Gerenciamento de Identidades e Acesso para Agentes Inteligentes

Andrey Brito[1], **Augusto Suruagy**[2], **Emerson Mello**[3], **Frederico Schardong**[4]

[1]Laboratório de Sistemas Distribuídos
Universidade Federal de Campina Grande (UFCG)
Campina Grande – PB – Brasil

[2]CESAR School
Recife – PE – Brasil

[3]Instituto Federal de Santa Catarina (IFSC)
São José – SC – Brasil

[4]Universidade Federal de Santa Catarina (UFSC)
Araranguá – SC – Brasil

andrey@computacao.ufcg.edu.br, jasm@cesar.org.br,
mello@ifsc.edu.br, frederico.schardong@ufsc.br

Resumo. *A ascensão de agentes de IA autônomos introduz uma nova classe de atores nos sistemas distribuídos, combinando a velocidade e a escalabilidade dos serviços com a autonomia e a imprevisibilidade dos humanos. Modelos de gestão de identidades e acesso (IAM) concebidos para usuários e serviços determinísticos falham em cenários com múltiplos domínios de confiança, operações assíncronas e delegação dinâmica de permissões. Este artigo apresenta o projeto GIAAI (Gerenciamento de Identidades e Acesso para Agentes Inteligentes), uma iniciativa do Centro de Competência EMBRAPII em Segurança Cibernética (CISSA), que avalia ferramentas tradicionais de IAM e propõe práticas, adaptações e extensões para o gerenciamento seguro de agentes.*

1. Introdução

As soluções atuais de gestão de identidade e acesso (IAM – Identity and Access Management), baseadas em protocolos como OAuth 2.0, OpenID Connect e SAML, foram concebidas para interações com usuários humanos ou contas de serviço. Em geral, o comportamento humano nesses sistemas é previsível, com um volume relativamente baixo e controlado de autenticações. Já os serviços atuam de forma determinística, com escopos de operação bem definidos e rigidamente controlados. Assim, tanto para humanos quanto para serviços, a autorização costuma ocorrer por meio de atribuições estáticas de papéis, permissões ou atributos.

A crescente capacidade de agentes inteligentes deu origem ao que tem sido chamado de IA Agêntica, que introduz uma nova categoria de entidades de software autônomas e proativas, capazes de planejar, raciocinar e executar tarefas complexas em múltiplas etapas, com mínima supervisão humana [OpenID Foundation 2025, Cloud Security Alliance 2025]. Embora possam ser vistas como uma forma avançada de serviço, os agentes agem de modo mais semelhante ao de um humano, com comportamento não determinístico. Para que tomem decisões de forma independente, muitas

vezes é necessário conceder-lhes amplos conjuntos de permissões ou atributos de acesso. Além disso, esses agentes podem criar outros agentes para executar tarefas específicas, o que pode levar à proliferação descontrolada de credenciais e identidades no ecossistema [Cloud Security Alliance 2025].

Diversos setores podem se beneficiar dessa tecnologia. No sistema financeiro, por exemplo, agentes autônomos poderiam executar operações críticas em nome de clientes, interagindo com sistemas de múltiplas instituições, ou realizar operações sensíveis sem intervenção humana direta, reduzindo o risco de fraudes internas cometidas por funcionários com credenciais privilegiadas.

Contudo, esses mesmos atributos de autonomia, escala e imprevisibilidade tornam a IA Agêntica um grande desafio para a gestão de identidade e de acesso [OpenID Foundation 2025]. Enquanto, para humanos, a gestão de identidade busca responder a quem pode fazer o quê no sistema, no caso de agentes, é necessário definir o que pode ser feito, em que sequência, a que velocidade e com quais limites contextuais e temporais. Iniciativas de padronização, como o grupo de trabalho WIMSE da IETF, já sinalizam a necessidade de estender os modelos de identidade de cargas de trabalho para contemplar agentes de IA [Ni and Liu 2025].

Este artigo apresenta o projeto GIAAI (Gerenciamento de Identidades e Acesso para Agentes Inteligentes), desenvolvido no âmbito do Centro de Competência EMBRAPII em Segurança Cibernética (CISSA – (Centro Integrado de Segurança em Sistemas e Aplicações), que investiga mecanismos de identidade, autenticação, delegação e confiança interdomínio para viabilizar a adoção segura de agentes autônomos.

2. Desafios

Os desafios do projeto seguem duas direções principais.

2.1. Identidades e Confiança

Assim como humanos e serviços de software, agentes precisam de identidades no momento em que começam a interagir com outros agentes e sistemas. Sem identidades, as ações não seriam observáveis, gerando problemas de depuração, atribuição de custos, autorização, etc. Assim, a primeira necessidade no ciclo de vida de um agente é a identidade. Esta identidade poderia ser simplesmente um identificador único, baseado apenas em propriedades que refletem o contexto básico do agente, como o local onde executa. Além disso, como os protocolos para comunicação entre agentes e suas ferramentas funcionam bem sobre TLS, formatos naturais para estas identidades seriam certificados X.509 ou tokens JWT. Sistemas como o SPIRE [The SPIRE Authors 2026] podem emitir certificados ou JWTs (seguindo o padrão SPIFFE [The SPIFFE Authors 2026]) como identidades a partir de informações sobre o namespace no cluster Kubernetes onde o agente executa (refletindo informações sobre a aplicação) e o projeto na nuvem (tenant), o que representaria um grupo ou projeto dentro da organização.

No entanto, a atribuição de uma identidade única baseada apenas nos aspectos acima perde a oportunidade de abstrair aspectos importantes do ciclo de vida do agente. De forma semelhante à emissão de um documento físico para um humano, o momento da emissão da identidade é uma oportunidade de avaliar se aquela entidade é

genuína e segue os padrões da organização. Assim, o primeiro desafio é que a identidade reflita não só o local, mas também aspectos de integridade e proveniência do agente [Birkholz et al. 2023, SLSA Community 2026]. Desta forma, pode-se afirmar, por exemplo, que o agente X, associado à aplicação Y, é íntegro e foi gerado seguindo os processos de *release* da organização.

O segundo desafio é o aspecto dinâmico. Um agente íntegro no momento inicial não necessariamente estará íntegro depois de algum tempo exposto a interações com o mundo real. Como a revogação de identidades é um problema complexo, seguimos a ideia de que identidades devem ter validades curtas. Desta forma, toda renovação é uma nova oportunidade de revalidar a integridade do agente.

Finalmente, o terceiro desafio é a colaboração entre organizações. Agentes frequentemente precisarão interagir com serviços e sistemas fora da organização. Assim, entidades precisam interagir de forma segura sem relacionamentos operacionais prévios ou federações de identidade previamente negociadas. Embora a federação tradicional de identidades estabeleça confiança entre domínios, ela depende de certificados estáticos, configurações manuais e acordos institucionais, sem verificar o estado do sistema em tempo real, tornando-se incompatível com interações dinâmicas em velocidade de máquina [OpenID Foundation 2025].

2.2. Autenticação, Delegação e Confiança Interdomínio

Estabelecida a identidade, o desafio seguinte é a autenticação e a delegação de autoridade. No modelo predominante hoje, o agente opera por personificação, isto é, com as mesmas credenciais do usuário que o iniciou, o que viola o princípio do menor privilégio e compromete a auditabilidade [OpenID Foundation 2025, Cloud Security Alliance 2025]. É preciso evoluir para um modelo de delegação explícita, no qual o agente apresenta uma identidade própria vinculada a uma autorização específica, permitindo ao sistema identificar não apenas em nome de quem a ação é executada, mas a relação de confiança que a autoriza. Padrões como o OAuth 2.0 e, em especial, o OAuth 2.0 Token Exchange [Jones et al. 2020] são pontos de partida naturais, mas foram concebidos para interações síncronas e escopos estáticos, e precisam ser adaptados para restringir, de forma granular e temporária, o que o agente pode fazer em nome do usuário.

Um agravante é a recursividade: agentes podem criar subagentes para executar tarefas específicas, formando cadeias de delegação. Cada elo da cadeia deve carregar permissões iguais ou mais restritas que o anterior, e a natureza assíncrona das operações não pode resultar em escalada de privilégios além do autorizado na iniciação da atividade. Garantir formalmente essa propriedade, por exemplo, por meio de modelos formais de delegação de confiança com auditabilidade, é um problema em aberto que o projeto GIAAI investiga.

Finalmente, retomando o terceiro desafio da seção anterior, autenticação e delegação precisam funcionar entre domínios administrativos distintos. Mecanismos como o OpenID Federation [Hedberg et al. 2026] permitem estabelecer confiança de forma mais dinâmica que as federações tradicionais, mas ainda é necessário compatibilizar diferentes padrões de autenticação e autorização entre ecossistemas. Além disso, a autonomia e a imprevisibilidade dos agentes exigem que a confiança concedida possa ser revista continuamente: frameworks como o Shared Signals Framework

[Tulshibagwale et al. 2025] viabilizam o compartilhamento de sinais de risco em tempo real entre domínios, permitindo decisões coordenadas de revogação e contenção quando um agente é comprometido [OpenID Foundation 2025].

3. Conclusão

Os padrões de identidade e delegação para agentes de IA ainda estão em formação em fóruns como IETF e OpenID Foundation, contribuir agora significa influenciar as tecnologias que serão consolidadas nos próximos anos. Nesse contexto, o projeto GIAAI articula a pesquisa científica e a demanda industrial para que agentes autônomos sejam adotados em setores críticos, com segurança, auditabilidade e responsabilização.

Agradecimentos

Este trabalho foi financiado pelo projeto GIAAI – Gerenciamento de Identidades e Acesso para Agentes Inteligentes (Fase 1), suportado pelo Centro Integrado de Segurança em Sistemas Avançados (CISSA), com recursos financeiros oriundos do PPI IoT/Manufatura 4.0 do MCTI, através do Termo de Cooperação 08/2024, firmado com a EMBRAPII).

Referências

Birkholz, H., Thaler, D., Richardson, M., Smith, N., and Pan, W. (2023). Remote ATtestation procedures (RATS) architecture. RFC 9334, Internet Engineering Task Force.

Cloud Security Alliance (2025). Agentic ai identity & access management: A new approach. Technical report, Cloud Security Alliance. Accessado: 2026-07-16.

Hedberg, R., Jones, M. B., Solberg, A. Å., Bradley, J., Marco, G. D., and Dzhuvinov, V. (2026). OpenID federation 1.0. OpenID Foundation. Accessado: 2026-07-16.

Jones, M. B., Nadalin, A., Campbell, B., Bradley, J., and Mortimore, C. (2020). OAuth 2.0 token exchange. RFC 8693, Internet Engineering Task Force.

Ni, Y. and Liu, C. P. (2025). Wimse applicability for ai agents. IETF Internet-Draft draft-ni-wimse-ai-agent-identity-01. Work in Progress.

OpenID Foundation (2025). Identity management for agentic ai: The new frontier of authorization, authentication, and security for an ai agent world. Technical report, OpenID Foundation. White Paper.

SLSA Community (2026). SLSA: Supply-chain levels for software artifacts. `https://slsa.dev/`. Accessado: 2026-07-16.

The SPIFFE Authors (2026). SPIFFE: Secure production identity framework for everyone. `https://spiffe.io/`. Accessado: 2026-07-16.

The SPIRE Authors (2026). SPIRE: The SPIFFE runtime environment. `https://spiffe.io/docs/latest/spire-about/`. Accessado: 2026-07-16.

Tulshibagwale, A., Cappalli, T., Deshpande, A., O'Dell, S., and Miel, S. (2025). OpenID shared signals framework specification 1.0. OpenID Foundation. Accessado: 2026-07-16.