REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

SUPERFÍCIE DE ATAQUE NÃO INVENTARIADA E NEGAÇÃO DE SERVIÇO NA CAMADA DE APLICAÇÃO: UM MODELO DE AVALIAÇÃO DA EXPOSIÇÃO DE APIs

UNINVENTORIED ATTACK SURFACE AND APPLICATION-LAYER DENIAL OF SERVICE: AN ASSESSMENT MODEL FOR API EXPOSURE

SUPERFICIE DE ATAQUE NO INVENTARIADA Y DENEGACIÓN DE SERVICIO EN LA CAPA DE APLICACIÓN: UN MODELO DE EVALUACIÓN DE LA EXPOSICIÓN DE APIs

Marcelo de Melo Marcon[1], Carlos Eduardo Costa Oliveira[2]

e2136
https://doi.org/10.67553/sol21.v.2i1.36

PUBLICADO: 10/2026

RESUMO

A migração das arquiteturas digitais para modelos orientados a interfaces de programação de aplicações (APIs) deslocou o vetor de negação de serviço da camada de rede para a camada de aplicação, na qual o ataque pode produzir efeito com volume comparativamente baixo, por explorar a assimetria entre o custo de emitir uma requisição e o de processá-la. O objetivo é propor um modelo de avaliação da exposição de APIs a essa classe de ataque, aplicável a organizações cuja superfície não esteja integralmente inventariada. Adota-se a Design Science Research (DSR), a partir de levantamento documental intencional cujo protocolo de busca, bases, descritores, período e critérios são explicitados. Os resultados compreendem uma taxonomia de cinco classes de exaustão de recursos; a caracterização de Shadow APIs e de endpoints zumbis como dívida de superfície de ataque; quatro pressupostos que limitam as plataformas de proteção; e o modelo proposto, em cinco dimensões e quatro níveis ordinais, com cinco métricas, demonstração analítica e mapeamento normativo. Argumenta-se que a completude do inventário é fator determinante da cobertura efetiva dos demais controles.

PALAVRAS-CHAVE: Negação de serviço. Segurança de APIs. Camada de aplicação. Superfície de ataque. Governança de segurança da informação. DDoS.

ABSTRACT

The migration of digital architectures toward application programming interface (API)-oriented models has shifted the denial-of-service vector from the network layer to the application layer, where attacks can achieve effect with comparatively low volume by exploiting the asymmetry between the cost of issuing a request and the cost of processing it. This article proposes a model for assessing API exposure to this class of attack, applicable to organizations whose attack surface is not fully inventoried. Design Science Research (DSR) is adopted, drawing on a purposive documentary survey whose search protocol, databases, descriptors, period, and inclusion and exclusion criteria are made explicit. The results comprise a taxonomy of five classes of resource exhaustion; the

[1] Doutor Honoris Causa em Ciências Exatas pela World Christian University (WCU), Mestre em Direção Estratégica em Tecnologias da Informação pela Universidad Europea del Atlántico, pós-graduado em MBIS – Segurança da Informação e em Engenharia de Redes e Sistemas de Telecomunicações. Possui mais de 20 anos de experiência profissional em Tecnologia da Informação e Comunicação (TIC), com atuação em ambientes críticos nos setores de saúde, governo e telecomunicações. Sua experiência abrange Segurança da Informação, Cibersegurança, Proteção de Dados, Gestão de Riscos, Datacenters, Virtualização, Linux, Redes, Governança, Continuidade de Negócios e Melhoria de Processos. É instrutor da Escola Superior de Guerra – ADESG/Campinas, membro do Comitê Avançado e Altos Estudos de Tecnologia e Inovação da Digital Law Academy, Diretor do Comitê de Segurança da APDADOS e Imortal da Academia Brasileira de Ciências, Artes, História e Literatura (ABRASCI).

[2] Fundador da vBold LLC (EUA) e vBold Ltda (Brasil), empresas que prestam serviços de inovação e segurança em tecnologia a clientes nacionais e internacionais. Graduação em Direito, pós-graduação em Computação em Nuvem e cursando mestrado em Negócios Globais nos EUA. Autor do livro Navegando Pelos Desafios da Segurança da Informação, membro da Mensa Internacional, Vice-Presidente da APDADOS e Imortal da ABRASCI, além de ter recebido o título Honoris Causa em Ciências Exatas pela WCU. Reconhecido por sua liderança e impacto, alia expertise em tecnologia à missão de transformar vidas e gerar oportunidades.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

characterization of shadow APIs and zombie endpoints as attack-surface debt; four assumptions that structurally limit protection platforms; and the proposed model, with five dimensions, four proposed ordinal levels, five metrics, an analytical demonstration, and a normative mapping.

***KEYWORDS**: Denial of service. API security. Application layer. Attack surface. Information security governance. DDoS.*

RESUMEN
La migración de las arquitecturas digitales hacia modelos orientados a interfaces de programación de aplicaciones (APIs) ha desplazado el vector de denegación de servicio desde la capa de red hacia la capa de aplicación, donde el ataque puede producir efecto con un volumen comparativamente bajo, al explotar la asimetría entre el costo de emitir una petición y el de procesarla. El objetivo es proponer un modelo de evaluación de la exposición de las APIs a esa clase de ataque, aplicable a organizaciones cuya superficie no esté íntegramente inventariada. Se adopta la Design Science Research (DSR), a partir de un levantamiento documental intencional cuyo protocolo de búsqueda, bases, descriptores, período y criterios se explicitan. Los resultados comprenden una taxonomía de cinco clases de agotamiento de recursos, la caracterización de las Shadow APIs y de los endpoints zombis como deuda de superficie de ataque, cuatro supuestos que limitan a las plataformas de protección, y el modelo propuesto, con cinco dimensiones, cuatro niveles ordinales, cinco métricas y una demostración analítica.

***PALABRAS CLAVE**: Denegación de servicio. Seguridad de APIs. Capa de aplicación. Superficie de ataque. Gobernanza de seguridad de la información. DDoS.*

INTRODUÇÃO

A arquitetura das aplicações digitais se reorganizou, ao longo da última década, em torno de interfaces de programação de aplicações (APIs): a superfície exposta deixou de ser o conjunto de páginas navegáveis por um usuário humano e passou a ser o de endpoints acessíveis por um cliente programático. Segundo a Imperva (2026), a atividade automatizada passou a responder por mais de 53% do tráfego da web em 2025, com 27% dos ataques automatizados dirigidos a endpoints de API.

Em paralelo, o vetor de negação de serviço distribuída (DDoS, distributed denial of service) deslocou-se. A Cloudflare (2026) reportou 29,64 trilhões de requisições por protocolo de transferência de hipertexto (HTTP) associadas a negação de serviço no primeiro semestre de 2026, e 90,60% dos ataques encerraram-se em menos de dez minutos, o que inviabiliza respostas dependentes de intervenção humana. A literatura documenta, ademais, a migração do ataque para a camada de aplicação, onde pode produzir efeito com volume comparativamente baixo (Praseed & Thilagam, 2019; Deressu et al., 2026).

O mercado respondeu com a convergência das defesas em plataformas de proteção de aplicações e APIs, categoria de mercado formulada sob a sigla WAAP, de web application and API protection. Essa convergência repousa sobre um pressuposto raramente explicitado: o de que a organização sabe quais endpoints existem, por onde roteiam e qual é o custo computacional normal de cada um. É esse pressuposto que as Shadow APIs e os endpoints zumbis invalidam — rotas de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

depuração esquecidas em produção e versões depreciadas, mas nunca desativadas permanecem alcançáveis e fora de toda política (OWASP, 2023).

Formulamos a proposição que orienta o trabalho: em ambientes orientados a API, a negação de serviço na camada de aplicação não é primariamente um problema de vazão, mas de assimetria entre o custo de emitir uma requisição e o de processá-la, e essa assimetria é governada pela completude do inventário. O adversário não necessita de poder volumétrico; necessita localizar o endpoint mais barato de alcançar que aciona a operação mais cara no backend. Endpoints não inventariados reúnem, por construção, as condições que elevam essa assimetria — ausência de limitação de consumo, de monitoramento e de manutenção —, ainda que este trabalho não produza evidência sobre sua frequência relativa como alvo efetivo.

O objetivo geral é propor um modelo de avaliação dessa exposição, orientado à governança e independente de tecnologia de detecção; os específicos são construir a taxonomia das classes aplicáveis a APIs, caracterizar Shadow APIs e endpoints zumbis, identificar os pressupostos que limitam as plataformas, propor o modelo e mapeá-lo contra referenciais normativos. A pesquisa se justifica porque, no corpus examinado, a produção concentra-se em detecção: Deressu et al. (2026) sistematizam 47 estudos primários nesse eixo, e Tanveer et al. (2025) organizam as vulnerabilidades de APIs REST em torno de autenticação, autorização e injeção, sem contemplar o consumo de recursos.

1 - REFERENCIAL TEÓRICO

1.1 Camada de aplicação e negação de serviço

O modelo de referência para interconexão de sistemas abertos (OSI, open systems interconnection), de Zimmermann (1980), estratifica a comunicação em sete camadas, das quais a sétima corresponde à aplicação; APIs baseadas em transferência de estado representacional (REST), em GraphQL ou em gRPC operam nessa faixa, sobre o protocolo de transferência de hipertexto (HTTP). O modelo descreve, porém, a interoperabilidade de protocolos, e não a semântica de negócio: uma requisição que consulta um saldo e outra que dispara a reconciliação de um extrato anual são protocolarmente indistinguíveis, embora difiram em ordens de grandeza no consumo de processamento. Daí a proteção insuficiente de controles como limites de taxa uniformes por endereço de origem.

Mirkovic e Reiher (2004) organizaram os ataques distribuídos segundo o grau de automação, a dinâmica de taxa e o impacto, e Zargar et al. (2013) sistematizaram os mecanismos de defesa. Praseed e Thilagam (2019) evidenciaram que os ataques na camada de aplicação podem ser conduzidos com volume comparativamente baixo e com maior furtividade; Tripathi e Hubballi (2021) ressaltaram seu caráter dissimulado; e Deressu et al. (2026) registram entre eles

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

uma classe de ataques assimétricos, que acionam operações de servidor custosas a custo mínimo para o atacante.

Desse conjunto extrai-se o conceito central adotado neste trabalho: a razão de assimetria, entendida como a proporção entre o custo computacional imposto ao servidor e o custo incorrido pelo cliente para provocá-lo. A formulação converge com a classe de ataques assimétricos de Deressu et al. (2026), e o que aqui se acrescenta é sua conversão em variável de avaliação, e não de detecção. Em arquiteturas orientadas a API, essa razão pode alcançar valores que tornam irrelevante a discussão sobre largura de banda.

1.2 Ameaças automatizadas e riscos específicos de APIs

O projeto Automated Threats to Web Applications, do OWASP, oferece ontologia com vinte e um eventos de ameaça automatizada (OAT, OWASP Automated Threat), de OAT-001 a OAT-021 (OWASP, 2018). Quatro interessam ao objeto: o OAT-015, Denial of Service, que cobre a exaustão de recursos da aplicação, do banco de dados ou de contas; o OAT-021, Denial of Inventory, que descreve a depleção de disponibilidade sem que a transação se complete; o OAT-003, Cost-Inflation Fraud, relativo à elevação artificial de custos; e o OAT-018, relativo ao mapeamento automatizado pelo qual endpoints não inventariados são localizados por terceiros.

A edição de 2023 do OWASP API Security Top 10 organiza os dez riscos mais críticos de APIs (OWASP, 2023), dos quais três estruturam a análise: o API4:2023, Unrestricted Resource Consumption, relativo à ausência de limitação sobre o consumo de processamento, memória, armazenamento ou banda por cliente; o API6:2023, relativo ao acesso automatizado a fluxos de negócio cuja exaustão prejudica a operação; e o API9:2023, relativo à manutenção inadequada do inventário de APIs ativas, de versões depreciadas e de ambientes expostos.

Duas categorias derivam desse último risco. Uma Shadow API é um endpoint em operação que não consta da especificação, não está registrado no gateway e não possui responsável designado. Um endpoint zumbi é aquele formalmente depreciado que permanece alcançável porque rotas, regras de ingresso ou balanceadores legados nunca foram removidos. Cumpre registrar uma limitação do corpus examinado: nele, praticamente toda a quantificação em circulação sobre essas categorias origina-se de fornecedores de soluções de proteção de APIs, cujo interesse comercial coincide com a magnitude do diagnóstico. Esses dados não são empregados aqui; as categorias são tratadas como construtos conceituais, atestados pelo reconhecimento normativo em API9:2023. Evidência indireta provém de Santos Filho et al. (2025): entre 1.955 especificações OpenAPI analisadas, apenas nove continham os elementos exigidos pelo método que propõem.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1.3 Plataformas de proteção de aplicações e APIs

A sigla WAAP, de web application and API protection, designa a convergência, em plataforma única, de firewall de aplicação web, proteção contra negação de serviço, gestão de bots e segurança de APIs. Trata-se de categoria de mercado, e não de construto acadêmico ou normativo — distinção que este artigo mantém explícita, dado que a literatura sobre o tema é predominantemente produzida por fornecedores da própria categoria. Firewalls operam sobre assinaturas de requisição; gateways aplicam limites de taxa, mas não detectam comportamento anômalo ao longo de uma sessão; soluções volumétricas atuam sobre vazão. Nenhuma isoladamente cobre o espectro que vai da injeção ao abuso de fluxo de negócio.

2 - METODOLOGIA

2.1 Delineamento da pesquisa

A pesquisa é qualitativa quanto à abordagem, aplicada quanto à natureza e documental quanto aos procedimentos. Adotamos a Design Science Research (DSR), delineamento apropriado quando o objetivo não é descrever um fenômeno, mas produzir um artefato que resolva um problema organizacional identificado (Hevner et al., 2004). O percurso segue Peffers et al. (2007), de cujas atividades foram conduzidas a identificação do problema, a definição dos objetivos, o projeto e desenvolvimento e a demonstração; a avaliação em ambiente de uso não foi executada.

2.2 Protocolo do levantamento documental

O levantamento é intencional, e não sistemático: a DSR requer fundamentação suficiente para projetar um artefato, e não estimativa da distribuição de resultados em um campo, finalidade própria da revisão sistemática. Daí uma consequência que explicitamos desde já: as afirmações deste artigo sobre escassez ou ausência de evidência referem-se ao corpus descrito nesta subseção e na seguinte, e não autorizam generalização sobre a totalidade da literatura.

As bases consultadas foram de três tipos. Para literatura revisada por pares, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect e Google Scholar, este último para rastreamento de citações. Para fontes normativas e de conhecimento aberto, os sítios oficiais do OWASP, da International Organization for Standardization e do National Institute of Standards and Technology, sem intermediação de base indexadora. Para telemetria, as publicações próprias dos operadores de infraestrutura, em seus domínios institucionais.

Os descritores foram combinados em inglês e em português por operadores booleanos, aplicados a título e resumo, com rastreamento posterior das referências dos trabalhos retidos. Em inglês: application layer DDoS, application-layer denial of service, low-rate DoS, API security, API inventory, shadow API, zombie endpoint, unrestricted resource consumption, rate limiting, economic

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

denial of sustainability e web application and API protection; em português, os termos correspondentes. O período foi definido por estrato: de 2019 a 2026 para a literatura substantiva; sem limite inferior para obras fundadoras de conceito, o que explica Zimmermann (1980) e Mirkovic e Reiher (2004); edição vigente em setembro de 2026 para as fontes normativas; e os dois ciclos mais recentes de cada operador para a telemetria.

Os critérios de inclusão, cumulativos, foram quatro: tratar de negação de serviço, de consumo de recursos ou da superfície de exposição de APIs; pertencer a um dos três primeiros níveis da hierarquia da subseção 2.3; estar disponível em texto integral em português, inglês ou espanhol; e permitir rastrear cada dado quantitativo até o documento que o produziu. Os de exclusão foram cinco: dado cuja fonte primária não pudesse ser alcançada; percentual atribuído a terceiro sem referência verificável; material de fornecedor como base quantitativa; preprint quando existente a versão revisada por pares; e reprodução secundária de dados já na fonte primária.

Uma lacuna deve ser registrada: não se manteve o registro do número de documentos inicialmente recuperados em cada base antes da triagem. O procedimento é reconstituível quanto às bases, aos descritores, ao período e aos critérios, mas não quanto ao quantitativo de triagem. O corpus retido está integralmente exposto na lista de referências, o que permite auditar o que foi utilizado, ainda que não o que foi descartado.

2.3 Qualificação das fontes e composição do corpus

Dada a natureza do tema, no qual grande parte da produção é gerada por fornecedores de solução, estabeleceu-se hierarquia de qualificação das fontes em quatro níveis: normas técnicas e projetos comunitários de conhecimento aberto; literatura revisada por pares, nela incluída a de fundamentação metodológica; relatórios de telemetria de operadores de infraestrutura, admitidos quando o dado decorre de observação direta da própria rede; e material técnico de fornecedores, admitido para caracterização conceitual e nunca como base quantitativa.

O corpus retido compreende dezenove documentos: quatro no primeiro nível, doze no segundo, três no terceiro e nenhum no quarto. A aplicação dos critérios teve consequência substantiva sobre os resultados: as quantificações em circulação acerca da prevalência de endpoints não governados foram integralmente excluídas, o que levou a tratar essas categorias como construtos conceituais e a converter a ausência de evidência independente em achado do trabalho.

2.4 Estratégia de avaliação do artefato

A avaliação de artefatos em DSR admite posições distintas quanto ao momento e ao paradigma. Venable et al. (2016) organizam-nas em duas dimensões: a finalidade, da formativa à somativa, e o paradigma, do artificial ao naturalístico. A avaliação artificial recorre a análise teórica,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

simulação ou experimento e oferece maior replicabilidade ao custo de abstrair o ambiente real; a naturalística examina o artefato em uso, em organização real. Distinguem ainda a avaliação ex ante, sobre o artefato projetado, da ex post.

Situamos este trabalho nesse quadro. A avaliação conduzida é ex ante, artificial e formativa, realizada por demonstração analítica em cenário ilustrativo, conforme a subseção 3.6, e corresponde à estratégia que Venable et al. (2016) denominam risco técnico e eficácia. Não foi conduzida avaliação naturalística nem somativa. Daí a restrição observada em todo o texto: o artefato é designado modelo proposto, seus níveis são níveis ordinais propostos e sua demonstração é analítica; não se afirma eficácia demonstrada nem utilidade empiricamente validada.

A avaliação naturalística que completaria o ciclo foi projetada, ainda que não executada, e fica registrada para permitir sua condução por outros pesquisadores. A unidade de análise seria a organização que opere superfície de APIs sob plataforma de proteção contratada. O procedimento compreenderia a aplicação do modelo por avaliador externo; a reconciliação entre inventário declarado e alcançabilidade observada em execução, de modo a obter o índice de completude por medição e não por declaração; e o confronto entre a leitura da regra de composição e a da média aritmética. A regra seria infirmada caso organizações com maturidade elevada em D3 a D5 e baixa em D1 apresentassem cobertura efetiva equivalente à de organizações com maturidade elevada em D1.

2.5 Limitações do estudo

Quatro limitações devem ser explicitadas. O ciclo da DSR não foi completado: o artefato foi demonstrado analiticamente, mas não submetido a avaliação naturalística, de modo que sua utilidade é argumentada e não demonstrada. O levantamento é intencional e não conserva o registro do quantitativo de triagem, razão pela qual as afirmações sobre escassez de evidência se restringem ao corpus retido. A ausência de evidência independente sobre a prevalência de endpoints não governados impede calibrar os níveis contra distribuições observadas. E a volatilidade do objeto levou a construir o modelo sobre propriedades de processo.

2.6 Declaração do uso de IA

Houve uso de IA: Os autores declaram que utilizaram o assistente Claude, da Anthropic, modelo Claude Opus 5, acessado entre setembro e outubro de 2026, sem ajuste fino e sem integração a sistemas dos autores, no levantamento e triagem de fontes, na definição do delineamento, na redação, na construção da taxonomia, na formulação das dimensões, níveis e métricas, na demonstração analítica, no mapeamento normativo e na formatação. O uso afetou a totalidade do texto, incluídas as Tabelas 1 a 6 e as referências, e abrange a revisão feita em

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

resposta aos pareceres de avaliação por pares. Coube exclusivamente aos autores a definição do tema e do recorte, a mudança de delineamento do estudo de caso para a proposição de artefato, os critérios de qualificação e exclusão de fontes, a decisão de prosseguir sem as bases quantitativas sobre endpoints não governados, a de manter na taxonomia as duas classes que não satisfazem a definição estrita de negação de serviço mediante critério declarado, e a validação final. Não foram fornecidos dados pessoais, sensíveis ou protegidos por sigilo, e nenhuma organização foi identificada, razão pela qual não houve tratamento de dados sujeitos à Lei nº 13.709/2018. Os prompts e registros metodologicamente relevantes estão arquivados pelos autores e disponíveis mediante solicitação. Cada referência foi conferida contra a fonte primária, e nenhum dado, citação ou conclusão foi aceito sem verificação humana. Os autores assumem responsabilidade integral pelo conteúdo final.

3 - RESULTADOS E DISCUSSÃO

3.1 Taxonomia da exaustão de recursos em APIs

A primeira contribuição organiza as classes de exaustão de recursos aplicáveis a APIs segundo o recurso exaurido, e não segundo o protocolo, uma vez que a classificação protocolar não distingue operações cujo custo difere em ordens de grandeza.

Tabela 1

Taxonomia das classes de exaustão de recursos com efeito de negação de serviço em APIs

Classe	Mecanismo de exaustão	Recurso alvo	Correspondência
Exaustão volumétrica	Grande número de requisições válidas contra endpoints de custo ordinário.	Processamento e conexões	OAT-015; API4:2023
Exaustão por conexão	Conexões abertas por longo tempo com banda mínima, incluída leitura e escrita lentas.	Descritores, threads e pool de conexões	OAT-015; API4:2023
Amplificação por custo de consulta	Operações cujo custo de processamento é desproporcional ao de solicitá-las: consultas aninhadas, agregações, paginação irrestrita.	Processamento, memória e banco de dados	API4:2023
Exaustão de fluxo de negócio	Consumo automatizado de disponibilidade finita sem conclusão da transação.	Estoque, reservas, cotas e filas	OAT-021; API6:2023
Inflação de custo, ou EDoS	Acionamento repetido de operações que encadeiam serviços tarifados ou infraestrutura elástica.	Orçamento de infraestrutura e de serviços	OAT-003; API4:2023

Nota. Elaborado pelos autores com base em OWASP (2018, 2023), Praseed e Thilagam (2019), Aldhyani e Alkahtani (2022) e Deressu et al. (2026).

A taxonomia admite estratificação quanto ao próprio conceito de negação de serviço. As três primeiras classes constituem negação de serviço em sentido estrito: produzem indisponibilidade por exaustão de recurso computacional. A quarta consome disponibilidade finita sem consumir recurso computacional, e o próprio OWASP a designa negação de inventário ao catalogá-la como

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

OAT-021. A quinta não produz indisponibilidade: corresponde ao construto que a literatura sobre computação em nuvem denomina negação econômica de sustentabilidade (EDoS, economic denial of sustainability), que explora o modelo de pagamento por uso de maneira que o dano se materialize na fatura enquanto o serviço permanece disponível (Aldhyani & Alkahtani, 2022).

Mantemos as duas últimas classes por razão que convém declarar, já que nenhuma satisfaz a definição estrita: ambas compartilham com as três primeiras o mecanismo gerador, que é a assimetria entre o custo de solicitar e o de atender, e os controles que as mitigam. Não as mantemos por produzirem indisponibilidade, pois a quinta não produz. Em arquiteturas elásticas, o dimensionamento automático que preserva a disponibilidade é precisamente o mecanismo que materializa o prejuízo, de modo que a defesa contra a terceira classe pode agravar a quinta.

3.2 Endpoints não governados como dívida de superfície de ataque

A segunda contribuição distingue duas categorias frequentemente tratadas como sinônimas, que têm origens distintas e demandam tratamentos distintos.

Tabela 2

Distinção conceitual entre Shadow APIs e endpoints zumbis

Critério	Shadow API	Endpoint zumbi
Origem	Publicação fora do processo formal de governança.	Alcançabilidade mantida após depreciação ou substituição.
Inventário	Nunca registrado.	Registrado como inativo, sem correspondência com o estado real.
Risco predominante	Ausência de limitação de consumo, autenticação e monitoramento.	Vulnerabilidades já corrigidas nas versões vigentes.
Descoberta	Análise de tráfego em execução e inspeção de artefatos de implantação.	Reconciliação entre inventário declarado e alcançabilidade observada.

Nota. Elaborado pelos autores com base em OWASP (2023).

As duas categorias diferem também no ciclo de vida — a Shadow API é ativa e em evolução, sem responsável designado; o endpoint zumbi é estático e sem correções. Designamos o conjunto como dívida de superfície de ataque, por analogia à dívida técnica: acumula-se como subproduto da velocidade de entrega e incide juros, pois cada endpoint não governado permanece exposto a vulnerabilidades já corrigidas em outros pontos.

3.3 Pressupostos que limitam as plataformas de proteção

A terceira contribuição resulta da análise da arquitetura de operação das plataformas. Identificaram-se quatro pressupostos cuja violação degrada a eficácia independentemente da qualidade da implementação. O de roteamento: a plataforma inspeciona apenas o tráfego que atravessa seu ponto de aplicação, e endpoints alcançáveis por caminhos que contornam o gateway ficam fora de seu alcance. O de especificação: a proteção baseada em modelo positivo compara a

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

requisição com o contrato declarado, e endpoints sem especificação recaem sobre regras genéricas — dependência reconhecida pela própria literatura que a propõe (Santos Filho et al., 2025).

O de linha de base: a detecção comportamental depende de histórico que caracterize o uso normal, e endpoints de baixo tráfego, recém-expostos ou intermitentes não o acumulam. E o de custo uniforme: limites expressos em requisições por unidade de tempo tratam como equivalentes operações cujo custo difere em ordens de grandeza. Os três primeiros são violados precisamente pelos endpoints não governados; o quarto, por qualquer endpoint cujo custo não tenha sido caracterizado. Desses enunciados decorre o desenho do modelo.

3.4 O modelo de avaliação proposto

O artefato consiste em um modelo de cinco dimensões, cada uma aferida em quatro níveis ordinais propostos. As três primeiras endereçam os pressupostos de roteamento, especificação e linha de base; a quarta, o de custo uniforme; a quinta trata da resposta, cuja tempestividade é condicionada pela duração observada dos ataques.

Tabela 3

Dimensões do modelo de avaliação da exposição de APIs

Dimensão	Objeto de avaliação	Evidência no nível gerenciado
D1 — Completude do inventário	Identificar, de forma contínua e em execução, os endpoints alcançáveis.	Descoberta automatizada, com reconciliação periódica entre observado e declarado.
D2 — Governança de ciclo de vida	Conduzir o endpoint do registro à desativação efetiva, com responsável designado.	Política de depreciação com prazo e verificação de alcançabilidade após a desativação.
D3 — Caracterização de custo	Conhecer o custo computacional de cada operação e de suas dependências.	Endpoints classificados por faixa de custo, com limites diferenciados por faixa.
D4 — Controle de consumo	Limites e cotas proporcionais ao custo e ao grau de autenticação.	Cotas por identidade e por custo agregado, com teto para tráfego não autenticado.
D5 — Detecção e resposta autônoma	Identificar automação anômala e conter sem intervenção humana.	Contenção automática por política, com registro auditável e reversão.

Nota. Elaborado pelos autores.

Os quatro níveis são uniformes entre as dimensões e ordinais: a progressão indica direção de melhoria, e não distância mensurada. No nível 1, Inicial, a dimensão não é objeto de processo definido. No nível 2, Definido, há processo documentado e responsável designado, mas execução periódica e manual. No nível 3, Gerenciado, o processo é contínuo e automatizado. No nível 4, Otimizado, integra-se ao ciclo de entrega, de modo que a violação da política impede a promoção da mudança ao ambiente produtivo.

O modelo adota uma regra de composição que expressa a proposição central do trabalho: o nível global não corresponde à média das dimensões, mas é limitado pelo nível de D1. A justificativa é lógica: controles de consumo, caracterização de custo e detecção comportamental aplicam-se a endpoints conhecidos, de modo que, se a organização desconhece parcela de sua superfície, esses controles operam sobre subconjunto de tamanho indeterminado e sua cobertura

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

agregada é indeterminada na mesma medida. Uma organização com maturidade máxima em D3, D4 e D5 e inicial em D1 não possui proteção elevada com uma lacuna pontual: possui proteção de extensão desconhecida. A regra é proposta, e não demonstrada.

3.5 Métricas operacionais

Para tornar o modelo aplicável, propõem-se cinco métricas, calculáveis a partir de dados que a operação já produz. A quarta delas desloca a pergunta de gestão — quantas requisições por segundo o sistema suporta — para outra, mais informativa: qual é a operação mais cara que um adversário sem credenciais consegue acionar.

Tabela 4

Métricas operacionais propostas

Métrica	Definição operacional	Interpretação
Índice de completude do inventário	Endpoints observados em execução que constam do inventário, sobre o total observado.	Fração da superfície sob governança; indicador limitante do modelo.
Razão de endpoints não roteados	Endpoints ativos alcançáveis por caminho que não atravessa o ponto de aplicação.	Mede a violação do pressuposto de roteamento.
Latência de aposentadoria	Tempo mediano entre a depreciação anunciada e a verificação de inalcançabilidade.	Estima a taxa de acumulação de endpoints zumbis.
Custo máximo por requisição não autenticada	Maior custo observado entre as operações acionáveis sem autenticação.	Aproxima o teto da razão de assimetria disponível a um adversário anônimo.
Cobertura de política de automação	Endpoints inventariados sob política de detecção e limite de consumo.	Distância entre a cobertura contratada e a efetiva.

Nota. Elaborado pelos autores.

3.6 Demonstração analítica do modelo

Para demonstrar a aplicabilidade do modelo, aplicamo-lo a uma configuração ilustrativa, construída para esse fim e que não corresponde a organização observada: os valores não constituem dado empírico, mas parâmetros escolhidos para exercitar o instrumento. A demonstração é, nos termos da subseção 2.4, avaliação ex ante, artificial e formativa. A configuração supõe uma organização de serviços financeiros que opera plataforma de proteção contratada, declara trezentos endpoints em seu gateway e apresenta trezentos e oitenta alcançáveis em observação de tráfego; a política de consumo cobre todo o inventário declarado; não há classificação por faixa de custo; e a reconciliação é manual e anual.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 5

Demonstração analítica: configuração ilustrativa e leitura do modelo

Dimensão	Nível	Evidência na configuração ilustrativa	Métrica
D1 — Completude do inventário	1 — Inicial	Reconciliação manual e anual; sem descoberta em execução.	Completude = 300/380 = 0,79
D2 — Governança de ciclo de vida	2 — Definido	Política de depreciação documentada, verificação manual.	Latência de aposentadoria = 180 dias
D3 — Caracterização de custo	1 — Inicial	Custo das operações não caracterizado.	Custo máximo não autenticado = não determinado
D4 — Controle de consumo	3 — Gerenciado	Cotas por identidade automatizadas, sobre todo o inventário declarado.	Cobertura = 1,00 do inventário; 0,79 do observado
D5 — Detecção e resposta	3 — Gerenciado	Contenção automática por política, com registro auditável.	Endpoints não roteados = 0,21
Leitura global	1 — Inicial	Regra de composição: limitada pelo nível de D1.	Média aritmética = 2,0

Nota. Elaborado pelos autores. Configuração ilustrativa, construída para fins de demonstração; não corresponde a organização observada e os valores não constituem dado empírico.

A demonstração produz um resultado que não é evidente antes de o instrumento ser aplicado. As duas leituras divergem: a média aritmética situa a organização no nível 2 e sugere esforço sobre D2; a regra de composição situa-a no nível 1 e o direciona para D1 e D3. A divergência não é de grau, e sim de objeto de investimento. Sua origem está na métrica de cobertura: a política de consumo alcança todo o inventário declarado e reporta cobertura integral sob o indicador que a organização acompanha, enquanto alcança 79% da superfície observada. Os 21% restantes não estão sob limite algum, e a ausência de caracterização de custo impede estimar o teto da razão de assimetria disponível a um adversário sem credenciais. A demonstração estabelece que o modelo é aplicável e que sua regra de composição é operativa; não estabelece que a recomendação seja correta em ambiente real.

3.7 Mapeamento normativo e discussão

Tabela 6

Mapeamento das dimensões do modelo contra referenciais normativos

Dimensão	OWASP API	ISO/IEC 27001:2022	NIST CSF 2.0
D1 — Completude do inventário	API9:2023	5.9 Inventário de ativos	Identificar
D2 — Governança de ciclo de vida	API9:2023; API8:2023	8.9 Gestão de configuração; 8.32 Gestão de mudanças	Governar; Proteger
D3 — Caracterização de custo	API4:2023	8.6 Gestão de capacidade	Identificar; Proteger

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

Dimensão	OWASP API	ISO/IEC 27001:2022	NIST CSF 2.0
D4 — Controle de consumo	API4:2023; API6:2023	8.26 Requisitos de segurança de aplicações	Proteger
D5 — Detecção e resposta	API4:2023	8.16 Monitoramento; 5.7 Inteligência de ameaças	Detectar; Responder

Nota. Elaborado pelos autores com base em OWASP (2023), International Organization for Standardization (2022) e National Institute of Standards and Technology (2024). Os identificadores de controle devem ser conferidos contra o texto integral da norma antes de uso em auditoria.

O modelo dialoga criticamente com a orientação predominante no campo, que concentra esforço na detecção, particularmente por aprendizado de máquina (Praseed & Thilagam, 2019; Tripathi & Hubballi, 2021; Deressu et al., 2026). Essa produção pressupõe resolvido o problema que este trabalho identifica como anterior: para que um detector observe tráfego anômalo em um endpoint, é necessário que o tráfego desse endpoint seja observado. A eficácia demonstrada em ambiente experimental, no qual o conjunto de endpoints é conhecido por construção, não se transfere a ambiente produtivo no qual essa premissa não se sustenta. Resta uma tensão: D1 demanda observabilidade justamente onde ela falta.

4 - CONSIDERAÇÕES

Este trabalho propôs um modelo de avaliação da exposição de APIs à negação de serviço na camada de aplicação, construído pela estratégia da Design Science Research e destinado a organizações cuja superfície de ataque não esteja integralmente inventariada.

Construímos uma taxonomia de cinco classes organizadas pelo recurso exaurido, evidenciando-se que apenas uma corresponde à concepção volumétrica corrente e que duas se situam no limite do conceito de negação de serviço. Distinguimos Shadow APIs e endpoints zumbis, propondo a designação do conjunto como dívida de superfície de ataque. Identificamos quatro pressupostos de eficácia das plataformas, dos quais os três primeiros são violados por endpoints não governados.

A contribuição teórica consiste na formulação da razão de assimetria como variável explicativa central e na proposição de que a completude do inventário é fator determinante da cobertura efetiva e da governança dos demais controles. Convém precisar seu alcance: o estudo sustenta que o inventário delimita a extensão sobre a qual os demais controles operam e que, desconhecida essa extensão, a cobertura agregada é indeterminada; não sustenta que a completude seja condição universal de eficácia. Com destaque para a ausência de avaliação naturalística, o modelo deve ser lido como proposta sujeita a exame empírico.

Como agenda futura, sugerimos a avaliação naturalística do modelo conforme o protocolo da subseção 2.4; a produção de evidência independente sobre a prevalência de endpoints não governados; a investigação experimental da razão de assimetria em interfaces REST e declarativas;

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

e a análise das implicações regulatórias da indisponibilidade de serviços mediados por API em setores regulados no Brasil.

REFERÊNCIAS

Aldhyani, T. H. H., & Alkahtani, H. (2022). Artificial intelligence algorithm-based economic denial of sustainability attack detection systems: Cloud computing environments. *Sensors, 22*(13), 4685. https://doi.org/10.3390/s22134685

Cloudflare. (2026). *DDoS threat report for 2026 Q2*. Cloudforce One. https://blog.cloudflare.com/ddos-threat-report-2026-h1/

Deressu, T. F., Beyene, A. M., & Gemechu, A. Y. (2026). A survey of application layer distributed denial of service detection: Approaches, models, and datasets. *Artificial Intelligence Review, 59*(6), 139. https://doi.org/10.1007/s10462-026-11528-3

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Imperva. (2025). *2025 bad bot report*. Thales Group. https://www.imperva.com/resources/wp-content/uploads/sites/6/reports/2025-Bad-Bot-Report.pdf

Imperva. (2026). *Bad bot report 2026: Bad bots in the agentic age*. Thales Group. https://www.imperva.com/blog/bad-bot-report-2026-bots-agentic-age/

International Organization for Standardization. (2022). *ISO/IEC 27001:2022 — Information security, cybersecurity and privacy protection: Information security management systems: Requirements*.

Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review, 34*(2), 39–53. https://doi.org/10.1145/997150.997156

National Institute of Standards and Technology. (2024). *The NIST Cybersecurity Framework (CSF) 2.0* (NIST CSWP 29). U.S. Department of Commerce. https://doi.org/10.6028/NIST.CSWP.29

OWASP. (2018). *OWASP automated threat handbook: Web applications* (Version 1.2). Open Worldwide Application Security Project. https://owasp.org/www-project-automated-threats-to-web-applications/

OWASP. (2023). *OWASP API security top 10 — 2023*. Open Worldwide Application Security Project. https://owasp.org/API-Security/editions/2023/en/0x00-toc/

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Praseed, A., & Thilagam, P. S. (2019). DDoS attacks at the application layer: Challenges and research perspectives for safeguarding web applications. *IEEE Communications Surveys & Tutorials, 21*(1), 661–685. https://doi.org/10.1109/COMST.2018.2870658

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Santos Filho, A., Rodríguez, R. J., & Feitosa, E. L. (2025). Automated broken object-level authorization attack detection in REST APIs through OpenAPI to colored Petri nets transformation. *International Journal of Information Security, 24*(2), 83. https://doi.org/10.1007/s10207-024-00970-5

Tanveer, F., Iradat, F., Iqbal, W., & Ahmad, A. (2025). Towards secure APIs: A survey on RESTful API vulnerability detection. *Computers, Materials & Continua, 84*(3), 4223–4257.

Tripathi, N., & Hubballi, N. (2021). Application layer denial-of-service attacks and defense mechanisms: A survey. *ACM Computing Surveys, 54*(4), 1–33. https://doi.org/10.1145/3448291

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems, 25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Zargar, S. T., Joshi, J., & Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials, 15*(4), 2046–2069. https://doi.org/10.1109/SURV.2013.031413.00127

Zimmermann, H. (1980). OSI reference model — The ISO model of architecture for open systems interconnection. *IEEE Transactions on Communications, 28*(4), 425–432. https://doi.org/10.1109/TCOM.1980.1094702

APÊNDICE A — SIGLAS E TERMOS TÉCNICOS

API, application programming interface; CSF, Cybersecurity Framework; DDoS, distributed denial of service; DSR, Design Science Research; EDoS, economic denial of sustainability; GraphQL, linguagem de consulta para APIs; gRPC, arcabouço de chamada de procedimento remoto; HTTP, hypertext transfer protocol; IA, inteligência artificial; ISO/IEC, International Organization for Standardization e International Electrotechnical Commission; NIST, National Institute of Standards and Technology; OAT, OWASP automated threat; OSI, open systems interconnection; OWASP, Open Worldwide Application Security Project; REST, representational state transfer; WAAP, web application and API protection.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.