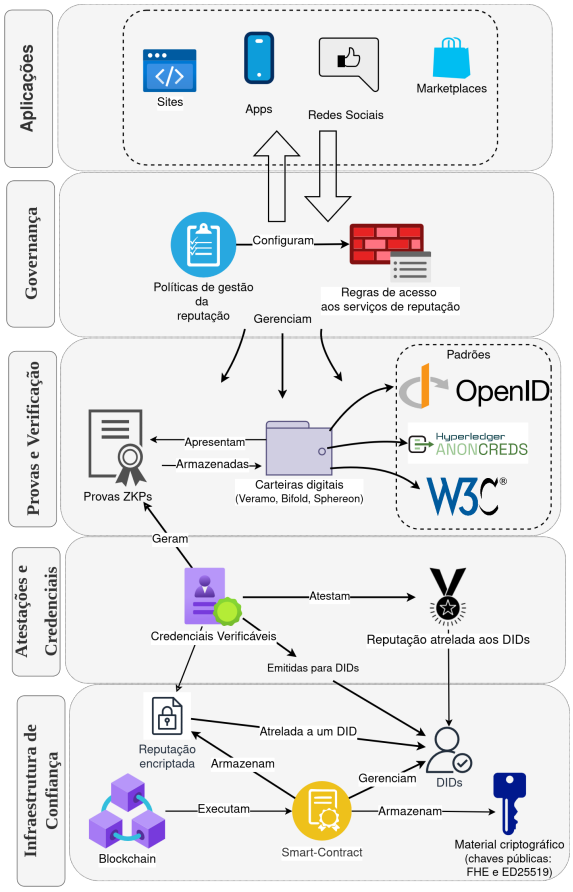

Figura 1. Arquitetura ShadowRep, projetada para provimento de serviço de reputação descentralizada com garantia de privacidade.

auto-organizável. A reputação local é baseada em provas de execução (PoE), mesmo com entrada e saída voluntária de nós e presença de falhas arbitrárias. Para garantir privacidade, a **ShadowRep** utiliza técnicas de *Privacy-Preserving Computation*, como criptografia homomórfica, que permite cálculos sobre dados cifrados, e provas de conhecimento zero (ZKPs), que viabilizam verificações sem revelar informações sensíveis. Isso assegura a confidencialidade dos dados reputacionais mesmo em decisões automatizadas. Um exemplo prático ocorre em contextos bancários: um cliente pode cifrar sua reputação com uma chave pública e permitir que a instituição opere sobre os dados cifrados. O resultado, também cifrado, é então descriptografado pelo cliente, mantendo a reputação privada durante todo o processo.

De forma alternativa, o banco também pode adotar uma postura proativa e requisitar a reputação do usuário para conceder acesso a determinados serviços. Contudo, o cliente não precisa revelar o valor exato de sua reputação. Neste caso, o banco solicita uma prova de que a reputação é superior ou inferior a um limiar previamente definido pela instituição. Para preservar a privacidade do usuário, utilizam-se provas de conhecimento zero.

3.3. Reputação anônima verificável baseado em computação confidencial

A arquitetura **ShadowRep** implementa um sistema de reputação anônima e descentralizada, onde agentes assumem papéis transitórios (desafiante/desafiado) em uma rede di-

nâmica. A reputação local é calculada mediante provas de execução (PoE), utilizando técnicas de *Privacy-Preserving Computation* como criptografia homomórfica e provas de conhecimento zero (ZKPs). Isso permite operações sobre dados cifrados e verificações sem exposição de informações sensíveis. Por exemplo, em transações bancárias, um cliente pode cifrar sua reputação para análise sem revelar seu valor real.

3.4. Reputação às cegas

A **ShadowRep** emprega reputação às cegas para prevenir manipulações, onde agentes não acessam diretamente sua própria reputação. O sistema utiliza controle de acesso em ambientes permissionados, criptografia avançada (FHE e MPC) e três mecanismos distribuídos principais: (1) derivação de subchaves FHE, (2) consenso para alterações de estado, e (3) auditoria criptográfica.

As operações requerem colaboração entre nós, garantindo resistência a ataques *Sybil*. Inicialmente adotando um esquema n-to-n, o sistema inclui uma camada de monitoramento para gerenciar nós operacionais, ajustando dinamicamente o conjunto de participantes ativos em um ambiente síncrono com detecção de falhas.

4. Prova de Conceito

Nesta seção, apresentamos uma prova de conceito (*Proof-of-Concept* - PoC) do sistema de reputação anônima da **ShadowRep**, implementada via emulação com contêineres Docker simulando dispositivos IoT heterogêneos. O ambiente foi configurado em uma máquina com 32 GB de RAM, processador Intel Core i7-1355U (13ª geração) e Ubuntu 22.04 LTS. A gestão de identidades (IDD) utilizou agentes Aries (ACA-Py), seguindo padrões W3C DID.

O sistema implementa um *daemon* de reputação que se integra ao ACA-Py, utilizando DIDComm v2 para comunicações criptografadas. Essa arquitetura estabelece interação contínua com um *smart contract* de orquestração, responsável por duas funções essenciais: (i) gestão dinâmica da reputação e (ii) verificação criptográfica das provas de execução (PoE-FHE). O modelo adotado prevê que validações bem-sucedidas resultem no incremento da reputação global, criando assim um mecanismo de incentivos que simultaneamente reforça a segurança computacional e promove a expansão orgânica da rede.

A geração das PoE-FHE emprega uma pilha tecnológica robusta envolvendo a OpenFHE-Python para operações totalmente homomórficas (FHE) e o Circom 2.0 em conjunto com o SnarkJS para construção de circuitos aritméticos ZKP otimizados. O protocolo PoE-FHE opera com base em três parâmetros fundamentais enviados pelo desafiante: o valor a, que consiste em um número aleatório cifrado com FHE; o nonce b, um identificador único determinado por um oráculo; e o hash h, obtido por meio da função Poseidon aplicada ao vetor $[c', b]$, sendo c' o resultado computado pelo desafiante. O cerne da prova reside na demonstração de conhecimento de um valor c'' tal que $H(c'' \parallel b) \equiv h$, em que $H(\cdot)$ representa a função hash Poseidon, $\parallel$ indica a concatenação dos dados, e a verificação dessa equivalência deve ser possível sem revelar o valor de c''. Este esquema garante que o desafiante realmente computou c' de forma válida, enquanto preserva a confidencialidade dos dados por meio da combinação de FHE e provas de conhecimento zero.

4.1. Modelo de reputação

O modelo de reputação da **ShadowRep**, implementado na PoC, baseia-se no desempenho dos agentes em desafios de *Proof of Execution using FHE* (PoE-FHE). Ele distingue entre reputação **global**, que reflete o desempenho dos desafiantes, e **local**, adaptável conforme o contexto.

A reputação global é atualizada com base na proporção de desafiados (α) que respondem corretamente em relação ao total de conexões (β), conforme a Equação 1. Se $\alpha \geq \frac{2}{3} \cdot \beta$, há **recompensa** com incremento δ; caso contrário, aplica-se **punição** com fator $\gamma > \delta$.

$$\Gamma_{\text{new}} = \begin{cases} \left(\frac{\alpha}{\beta}\right) \cdot \delta + \Gamma_{\text{old}}, & \text{se } \alpha \geq \frac{2}{3} \cdot \beta \\ \Gamma_{old} - (\beta - \alpha) \cdot \gamma, & \text{se } \alpha < \frac{2}{3} \cdot \beta \end{cases} \tag{1}$$

O PoE-FHE visa prevenir fraudes, como plágio de resultados. Um oráculo (*Smart Contract*) seleciona aleatoriamente um desafiante e um *nonce* b, exigindo uma entrada cifrada a que produza uma saída válida c'. A integridade é garantida por provas de conhecimento zero (ZKPs); falhas resultam em penalizações reputacionais exponenciais. Por fim, nós aleatórios emitem uma nova credencial de reputação calculada via a Equação 1.

4.2. Experimentos

A PoC foi avaliada em um cenário de IoT industrial com computação na borda, onde dispositivos executam tarefas críticas (e.g., tomada de decisão ou *offloading*). Neste contexto, a reputação atua como camada de segurança contra três comportamentos maliciosos: (i) nós desonestos (e.g., ataques DoS ou tráfego falso); (ii) personalidades dinâmicas (alternância imprevisível entre comportamentos); e (iii) ataques *churn* (troca de identidade para apagar histórico reputacional).

Os experimentos foram realizados em uma rodada de 40 minutos de duração, tendo um intervalo aleatório para cada tarefa (PoE-FHE) de 10 a 20 segundos. A partir dos resultados advindos da emulação, foi avaliada a evolução da reputação dos nós escolhidos como desafiantes ao longo do tempo. Na Figura 2 é mostrada a comparação entre os nós com comportamento regular (nós 1 e 2) e os nós com comportamento anômalo (nós 3 e 4). O Nó 3 foi configurado para agir como um nó oscilante, uma alusão a falhas de rede ou falhas do dispositivo em si que levam o dispositivo a não responder adequadamente aos desafios. Assim, o Nó 3 tem uma evolução de quedas contínuas na reputação, mantendo-se sempre abaixo de 0,10%.

Já o Nó 4 atua como um agente malicioso que, intencionalmente, constrói gradualmente sua reputação até atingir um nível considerado confiável (0,70%). Ao alcançar esse patamar, o agente inicia seu ataque, o que pode ser observado no intervalo entre os tempos 5–10. A partir desse ponto, a reputação do Nó 4 sofre uma queda acentuada e não volta a atingir os 70% ao longo do restante da emulação. O Nó 1, por sua vez, apresentou uma leve redução de reputação ao atuar como desafiante, possivelmente devido à sua lentidão na consulta ao oráculo. Embora esse comportamento afete temporariamente sua reputação, diferentemente dos nós com falhas reais, o Nó 1 tende a recuperar sua credibilidade com o tempo. É importante destacar que, do ponto de vista da rede, tanto os Nós

Figura 2. Reputação ao longo do tempo para nós com comportamento esperado (1 e 2) e para o nó oscilante (3), que simula falhas intermitentes de resposta.

3 quanto 4 são interpretados como agentes maliciosos, uma vez que não há informações suficientes para distinguir a causa exata de seus comportamentos anômalos.

4.3. Discussão dos resultados

Com base nos resultados apresentados, é possível afirmar que a **ShadowRep** se mostra eficaz na mitigação de comportamentos maliciosos por parte dos agentes na rede. Agentes cuja reputação atinge um limiar inferior (e.g. 0%) podem ser ignorados pela rede, dependendo da política adotada. Além do modelo de punição baseado na redução da reputação, pode-se incorporar um sistema de incentivos por meio de *tokens*, recompensando ações executadas corretamente. O modelo de reputação apresentado, apesar de simples, é modular e pode ser ajustado para influenciar recompensas ou punições. A calibragem dos fatores varia conforme a aplicação que consome a reputação.

5. Conclusão e trabalhos futuros

Este trabalho apresentou a **ShadowRep**, que integra identificadores descentralizados (DIDs), provas de conhecimento zero (ZKPs) e criptografia homomórfica (FHE) para a construção de sistemas de reputação que conciliam verificabilidade com preservação da privacidade. Foram propostas duas abordagens: uma para reputação anônima e outra para reputação às cegas, ambas voltadas a mitigar ataques baseados em exploração oportunista, manipulação da reputação e revelação indevida de dados sensíveis. A prova de conceito demonstrou a viabilidade técnica da abordagem, evidenciando a capacidade do modelo de reagir a comportamentos maliciosos por meio de penalizações proporcionais e auditáveis. Como trabalhos futuros, pretende-se implementar a prova de conceito do sistema de reputação às cegas, estratégia baseada em MPC e FHE. Além disso, é importante realizar uma avaliação do desempenho computacional para validar a viabilidade da **ShadowRep** em dispositivos com recursos limitados. Outro direcionamento relevante é a adaptação da proposta para cenários móveis, por meio da integração com carteiras digitais, como Sphereon ou Bifold.

6. Agradecimentos

Os autores agradecem o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - 316208/2021-3), pela Fundação de Amparo à Pesquisa do Estado da

Bahia (FAPESB - TIC0004/2015) e pelo Air Force Office of Scientific Research (AFOSR - FA9550-23-1-0631).

Referências

Allen, C. (2018). Decentralized identity: Where did it come from and where is it going? `https://www.lifewithalacrity.com/2018/04/decentralized-identity.html`. Acessado: 2025-05-14.

Arshad, J., Azad, M. A., Prince, A., Ali, J., and Papaioannou, T. G. (2022). Reputable–a decentralized reputation system for blockchain-based ecosystems. *IEEE Access*, 10:79948–79961.

Ernstberger, J., Lauinger, J., Elsheimy, F., Zhou, L., Steinhorst, S., Canetti, R., Miller, A., Gervais, A., and Song, D. (2023). Sok: data sovereignty. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*, pages 122–143. IEEE.

Feraudo, A., Romandini, N., Mazzocca, C., Montanari, R., and Bellavista, P. (2024). Diva: A did-based reputation system for secure transmission in vanets using iota. *Computer Networks*, page 110332.

Hoang, A.-T., Ileri, C. U., Sanders, W., and Schulte, S. (2024). zkssi: A zero-knowledge-based self-sovereign identity framework. In *2024 IEEE International Conference on Blockchain (Blockchain)*, pages 276–285.

Kerschbaum, F. (2012). Privacy-preserving computation: (position paper). In *Annual privacy forum*, pages 41–54. Springer.

Liu, D., Alahmadi, A., Ni, J., Lin, X., and Shen, X. (2019). Anonymous reputation system for iiot-enabled retail marketing atop pos blockchain. *IEEE Transactions on Industrial Informatics*, 15(6):3527–3537.

Liu, Y., Xiong, Z., Hu, Q., Niyato, D., Zhang, J., Miao, C., Leung, C., and Tian, Z. (2022). Vrepchain: A decentralized and privacy-preserving reputation system for social internet of vehicles based on blockchain. *IEEE Transactions on Vehicular Technology*, 71(12):13242–13253.

Mazzocca, C., Acar, A., Uluagac, S., Montanari, R., Bellavista, P., and Conti, M. (2025). A survey on decentralized identifiers and verifiable credentials. *IEEE Communications Surveys & Tutorials*, pages 1–1.

Nalini, N., Kumar, A., Sharma, M., Sil, A., and Khan, W. (2023). Pseudonymous decentralised reputation system. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, pages 682–687.

Solomon, R., Weber, R., and Almashaqbeh, G. (2023). smartfhe: Privacy-preserving smart contracts from fully homomorphic encryption. In *2023 IEEE 8th European symposium on security and privacy (euroS&p)*, pages 309–331. IEEE.

Sporny, M., Longley, D., Chadwick, D., Reed, D., and Sabadello, M. (2022). Decentralized identifiers (dids) v1.0. Technical report, W3C. Acessado: 2025-05-02.

Zhou, Z., Wang, M., Yang, C.-N., Fu, Z., Sun, X., and Wu, Q. J. (2021). Blockchain-based decentralized reputation system in e-commerce environment. *Future Generation Computer Systems*, 124:155–167.