REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

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

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

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

PUBLISHED: 10/2026

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 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. The article argues that inventory completeness is a determining factor for the effective coverage of all other controls.

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

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

[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.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

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.

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.

INTRODUCTION

Over the past decade, the architecture of digital applications has reorganized itself around application programming interfaces (APIs): the exposed surface is no longer the set of pages a human user can browse, but the set of endpoints reachable by a programmatic client. According to Imperva (2026), automated activity came to account for more than 53% of web traffic in 2025, with 27% of automated attacks directed at API endpoints.

In parallel, the distributed denial-of-service (DDoS) vector has shifted. Cloudflare (2026) reported 29.64 trillion Hypertext Transfer Protocol (HTTP) requests associated with denial of service in the first half of 2026, and 90.60% of attacks ended in under ten minutes, which renders response architectures dependent on human intervention unworkable. More significantly, the literature documents the migration of the attack to the application layer, where it can achieve effect with comparatively low volume (Praseed & Thilagam, 2019; Deressu et al., 2026).

The market responded by converging defenses into web application and API protection platforms, a market category formulated under the acronym WAAP. That convergence rests on a rarely stated assumption: that the organization knows which endpoints exist, where they route, and what the normal computational cost of each one is. It is this assumption that shadow APIs and

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

zombie endpoints invalidate — debug routes left in production and versions deprecated but never decommissioned remain reachable and outside every policy (OWASP, 2023).

We state the proposition that guides this work: in API-oriented environments, application-layer denial of service is not primarily a throughput problem but one of asymmetry between the cost of issuing a request and the cost of processing it, and that asymmetry is governed by inventory completeness. The adversary does not need volumetric power; it needs to locate the endpoint that is cheapest to reach and triggers the most expensive backend operation. Uninventoried endpoints combine, by construction, the conditions that raise that asymmetry — absence of consumption limits, of monitoring, and of maintenance — although this work produces no evidence as to their relative frequency as actual targets.

The general objective is to propose a model for assessing that exposure, oriented toward governance and independent of any detection technology; the specific objectives are to build a taxonomy of the classes applicable to APIs, to characterize shadow APIs and zombie endpoints, to identify the assumptions that limit protection platforms, to propose the model, and to map it against normative frameworks. The research is warranted because, within the examined corpus, scholarly output concentrates on detection: Deressu et al. (2026) systematize 47 primary studies along that axis, and Tanveer et al. (2025) organize REST API vulnerabilities around authentication, authorization, and injection, without addressing resource consumption.

1 - THEORETICAL FRAMEWORK

1.1 The application layer and denial of service

The Open Systems Interconnection (OSI) reference model, formalized by Zimmermann (1980), stratifies communication into seven layers, the seventh of which corresponds to the application; APIs based on Representational State Transfer (REST), on GraphQL, or on gRPC operate in that band, over the Hypertext Transfer Protocol (HTTP). The model describes protocol interoperability, however, not business semantics: a request that retrieves an account balance and one that triggers the reconciliation of an annual statement are indistinguishable at the protocol level, even though they may differ by orders of magnitude in processing consumption. Hence the insufficient protection afforded by controls such as uniform rate limits per source address.

Mirkovic and Reiher (2004) organized distributed attacks according to degree of automation, rate dynamics, and impact, and Zargar et al. (2013) systematized defense mechanisms. Praseed and Thilagam (2019) showed that application-layer attacks can be carried out with comparatively low volume and with greater stealth; Tripathi and Hubballi (2021) emphasized their dissimulated character; and Deressu et al. (2026) record among them a class of asymmetric attacks, which trigger costly server operations at minimal cost to the attacker.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

From this body of work we draw the central concept adopted here: the asymmetry ratio, understood as the proportion between the computational cost imposed on the server and the cost incurred by the client to provoke it. The formulation converges with the class of asymmetric attacks recorded by Deressu et al. (2026), and what is added here is its conversion into a variable for assessment rather than for detection. In API-oriented architectures, that ratio can reach values that render the discussion of bandwidth irrelevant.

1.2 Automated threats and API-specific risks

The OWASP Automated Threats to Web Applications project offers a vendor-independent ontology of twenty-one automated threat events (OAT, OWASP Automated Threat), from OAT-001 to OAT-021 (OWASP, 2018). Four are pertinent here: OAT-015, Denial of Service, which covers the exhaustion of application, database, or account resources; OAT-021, Denial of Inventory, which describes the depletion of availability without the transaction being completed; OAT-003, Cost-Inflation Fraud, concerning the artificial inflation of costs; and OAT-018, concerning the automated mapping by which uninventoried endpoints are located by third parties.

The 2023 edition of the OWASP API Security Top 10 organizes the ten most critical API risks (OWASP, 2023), three of which structure the analysis: API4:2023, Unrestricted Resource Consumption, concerning the absence of limits on the consumption of processing, memory, storage, or bandwidth per client; API6:2023, concerning automated access to business flows whose exhaustion impairs operations; and API9:2023, concerning inadequate maintenance of the inventory of active APIs, deprecated versions, and exposed environments.

Two categories derive from that last risk. A shadow API is an endpoint in operation that does not appear in the specification, is not registered in the gateway, and has no designated owner. A zombie endpoint is one that has been formally deprecated yet remains reachable because legacy routes, ingress rules, or load balancers were never removed. A limitation of the examined corpus must be recorded: within it, virtually all circulating quantification of these categories originates from vendors of API protection solutions, whose commercial interest coincides with the magnitude of the diagnosis. Those figures are not used here; the categories are treated as conceptual constructs, attested by their normative recognition in API9:2023. Indirect evidence comes from Santos Filho et al. (2025): among 1,955 OpenAPI specifications analyzed, only nine contained the elements required by the method they propose.

1.3 Web application and API protection platforms

The acronym WAAP, for web application and API protection, designates the convergence, within a single platform, of a web application firewall, denial-of-service protection, bot management, and API security. It is a market category, not an academic or normative construct — a distinction this article keeps explicit, given that the literature on the subject is predominantly produced by vendors

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

within that same category. Firewalls operate on request signatures; gateways apply rate limits but do not detect anomalous behavior across a session; volumetric solutions act on throughput. None of them alone covers the spectrum running from injection to business-flow abuse.

2 - METHODOLOGY

2.1 Research design

The research is qualitative in approach, applied in nature, and documentary in procedure. We adopt Design Science Research (DSR), a design appropriate when the objective is not to describe a phenomenon but to produce an artifact that solves an identified organizational problem (Hevner et al., 2004). The path follows Peffers et al. (2007), of whose activities we carried out problem identification, definition of the objectives of a solution, design and development, and demonstration; evaluation in a use environment was not performed.

2.2 Documentary survey protocol

The survey is purposive rather than systematic: DSR requires grounding sufficient to design an artifact, not an estimate of the distribution of findings in a field, which is the purpose proper to a systematic review. Hence a consequence we state at the outset: this article's claims about the scarcity or absence of evidence refer to the corpus described in this subsection and the next, and do not license generalization over the literature as a whole.

The databases consulted were of three kinds. For peer-reviewed literature, IEEE Xplore, the ACM Digital Library, SpringerLink, ScienceDirect, and Google Scholar, the last of these for citation tracking. For normative and open-knowledge sources, the official sites of OWASP, the International Organization for Standardization, and the National Institute of Standards and Technology, without the intermediation of an indexing database. For telemetry, the infrastructure operators' own publications, on their institutional domains.

Descriptors were combined in English and Portuguese by Boolean operators, applied to title and abstract, with subsequent backward tracking of the references of retained works. In English: 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 denial of sustainability, and web application and API protection; in Portuguese, the corresponding terms. The period was defined by stratum: 2019 to 2026 for the substantive literature; no lower bound for concept-founding works, which accounts for Zimmermann (1980) and Mirkovic and Reiher (2004); the edition in force as of September 2026 for normative sources; and the two most recent cycles from each operator for telemetry.

The inclusion criteria, applied cumulatively, were four: to address denial of service, resource consumption, or the API exposure surface; to belong to one of the first three tiers of the hierarchy

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

set out in subsection 2.3; to be available in full text in Portuguese, English, or Spanish; and to permit tracing every quantitative datum back to the document that produced it. The exclusion criteria were five: a datum whose primary source could not be reached; a percentage attributed to a third party without a verifiable reference to the original; vendor material used as a quantitative basis; a preprint where a peer-reviewed version exists; and the secondary reproduction of data already present in the primary source.

One gap must be recorded: no record was kept of the number of documents initially retrieved from each database before screening. The procedure is reconstructible as to databases, descriptors, period, and criteria, but not as to the screening count. The retained corpus is set out in full in the reference list, which allows the reader to audit what was used, if not what was discarded.

2.3 Source qualification and corpus composition

Given the nature of the subject, in which much of the available output is produced by solution vendors, an explicit four-tier hierarchy of source qualification was established: technical standards and open community knowledge projects; peer-reviewed literature, including works of methodological grounding; telemetry reports from infrastructure operators, admitted when the datum derives from direct observation of the operator's own network; and vendor technical material, admitted for conceptual characterization and never as a quantitative basis.

The retained corpus comprises nineteen documents: four in the first tier, twelve in the second, three in the third, and none in the fourth. Applying these criteria had a substantive consequence for the results: the circulating quantifications of the prevalence of ungoverned endpoints were excluded in full, which led us to treat those categories as conceptual constructs and to convert the absence of independent evidence into a finding of the study.

2.4 Artifact evaluation strategy

Artifact evaluation in DSR admits distinct positions as to timing and paradigm. Venable et al. (2016) organize them along two dimensions: purpose, running from formative to summative, and paradigm, running from artificial to naturalistic. Artificial evaluation relies on theoretical analysis, simulation, or experiment and offers greater replicability at the cost of abstracting from the real environment; naturalistic evaluation examines the artifact in use, within a real organization. They further distinguish ex ante evaluation, bearing on the designed artifact, from ex post evaluation.

We situate this work within that frame. The evaluation conducted is ex ante, artificial, and formative, carried out by analytical demonstration in an illustrative scenario, as set out in subsection 3.6, and corresponds to the strategy Venable et al. (2016) term technical risk and efficacy. No naturalistic or summative evaluation was conducted. Hence the restraint observed throughout the text: the artifact is designated a proposed model, its levels are proposed ordinal levels, and its

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

demonstration is analytical; no claim is made of demonstrated efficacy or of empirically validated usefulness.

The naturalistic evaluation that would complete the cycle has been designed, though not executed, and is recorded here so that other researchers may carry it out. The unit of analysis would be an organization operating an API surface under a contracted protection platform. The procedure would comprise application of the model by an external assessor; reconciliation between the declared inventory and reachability observed at runtime, so as to obtain the completeness index by measurement rather than by declaration; and comparison between the reading produced by the composition rule and that produced by the arithmetic mean. The rule would be disconfirmed if organizations with high maturity in D3 through D5 and low maturity in D1 showed effective coverage equivalent to that of organizations with high maturity in D1.

2.5 Limitations of the study

Four limitations must be made explicit. The DSR cycle was not completed: the artifact was demonstrated analytically but not submitted to naturalistic evaluation, so that its usefulness is argued rather than demonstrated. The survey is purposive and does not preserve a record of the screening count, which is why claims about the scarcity of evidence are confined to the retained corpus. The absence of independent evidence on the prevalence of ungoverned endpoints precludes calibrating the levels against observed distributions. And the volatility of the subject led us to build the model on process properties.

2.6 Declaration of artificial intelligence use

AI was used: The authors declare that they used the assistant Claude, from Anthropic, model Claude Opus 5, accessed between September and October 2026, without fine-tuning and without integration into the authors' systems, in the survey and screening of sources, the definition of the research design, the writing, the construction of the taxonomy, the formulation of the dimensions, levels and metrics, the analytical demonstration, the normative mapping, and the formatting. The use affected the entirety of the text, including Tables 1 through 6 and the references, and extends to the revision made in response to the peer-review reports. The following were exclusively the authors' responsibility: the definition of the topic and its delimitation; the change of design from a case study to the proposition of an artifact; the source qualification and exclusion criteria; the decision to proceed without the quantitative bases on ungoverned endpoints; the decision to retain in the taxonomy the two classes that do not satisfy the strict definition of denial of service, subject to a declared criterion; and the final validation. No personal, sensitive, or legally protected data were supplied, and no organization was identified, which is why there was no processing of data subject to Brazilian Law No. 13,709/2018 (Brazilian General Data Protection Law). The prompts and methodologically relevant records are archived by the authors and available

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

upon request. Every reference was checked against the primary source, and no datum, citation, or conclusion was accepted without human verification. The authors assume full responsibility for the final content.

3 - RESULTS AND DISCUSSION

3.1 Taxonomy of resource exhaustion in APIs

The first contribution organizes the classes of resource exhaustion applicable to APIs according to the resource exhausted rather than the protocol, since protocol-level classification does not distinguish operations whose cost differs by orders of magnitude.

Table 1

Taxonomy of resource-exhaustion classes producing denial-of-service effects in APIs

Class	Exhaustion mechanism	Target resource	Correspondence
Volumetric exhaustion	Large number of protocol-valid requests against endpoints of ordinary cost.	Processing and connections	OAT-015; API4:2023
Connection exhaustion	Connections held open at minimal bandwidth, including deliberately slow reads and writes.	Descriptors, threads, and connection pool	OAT-015; API4:2023
Query-cost amplification	Operations whose processing cost is disproportionate to the cost of requesting them: nested queries, aggregations, unrestricted pagination.	Processing, memory, and database	API4:2023
Business-flow exhaustion	Automated consumption of finite availability without completing the transaction.	Inventory, reservations, quotas, and queue slots	OAT-021; API6:2023
Cost inflation, or EDoS	Repeated triggering of operations that chain to metered services or elastic infrastructure.	Infrastructure and contracted-service budget	OAT-003; API4:2023

Note. Compiled by the authors based on OWASP (2018, 2023), Praseed and Thilagam (2019), Aldhyani and Alkahtani (2022), and Deressu et al. (2026).

The taxonomy admits stratification with respect to the concept of denial of service itself. The first three classes constitute denial of service in the strict sense: they produce unavailability through the exhaustion of a computational resource. The fourth consumes finite availability without consuming a computational resource, and OWASP itself designates it denial of inventory in cataloging it as OAT-021. The fifth produces no unavailability: it corresponds to the construct that the cloud computing literature calls economic denial of sustainability (EDoS), which exploits the pay-per-use model so that the harm materializes on the invoice while the service remains available (Aldhyani & Alkahtani, 2022).

We retain the last two classes on grounds worth declaring, since neither satisfies the strict definition: both share with the first three the generative mechanism, which is the asymmetry between the cost of requesting and the cost of serving, and the controls that mitigate them. We do not retain them because they produce unavailability, since the fifth does not. In elastic architectures, the

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

automatic scaling that preserves availability is precisely the mechanism that materializes the loss, so that defending against the third class may aggravate the fifth.

3.2 Ungoverned endpoints as attack-surface debt

The second contribution distinguishes two categories often treated as synonymous, which have distinct origins and call for distinct treatment.

Table 2

Conceptual distinction between shadow APIs and zombie endpoints

Criterion	Shadow API	Zombie endpoint
Origin	Published outside the formal governance process.	Reachability retained after deprecation or replacement.
Inventory	Never registered.	Registered as inactive, with no correspondence to the actual state.
Predominant risk	Absence of consumption limits, authentication, and monitoring.	Vulnerabilities already fixed in current versions.
Discovery	Runtime traffic analysis and inspection of deployment artifacts.	Reconciliation between declared inventory and observed reachability.

Note. Compiled by the authors based on OWASP (2023).

The two categories also differ in life cycle — the shadow API is active and evolving, with no designated owner; the zombie endpoint is static and unpatched. We designate the pair attack-surface debt, by analogy with technical debt: it accrues as a by-product of delivery speed and bears interest, since every ungoverned endpoint remains exposed to vulnerabilities already fixed elsewhere.

3.3 Assumptions that limit protection platforms

The third contribution results from analyzing the operating architecture of the platforms. Four assumptions were identified whose violation degrades efficacy regardless of implementation quality. The routing assumption: the platform inspects only the traffic that traverses its enforcement point, and endpoints reachable by paths that bypass the gateway fall outside its reach. The specification assumption: positive-model protection compares the request against the declared contract, and endpoints without a specification fall back on generic rules — a dependency acknowledged by the very literature that proposes such protection (Santos Filho et al., 2025).

The baseline assumption: behavioral detection depends on a history that characterizes normal use, and endpoints with low traffic, newly exposed, or intermittent do not accumulate one. And the uniform-cost assumption: limits expressed in requests per unit of time treat as equivalent operations whose cost differs by orders of magnitude. The first three are violated precisely by ungoverned endpoints; the fourth, by any endpoint whose cost has not been characterized. The design of the model follows from these statements.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

3.4 The proposed assessment model

The artifact consists of a model of five dimensions, each measured on four proposed ordinal levels. The first three address the routing, specification, and baseline assumptions; the fourth, the uniform-cost assumption; the fifth deals with response, whose timeliness is conditioned by the observed duration of attacks.

Table 3

Dimensions of the model for assessing API exposure

Dimension	Object of assessment	Evidence at the managed level
D1 — Inventory completeness	Identifying, continuously and at runtime, the reachable endpoints.	Automated discovery, with periodic reconciliation between observed and declared.
D2 — Life-cycle governance	Carrying the endpoint from registration to effective decommissioning, with a designated owner.	Deprecation policy with a deadline and verification of reachability after decommissioning.
D3 — Cost characterization	Knowing the computational cost of each operation and of its dependencies.	Endpoints classified by cost band, with limits differentiated by band.
D4 — Consumption control	Limits and quotas proportional to cost and to the requester's level of authentication.	Quotas by identity and by aggregate cost, with a ceiling for unauthenticated traffic.
D5 — Autonomous detection and response	Identifying anomalous automation and containing it without human intervention.	Automatic containment by policy, with an auditable record and rollback.

Note. Compiled by the authors.

The four levels are uniform across the dimensions and ordinal: progression indicates a direction of improvement, not a measured distance. At level 1, Initial, the dimension is not the object of a defined process. At level 2, Defined, there is a documented process and a designated owner, but execution is periodic and manual. At level 3, Managed, the process is continuous and automated. At level 4, Optimized, it is integrated into the delivery cycle, so that a policy violation prevents the change from being promoted to production.

The model adopts a composition rule that expresses the central proposition of this work: the overall level does not correspond to the mean of the dimensions but is capped by the level of D1. The justification is logical: consumption controls, cost characterization, and behavioral detection apply to known endpoints, so that if the organization is unaware of part of its surface, those controls operate over a subset of indeterminate size and their aggregate coverage is indeterminate to the same degree. An organization with maximum maturity in D3, D4, and D5 and initial maturity in D1 does not have strong protection with one localized gap: it has protection of unknown extent. The rule is proposed, not demonstrated.

3.5 Operational metrics

To make the model applicable, five metrics are proposed, computable from data that operations already produce. The fourth of them shifts the management question — how many

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

requests per second the system can withstand — to a more informative one: what is the most expensive operation an adversary without credentials can trigger.

Table 4

Proposed operational metrics

Metric	Operational definition	Interpretation
Inventory completeness index	Endpoints observed at runtime that appear in the inventory, over the total observed.	Fraction of the surface under governance; the limiting indicator of the model.
Unrouted endpoint ratio	Active endpoints reachable by a path that does not traverse the enforcement point.	Measures violation of the routing assumption.
Retirement latency	Median time between announced deprecation and verification of unreachability.	Estimates the accumulation rate of zombie endpoints.
Maximum cost per unauthenticated request	Highest cost observed among the operations triggerable without authentication.	Approximates the ceiling of the asymmetry ratio available to an anonymous adversary.
Automation policy coverage	Inventoried endpoints under a detection policy and consumption limit.	Distance between contracted and effective coverage.

Note. Compiled by the authors.

3.6 Analytical demonstration of the model

To demonstrate the applicability of the model, we apply it to an illustrative configuration, constructed for that purpose and not corresponding to any observed organization: the values are not empirical data but parameters chosen to exercise the instrument. The demonstration is, in the terms of subsection 2.4, an ex ante, artificial, and formative evaluation. The configuration posits a financial services organization that operates a contracted protection platform, declares three hundred endpoints in its gateway, and presents three hundred and eighty reachable under traffic observation; the consumption policy covers the entire declared inventory; there is no classification by cost band; and reconciliation is manual and annual.

Table 5

Analytical demonstration: illustrative configuration and reading of the model

Dimension	Level	Evidence in the illustrative configuration	Metric
D1 — Inventory completeness	1 — Initial	Manual, annual reconciliation; no runtime discovery.	Completeness = 300/380 = 0.79
D2 — Life-cycle governance	2 — Defined	Documented deprecation policy, manual verification.	Retirement latency = 180 days
D3 — Cost characterization	1 — Initial	Cost of operations not characterized.	Maximum unauthenticated cost = undetermined
D4 — Consumption control	3 — Managed	Automated quotas by identity, over the entire declared inventory.	Coverage = 1.00 of inventory; 0.79 of observed
D5 — Detection and response	3 — Managed	Automatic containment by policy, with an auditable record.	Unrouted endpoints = 0.21
Overall reading	1 — Initial	Composition rule: capped by the level of D1.	Arithmetic mean = 2.0

Note. Compiled by the authors. Illustrative configuration, constructed for purposes of demonstration; it does not correspond to any observed organization and the values are not empirical data.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

The demonstration yields a result that is not evident before the instrument is applied. The two readings diverge: the arithmetic mean places the organization at level 2 and suggests effort on D2; the composition rule places it at level 1 and directs effort to D1 and D3. The divergence is not one of degree but of the object of investment. Its origin lies in the coverage metric: the consumption policy reaches the entire declared inventory and reports full coverage under the indicator the organization tracks, while reaching 79% of the observed surface. The remaining 21% are under no limit at all, and the absence of cost characterization makes it impossible to estimate the ceiling of the asymmetry ratio available to an adversary without credentials. The demonstration establishes that the model is applicable and that its composition rule is operative; it does not establish that the recommendation is correct in a real environment.

3.7 Normative mapping and discussion

Table 6

Mapping of the model's dimensions against normative frameworks

Dimension	OWASP API	ISO/IEC 27001:2022	NIST CSF 2.0
D1 — Inventory completeness	API9:2023	5.9 Inventory of assets	Identify
D2 — Life-cycle governance	API9:2023; API8:2023	8.9 Configuration management; 8.32 Change management	Govern; Protect
D3 — Cost characterization	API4:2023	8.6 Capacity management	Identify; Protect
D4 — Consumption control	API4:2023; API6:2023	8.26 Application security requirements	Protect
D5 — Detection and response	API4:2023	8.16 Monitoring activities; 5.7 Threat intelligence	Detect; Respond

Note. Compiled by the authors based on OWASP (2023), International Organization for Standardization (2022), and National Institute of Standards and Technology (2024). The control identifiers should be checked against the full text of the standard before use in an audit.

The model engages critically with the prevailing orientation in the field, which concentrates effort on detection, particularly through machine learning (Praseed & Thilagam, 2019; Tripathi & Hubballi, 2021; Deressu et al., 2026). That body of work presupposes as solved the problem this study identifies as prior: for a detector to observe anomalous traffic on an endpoint, the traffic of that endpoint must be observed. Efficacy demonstrated in an experimental setting, in which the set of endpoints is known by construction, does not transfer to a production environment in which that premise does not hold. One tension remains: D1 demands observability precisely where it is lacking.

4 - FINAL CONSIDERATIONS

This work proposed a model for assessing API exposure to application-layer denial of service, built through the Design Science Research strategy and intended for organizations whose attack surface is not fully inventoried.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

We built a taxonomy of five classes organized by the resource exhausted, showing that only one corresponds to the prevailing volumetric conception and that two sit at the boundary of the concept of denial of service. We distinguished shadow APIs from zombie endpoints, proposing the designation attack-surface debt for the pair. We identified four efficacy assumptions of the platforms, of which the first three are violated by ungoverned endpoints.

The theoretical contribution consists in formulating the asymmetry ratio as the central explanatory variable and in the proposition that inventory completeness is a determining factor for the effective coverage and governance of all other controls. The scope of that proposition bears stating precisely: the study holds that the inventory delimits the extent over which the other controls operate and that, with that extent unknown, aggregate coverage is indeterminate; it does not hold that completeness is a universal condition of efficacy. Given the absence of naturalistic evaluation, the model should be read as a proposal subject to empirical examination.

As a future agenda, we suggest the naturalistic evaluation of the model following the protocol set out in subsection 2.4; the production of independent evidence on the prevalence of ungoverned endpoints; experimental investigation of the asymmetry ratio in REST and declarative interfaces; and analysis of the regulatory implications of the unavailability of API-mediated services in regulated sectors in Brazil.

REFERENCES

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/

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

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

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. https://doi.org/10.32604/cmc.2025.067536

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

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

REVISTA CIENTÍFICASOL21
STANDARD OPEN LITERATURE - ISSN 3086-089X

APPENDIX A — ABBREVIATIONS AND TECHNICAL TERMS

API, application programming interface; CSF, Cybersecurity Framework; DDoS, distributed denial of service; DSR, Design Science Research; EDoS, economic denial of sustainability; GraphQL, query language for APIs; gRPC, open-source remote procedure call framework; HTTP, Hypertext Transfer Protocol; ISO/IEC, International Organization for Standardization and 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.

This article is published on an open-access basis under the **Creative Commons Attribution 4.0 International License (CC BY)**, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.