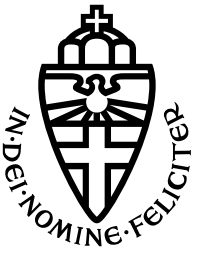
RADBOUD UNIVERSITY



and

Explainable Ensemble-Based Anomaly Detection for DNS Registry Mutations in Hybrid Registry Infrastructures

June 30, 2026

Author:
Ina Dan (s1151804)

Thesis M.Sc. Computing Science - Cybersecurity and AI

Supervisors:
Dr. Stjepan Picek (Radboud University)
Dr. Elmer Lastdrager (SIDN Labs)
Dr. Ralph Koning (SIDN Labs)

Second reader:
Dr. Gunes Acar (Radboud University)

Abstract

Modern Top-Level Domain (TLD) registries increasingly adopt hybrid and cloud-integrated architectures in which registrar-facing operational systems are partially deployed in public cloud environments while authoritative DNS infrastructure and registry data remain within sovereign or highly controlled environments. Although this architectural transition improves scalability, availability and operational flexibility, it also introduces new security challenges at the replication boundary between cloud-hosted systems and authoritative registry infrastructure.

This thesis investigates how explainable ensemble-based anomaly detection techniques can be used to detect malicious or suspicious DNS registry mutations before they are incorporated into authoritative DNS zones. The proposed framework operates as an additional behavioral security layer alongside authoritative replication pipelines, monitoring mutation activity without interrupting synchronization between operational systems and the authoritative registry database.

The framework combines heterogeneous Graph Neural Network (GNN) representation learning with feature-based anomaly detection using Isolation Forest. Registry mutations are modeled as relational infrastructure behavior involving domains, nameservers, registrars, IP subnets and DNSSEC-related entities, while additional statistical and temporal features capture operational mutation behavior. To support operational usability, the framework additionally incorporates post-hoc explainability mechanisms combining graph-based relational interpretation with feature-level anomaly explanations.

Because publicly available labeled datasets for malicious authoritative registry mutations are effectively unavailable, evaluation is conducted using synthetic schema-valid attack scenarios derived from realistic DNS abuse patterns, including delegation manipulation, registrar compromise, coordinated infrastructure reuse, DNSSEC-related attacks and long-term mutation abuse.

Experimental results demonstrate that the proposed ensemble framework is highly effective for detecting structurally significant attacks involving coordinated infrastructure relationships and anomalous trust configurations. Most implemented attack scenarios achieved strong anomaly separation, high ROC-AUC values and stable ranking behavior across multiple random seeds. However, the evaluation also revealed an important limitation: subtle glue-record manipulation attacks that preserve graph topology and operational plausibility remained substantially more difficult to distinguish from benign infrastructure behavior.

The findings demonstrate both the strengths and limitations of graph-based anomaly detection within authoritative DNS registry environments. Overall, the research shows that explainable ensemble-based anomaly detection constitutes a promising approach for strengthening behavioral monitoring and security visibility within modern hybrid DNS registry infrastructures.

Contents

1 Introduction

The Domain Name System (DNS) is one of the fundamental components of the Internet infrastructure by mapping domain names, which are easily readable by humans, to network resources [11]. At the core of this ecosystem lie Top-Level Domain (TLD) registries [33], i.e., organizations responsible for maintaining the authoritative data and operational infrastructure of a TLD. These registries maintain authoritative data containing domain ownership information, registrar associations (i.e., accredited entities that provide domain registration and management services to registrants and interact directly with the registry on their behalf), delegation records and DNS Security Extensions (DNSSEC) trust anchors. The correctness and integrity of registry data are essential to the global stability and security of the Internet [74]. Unauthorized modifications to registry data may result in domain hijacking, traffic redirection, service disruption, large-scale phishing campaigns or the undermining of cryptographic trust chains [79].

Among all registry data fields, delegation information represents one of the most security-critical subsets [68]. Name server (NS) records define which infrastructure is authoritative for resolving a domain. Glue records, i.e., DNS records that provide the IP addresses of name servers when those servers are located within the domain they serve, thereby preventing circular resolution dependencies, specify the corresponding network endpoints. Delegation signer (DS) records establish the DNSSEC chain of trust. Unauthorized modification of any of these elements can significantly affect how DNS resolution is performed. DNSSEC provides an important security mechanism by enabling validating resolvers to detect forged, modified or otherwise unauthorized DNS data [7]. As a result, many attacks involving DNS response manipulation or inconsistent delegation information can be prevented through proper DNSSEC deployment. However, DNSSEC primarily protects the integrity and authenticity of published DNS data rather than the legitimacy of the registry mutations that produced that data.

If an attacker succeeds in introducing formally valid delegation, glue or DNSSEC-related changes through legitimate registry workflows, those changes may ultimately be incorporated into a correctly signed zone and distributed as authentic DNS data. In such cases, DNSSEC continues to function as intended, but it cryptographically protects a state that may itself have been produced through malicious or unauthorized mutations. Consequently, additional monitoring and validation mechanisms capable of assessing the legitimacy of registry mutations before publication can provide a valuable complementary security layer. Prior empirical work has shown that inconsistencies and unexpected behavior in delegation relationships can have wide effects, illustrating how sensitive this part of the DNS control plane is to integrity failures [67].

In recent years, organizations have increasingly adopted hybrid cloud architectures [42] to accommodate growing scalability demands, automation requirements and operational flexibility, while maintaining compliance with regulatory and sovereignty requirements. In such architectures, operational services, application programming interfaces (APIs) and management platforms are increasingly deployed across external and cloud-hosted environments, while authoritative infrastructure remains within highly trusted environments. This separation allows organizations to benefit from cloud elasticity without relinquishing control over critical assets.

This architectural transition is also reflected in ongoing modernization efforts within TLD registry environments. For example, SIDN, the registry operator of the `.nl` top-level domain, is undertaking a gradual transition toward a hybrid cloud-based operational model [64]. Although this separation improves scalability, resilience and operational flexibility, it also introduces new trust boundaries between cloud-hosted operational systems and the authoritative registry core.

While this modernization enables improved service availability and operational flexibility, it also intro-

duces new security challenges that were largely absent in earlier monolithic deployments [9]. In particular, the introduction of automated replication pipelines [47], [49] between cloud-hosted operational systems and the authoritative registry database fundamentally alters the threat landscape, creating new attack vectors in which cloud-originated mutations, i.e., state-changing operations such as insertions, updates and deletions applied to registry data, may directly affect the integrity of the registry core, including delegation mappings, glue endpoints and DNSSEC trust configuration that ultimately feeds zone generation and signing processes. Ensuring that this architectural transition does not weaken the integrity of authoritative registry data, the availability of DNS services or the protection of delegation and DNSSEC trust relationships therefore constitutes a critical engineering and research challenge.

To ensure consistency between operational systems and the authoritative database, mutations originating in the cloud environment are automatically propagated to the registry database using replication mechanisms or change data capture pipelines [46]. These mechanisms typically replicate low-level database operations without evaluating the semantic intent or security implications of the applied changes. Consequently, syntactically valid yet malicious mutations may propagate from a less trusted environment into the authoritative registry core. Although multiple preventative and detective controls already exist, including identity and access management, network segmentation, encryption, auditing and infrastructure hardening, these controls do not provide semantic validation at the replication boundary.

Public cloud environments may additionally introduce broader and more dynamic attack surfaces due to increased API exposure, configuration complexity and shared responsibility models [20], [21]. Credential compromise, vulnerable APIs, misconfigurations or insider threats may therefore allow adversaries to manipulate registry data within cloud-hosted systems [23], [52]. Once accepted at the operational layer, such mutations may be replicated automatically toward the authoritative registry database and subsequently incorporated into generated and signed zone files. In this context, the boundary between cloud-hosted systems and the sovereign authoritative infrastructure becomes a critical security boundary within hybrid registry architectures.

At the same time, replication pipelines must preserve continuous synchronization [47] between operational and authoritative environments to maintain registry availability and correctness. A mechanism that blocks or delays mutations risks introducing failures that are unacceptable for critical infrastructure operators. Consequently, the practical goal of this thesis is detection and risk scoring of suspicious mutations rather than hard prevention through mutation rejection.

Existing security controls in operational DNS environments primarily rely on authentication, DNSSEC-based integrity protection, syntactic validation and static policy enforcement [60]. While these controls significantly reduce the likelihood of malformed or unauthorized transactions, they remain insufficient against threats operating within valid protocol and schema constraints. Compromised credentials, abused automation pipelines or manipulated operational components may generate delegation or glue changes that are fully compliant yet malicious in intent. Rule-based validation also struggles to identify subtle low-and-slow attacks, coordinated mutation campaigns or deviations from historical operational behavior.

This thesis proposes an explainable ensemble-based anomaly detection framework that operates as a pre-zone behavioral risk filter at the authoritative replication boundary, analyzing registry mutations before they become incorporated into the signed authoritative DNS zone. The proposed framework combines heterogeneous graph representation learning with feature-based anomaly scoring to detect suspicious mutations, while a threat-model-driven synthetic attack methodology is introduced to evaluate realistic registry abuse scenarios in the absence of labeled malicious datasets.

Machine learning techniques have demonstrated considerable success in related cybersecurity domains, including intrusion detection systems [66], cloud infrastructure monitoring [70] and anomaly detection

in network traffic [50]. Supervised learning approaches can identify known attack patterns when labeled data is available [35], while unsupervised methods can model normal operational behavior and highlight deviations indicative of novel or stealthy threats [51]. However, much of the existing ML-based security anomaly detection literature is developed for network traffic telemetry [22] or for system log streams [12], rather than structured DNS registry mutation data. Moreover, labeled datasets of real registry mutation attacks are typically unavailable due to confidentiality constraints and the rarity of publicly documented incidents. This lack of labeled and publicly available datasets for registry mutation attacks constitutes an open research problem.

The structured and domain-specific nature of registry mutation data, combined with the lack of labeled datasets, motivates an unsupervised anomaly detection approach as the primary machine learning component. The working assumption is that historical mutation data reflects benign operational behavior. Suspicious scenarios can then be introduced in a controlled manner for evaluation by constructing threat-model-driven synthetic mutations. This approach allows performance to be evaluated without requiring access to confidential incident labels, while preserving realistic operational distributions.

The proposed framework is intended to operate within the trusted environment hosting the registry database as a detection layer evaluating incoming mutations before authoritative zone generation. While it primarily focuses on delegation and glue-related mutations due to their direct impact on DNS resolution, it also incorporates DNSSEC-related and broader mutation behavior. Because the framework targets critical infrastructure environments, explainability is essential for supporting incident response, registrar communication and operational auditing.

Due to the absence of publicly available labeled registry attack datasets, this research adopts a threat-model-driven synthetic scenario methodology for validation. Realistic mutation events are systematically modified to simulate delegation hijacking, glue manipulation, coordinated infrastructure abuse and stealthy takeover scenarios while preserving operational plausibility and statistical realism. In addition, the framework incorporates explainability mechanisms that highlight dominant anomaly-contributing features and deviations from historical baselines.

By addressing the largely unexplored problem of suspicious mutation propagation in DNS registry infrastructures, this research contributes both to the academic understanding of domain-specific anomaly detection and to the practical strengthening of Internet critical infrastructure. The proposed framework aims to reduce the risk that high-impact delegation or glue manipulations originating from less trusted operational environments remain unnoticed until after global publication, thereby strengthening confidence in modern hybrid registry deployments.

1.1 Motivation

DNS registries form part of the Internet's critical infrastructure and are entrusted with maintaining the correctness of authoritative domain data for millions of users and services. Even short-lived corruption of registry data may propagate globally through DNS caching and DNSSEC signing, resulting in large-scale service disruption, financial losses, reputational damage and undermining of trust in digital systems [62]. High-profile incidents involving domain hijacking and registrar compromise have demonstrated that attacks targeting registry data can be more impactful than traditional network-level intrusions [3], [30].

The increasing adoption of hybrid, cloud-integrated and externally managed operational environments has further amplified these concerns [42]. While such architectures provide significant benefits in terms of scalability, resilience and operational flexibility, they also introduce additional potential attack surfaces. Automated replication pipelines remain essential for maintaining consistency between operational and authoritative environments, but they are designed to ensure synchronization rather than evaluate the

legitimacy of propagated changes.

This architectural transition is also reflected within DNS registry environments, where operational platforms are increasingly integrated with cloud-hosted and externally managed services while authoritative infrastructure remains within sovereign or highly controlled environments [64]. Although this separation improves scalability, resilience and agility, it also introduces new trust boundaries between operational systems and the authoritative registry core.

1.2 Problem Statement

In hybrid DNS registry architectures, domain mutations originating from cloud-hosted platforms are automatically propagated to the authoritative registry database through replication mechanisms. These mechanisms preserve database consistency but remain agnostic to the operational legitimacy or security implications of applied changes. Although various preventive and corrective security controls exist upstream, formally valid mutations might still be replicated without contextual evaluation. Consequently, formally compliant yet suspicious delegation or glue modifications may propagate into the authoritative state if not detected through additional semantic or behavioral validation mechanisms. Once committed, such changes become indistinguishable from legitimate registry updates and are subsequently incorporated into signed zone files and distributed everywhere.

The core issue addressed in this thesis is therefore:

> How can explainable ensemble-based anomaly detection techniques be used to detect malicious or suspicious DNS registry mutations at the authoritative replication boundary in hybrid registry infrastructures, with particular emphasis on delegation and glue-related mutations?

To address this research question, the following sub-questions are considered:

- What attack scenarios and malicious mutation patterns are relevant within hybrid DNS registry infrastructures?
- Which structural, behavioral and DNS-specific features most effectively characterize suspicious registry mutations?
- How effectively can an ensemble-based anomaly detection framework distinguish suspicious mutations from benign behavior?
- How robust is the proposed framework under highly imbalanced and operationally realistic registry conditions?
- How can anomaly detection outputs be made interpretable and actionable for registry operators?

1.3 Contributions of This Research

This research contributes to the field of DNS security and anomaly detection by addressing the emerging security challenges introduced by hybrid and cloud-integrated DNS registry architectures. As organizations increasingly adopt hybrid, cloud-integrated and externally managed operational environments, ensuring that this architectural evolution does not introduce new security risks becomes an important operational and research challenge [9], [42]. This trend is also evident within DNS registry environments, where operational systems are increasingly distributed across external environments while retaining authoritative DNS functions within sovereign or highly controlled environments [64].

Within this context, this thesis proposes an explainable anomaly detection framework designed specifically for authoritative DNS registry mutation monitoring. The framework combines heterogeneous graph representation learning with feature-based anomaly detection to identify suspicious mutation behavior at the authoritative replication boundary before mutations are incorporated into signed DNS zones.

An additional contribution of this research is the development of a threat-model-driven methodology for authoritative DNS registry environments. Through literature review, expert discussions with SIDN and NLnet Labs, and analysis of hybrid registry architectures, twelve realistic attack scenarios were identified and modeled. These scenarios cover delegation abuse, glue manipulation, registrar compromise, DNSSEC misuse, coordinated multi-domain attacks, infrastructure reuse, contact hijacking and long-term low-and-slow manipulation. Beyond supporting the evaluation itself, these threat models contribute to the broader discussion of how new attack classes may emerge as DNS registries increasingly adopt hybrid and cloud-integrated operational workflows.

Because publicly available labeled datasets containing malicious registry mutations are effectively unavailable, this research additionally develops a synthetic attack injection methodology based on realistic and schema-valid mutation behavior. This enables controlled and reproducible evaluation under multiple operationally relevant attack scenarios while preserving temporal consistency and operational plausibility.

The main contributions of this thesis can therefore be summarized as follows:

- Development of an explainable ensemble-based anomaly detection framework for authoritative DNS registry mutation monitoring (Section 6).
- Application of heterogeneous graph representation learning to model complex relationships between domains, nameservers, registrars, IP subnets and DNSSEC-related entities (Section 6.5).
- Design of an ensemble architecture combining Graph Neural Networks and Isolation Forest to jointly capture structural and feature-level anomalies (Section 6.7).
- Development of a DNS registry threat model comprising 12 attack scenarios relevant to hybrid and externally hosted registry infrastructures (Section 4).
- Creation of a threat-model-driven synthetic attack injection methodology for evaluating anomaly detection performance in the absence of labeled attack datasets (Section 6.9).
- Experimental evaluation of the strengths and limitations of graph-based anomaly detection for DNS registry security, including the challenges associated with subtle infrastructure-level attacks such as glue record manipulation (Section 7).
- Introduction of a pre-zone behavioral risk filter concept that provides an additional monitoring layer before suspicious mutations become incorporated into authoritative DNS zones (Section 8).

2 Background

DNS constitutes one of the most indispensable components of the worldwide Internet infrastructure [62]. Although frequently perceived as a simple directory service that translates human-readable domain names into IP addresses, DNS in reality represents a globally distributed coordination and trust infrastructure. [13] It enables stable digital identity, service discoverability and communication routing at planetary scale. Virtually every Internet transaction, including web browsing, email delivery, software updates, certificate validation processes and cloud API interactions, depends directly or indirectly on the correctness, availability and integrity of DNS resolution.

DNS was designed as a hierarchical and decentralized system in order to provide scalability, delegation of administrative authority and resilience against localized failures. Its architecture distributes responsibility across multiple organizational actors while preserving a coherent and globally consistent namespace. At an upper level of this hierarchy reside Top-Level Domain (TLD) registries, positioned directly below the DNS root, which maintain authoritative control over defined portions of the namespace. These registries serve not merely as service operators, but as custodians of critical digital infrastructure whose correctness directly affects global Internet stability.

Because DNS forms the foundation of nearly all Internet-based services, its correctness is not solely a matter of reliability but of systemic stability. Errors, inconsistencies or malicious manipulations at upper levels of the hierarchy may propagate widely and rapidly. Consequently, TLD registries occupy a uniquely sensitive position within the Internet ecosystem: they are simultaneously operational service providers, contractual authorities and stewards of globally trusted state.

2.1 Authoritative Registry Databases and Global Trust

Each TLD registry maintains an authoritative database representing the ground truth for its managed namespace [27]. Registrars act as intermediaries between registrants and the registry by submitting domain registrations and updates. Consequently, compromise at the registrar level may enable attackers to introduce formally valid yet malicious mutations into the authoritative registry state.

The registry database defines domain ownership, registrar relationships, authoritative name server delegations, glue records, DNSSEC trust configuration and policy states. Together, these elements determine the resolution behavior and cryptographic trust relationships of millions of domains.

Zone generation processes transform this authoritative state into DNS zone files that are distributed to authoritative name servers worldwide. When DNSSEC is enabled, zone data is cryptographically signed before distribution [8]. Through recursive resolution and caching, these signed records rapidly propagate across the global DNS ecosystem.

Consequently, any mutation committed to the authoritative registry database may become globally visible within minutes. Even short-lived inconsistencies or malicious changes may result in traffic redirection, service disruption or loss of trust, making registry integrity a critical component of Internet infrastructure security.

2.2 Domain Mutations as High-Impact Control Operations

Registry databases evolve continuously through domain mutations. These mutations constitute discrete control operations that alter the authoritative state of the namespace. They include both routine lifecycle events and infrastructure-defining configuration changes.

Typical mutation categories include:

- Domain lifecycle events (creation, renewal, expiration, deletion),
- Registrar transfers and administrative updates,
- Contact information modifications,
- Delegation updates to authoritative name servers,
- Glue record additions or removals,
- DNSSEC key rollovers and DS record updates,
- Status flag transitions enforcing policy or dispute outcomes.

While many registry mutations are operationally benign, delegation and glue-related changes are particularly security-sensitive [67] because they define the authoritative infrastructure trusted to resolve a domain. In hybrid cloud environments, adversaries may gain the ability to introduce formally valid yet malicious delegation or glue updates through compromised registrar accounts, abused provisioning APIs, misconfigured cloud services, vulnerable automation pipelines or other weaknesses. Such mutations may redirect traffic toward attacker-controlled infrastructure, enabling hijacking, phishing or service disruption [30], [77]. Once incorporated into signed zone output, these changes become globally distributed and cryptographically authenticated despite potentially malicious intent, highlighting the distinction between structural validity and operational legitimacy.

2.3 Architectural Evolution Toward Hybrid Registry Infrastructures

Historically, TLD registries operated primarily within centralized and tightly controlled sovereign infrastructures [27], where registrar interfaces, operational systems and authoritative databases were co-located and protected through perimeter defenses, network segmentation and strict physical control.

However, increasing transaction throughput requirements, expectations of near real-time responsiveness, demands for geographic redundancy and high availability, economic incentives for elastic infrastructure, DevOps adoption, expanding service integration and perceived cloud security benefits have driven the adoption of distributed and hybrid cloud architectures [5], [63].

In modern hybrid architectures, registrar-facing services and platforms are frequently deployed in public cloud environments, while authoritative registry databases and DNSSEC signing infrastructure remain within sovereign or tightly controlled environments due to security and regulatory requirements.

Although this separation preserves formal control over critical infrastructure, it also introduces new trust boundaries between cloud-hosted systems and authoritative registry infrastructure. Consequently, the attack surface extends beyond sovereign environments to include cloud-hosted APIs, identity management systems and automation pipelines, requiring registry security models to manage trust transitions across architectural boundaries rather than relying solely on perimeter protection.

2.4 Mutation Propagation and Replication Pipelines

To ensure consistency between operational platforms and the authoritative registry database, registries employ replication pipelines that propagate mutations from cloud-hosted environments toward sovereign infrastructure [46]. These mechanisms are designed to provide low-latency synchronization, transactional consistency and fault tolerance while minimizing transformation of replicated data.

In practice, this design enables near real-time synchronization of registry state while decoupling frontend systems from authoritative infrastructure. High availability and low latency are essential, as registry services must process large volumes of registrar transactions without introducing degradation or inconsistency.

From a security perspective, however, replication architectures introduce structural limitations. Replication mechanisms validate syntactic correctness, schema compliance and transactional integrity, but they do not assess semantic legitimacy or contextual risk. If a mutation satisfies formal validation constraints at the cloud-hosted layer, it will typically be replicated faithfully into the authoritative registry database.

Because continuous synchronization is a core requirement, purely preventive mechanisms that block or delay replication may be infeasible or risky. Interrupting replication can violate consistency guarantees between cloud and sovereign components, potentially destabilizing critical infrastructure. This constraint motivates detection-centric approaches that preserve synchronization continuity while introducing semantic scrutiny at the replication boundary.

2.5 Threat Landscape and Security Implications

Public cloud environments introduce a broader and more dynamic threat surface for registry infrastructures [23]. Potential attack vectors include compromised registrar credentials, abuse of provisioning APIs, misconfigured cloud services, supply-chain compromise and insider threats. Rather than compromising the authoritative database directly, attackers may inject malicious yet formally valid mutations into cloud-hosted systems, from where they propagate automatically through replication pipelines.

If malicious delegation, glue or DNSSEC-related updates are committed to the authoritative database, they become incorporated into zone generation and cryptographically signed using DNSSEC, transforming a localized compromise into a globally distributed and authenticated false state. Recursive resolvers and end users subsequently accept the signed output as authoritative, amplifying the impact of a single successful mutation.

DNSSEC configuration changes may also independently introduce security risks. Incorrect or unexpected modifications, such as DS mismatches or anomalous key changes, can weaken trust guarantees or render domains inaccessible to validating resolvers due to DNS caching effects [6], [45]. Although such changes may remain formally valid under protocol and schema rules, they can still negatively affect the cryptographic integrity of the namespace. Consequently, DNSSEC-related updates should be treated as security-relevant components of the authoritative control plane [8].

Many high-impact scenarios therefore exploit the gap between structural validity and contextual legitimacy rather than violating explicit authorization rules. Sophisticated adversaries can operate entirely within valid registry workflows.

2.6 Limitations of Existing Validation Mechanisms

Current DNS security mechanisms primarily rely on preventive controls, including authentication mechanisms, configuration validation, encryption and cryptographic integrity protection through DNSSEC [7], [60]. These mechanisms significantly reduce the likelihood of unauthorized access, forged DNS data and structurally invalid changes, but they are not designed to evaluate whether formally valid mutations are operationally expected or behaviorally consistent over time.

As a result, existing validation approaches remain limited in their ability to detect subtle or coordinated malicious behavior that operates within legitimate workflows, such as anomalous mutation patterns, coordinated infrastructure changes or statistically unusual operational behavior across domains and registrars.

Existing machine learning-based security approaches frequently prioritize predictive performance with-

out sufficient emphasis on explainability and operational integration. In critical infrastructure environments, opaque anomaly scores are insufficient because operators must understand why mutations are considered suspicious in order to support investigation, incident response and operational decision-making.

2.7 Motivation for Intelligent Mutation Validation

The replication boundary between cloud-hosted systems and the authoritative registry database represents a strategically critical observation and validation point. At this boundary, mutations have passed formal validation but have not yet been incorporated into globally distributed signed zone outputs.

A protection mechanism deployed at this boundary must satisfy demanding requirements:

- Near real-time processing,
- Deterministic and explainable outputs,
- Compatibility with strict consistency guarantees,
- Robustness against evolving attack strategies,
- Minimal operational disruption from false positives.

Machine learning techniques, when combined with domain-specific feature engineering and deterministic policy enforcement, offer the ability to model complex behavioral patterns across registrars, domains and mutation types. Such an approach can bridge the gap between structural correctness and contextual legitimacy.

This thesis builds upon the architectural and threat landscape described above to design and evaluate a machine-learning-assisted mutation detection framework operating at the authoritative boundary of hybrid DNS registry infrastructures. The objective is to reduce the risk that high-impact delegation or glue manipulations originating from less trusted environments become authoritative and globally propagated, while preserving the reliability, consistency and performance requirements of production registry systems.

3 Related Work

This section reviews research relevant to the design of an anomaly detection layer for registry mutations within hybrid DNS registry infrastructures. The discussion focuses on four main areas: anomaly detection in cloud and database environments, sequential and relational anomaly detection for transactional data, robustness and evaluation under limited labeling conditions, and DNS and domain infrastructure security.

Across these themes, a recurring gap is that most proposed methods are validated on logs, network telemetry or access traces, whereas this thesis operates on structured, temporally versioned registry tables that represent authoritative state transitions rather than raw system events.

3.1 Anomaly Detection in Cloud and Database Systems

Machine learning for anomaly detection in cloud environments and databases has been studied extensively, particularly in the context of insider threats, misconfigurations, unauthorized access, and abnormal workload patterns. Surveys and systematic reviews describe the landscape of supervised, semi-supervised and unsupervised methods and their challenges in hybrid and multi-cloud deployments [26], [32], [59]. A consistent conclusion is that unsupervised detection is often favored when labeled attack traces are scarce or sensitive, but evaluation must be carefully designed to avoid overly optimistic results.

Isolation Forest [40] and reconstruction-based approaches such as autoencoders [61] are widely used due to their ability to learn normal behavior from unlabeled data. Recent works apply these approaches to database monitoring and cloud security, often using access logs, query traces, or user activity data as input [1], [44], [65]. However, these inputs differ fundamentally from registry mutation data: query logs describe how data is accessed, whereas registry mutation streams describe how authoritative state is changed.

Hybrid detectors and ensembles have been proposed to improve stability and reduce false positives in dynamic cloud settings by combining multiple model families (e.g., Isolation Forest + deep models) and by aggregating signals across time [37]. Likewise, generative approaches and richer feature fusion methods have been explored to better model complex cloud behaviors [15], [55]. Although these techniques are relevant conceptually, their direct application to registry mutation validation requires adaptation because registry data is highly structured, semantics-heavy, and subject to strict correctness constraints and latency requirements.

3.2 Sequential and Transaction-Stream Anomaly Detection

Registry mutations are naturally ordered and often correlated across short time spans, making the problem related to anomaly detection in logs and transaction streams. Early log anomaly detection approaches such as DeepLog use LSTMs to learn normal event sequences and detect deviations [18]. Transformer-based models have been proposed to capture longer-range dependencies and richer semantics in distributed logs, such as LogFormer [24]. Systematic studies emphasize that sequence length, representation choices, and workload characteristics strongly affect performance and generalization [25].

Despite their strong detection capabilities, sequence models introduce trade-offs. They can be computationally expensive, sensitive to drift and comparatively less transparent than classical detectors. For registry operations, explainability and predictable failure modes are important, because false alarms can disrupt registrar operations and because incident response requires evidence-based explanations. As a result, sequence modeling serves primarily as a reference point in this thesis, while the proposed approach emphasizes interpretable and computationally efficient features that still capture important temporal behavior,

including mutation burstiness, time since previous changes and correlations between related mutation events.

3.3 Relational and Graph-Based Security Analytics

A DNS registry is intrinsically relational: domains are linked to registrars and resellers, domains are linked to authoritative nameservers, and nameservers are linked to IP addresses (glue). These relationships naturally form a graph structure in which entities correspond to nodes and their interactions define edges.

This structural representation makes the problem well-suited for graph-based security analytics and Graph Neural Network (GNN) approaches, which have increasingly been applied to anomaly detection and cyber threat modeling [36], [73]. Unlike traditional machine learning methods that operate primarily on independent feature vectors, graph-based models explicitly incorporate dependencies between entities and their surrounding relational context.

Prior research has shown that many cyber attacks emerge through coordinated interactions, shared infrastructure or anomalous connectivity patterns rather than isolated feature deviations alone [19], [36], [73]. In DNS ecosystems, this may include sudden delegation changes toward previously unseen nameserver infrastructure, coordinated multi-domain mutation campaigns, unusual infrastructure reuse across unrelated domains or anomalous DNSSEC trust relationships [30], [77], [79].

GNNs are particularly attractive in this setting because they can propagate information across neighboring entities and learn representations that jointly capture structural and contextual behavior [56], [81]. This enables detection of anomalies that may remain difficult to identify using purely feature-based approaches.

Recent work on graph anomaly detection further suggests that graph-based models are especially effective when anomalies manifest through irregular relational structure, evolving connectivity patterns or coordinated multi-entity behavior [17], [19], [80]. These characteristics closely align with several DNS abuse and infrastructure takeover scenarios discussed throughout this thesis.

At the same time, existing literature also highlights important practical considerations for deployment, including scalability, temporal graph dynamics and explainability of learned anomaly signals [38]. Recent research therefore increasingly emphasizes the importance of combining relational graph structure with richer contextual and feature-level information in order to improve anomaly detection across complex and evolving cyber environments.

Overall, prior research strongly suggests that graph-based security analytics provide a promising foundation for detecting coordinated DNS infrastructure abuse and relational anomalies within complex registry environments.

3.4 Robustness, Concept Drift and Evaluation Without Labels

One of the main challenges in security anomaly detection is that real attacks are rare, labels are incomplete and normal behavior evolves. Recent work highlights common pitfalls in evaluation (e.g., contaminated training data, unrealistic synthetic anomalies, and distribution shifts) and stresses the need for careful experimental design [57]. In hybrid cloud deployments, concept drift is especially relevant because registrar automation, policy changes and infrastructure migrations can substantially alter baseline behavior over time [41].

The absence of labeled malicious registry mutations is a defining constraint in this thesis. While supervised learning can be used when realistic attack traces exist, unsupervised or semi-supervised anomaly detection is often the most practical baseline in operational settings. The evaluation then hinges on constructing realistic suspicious scenarios and measuring detection under controlled conditions, without over-claiming

generalization to all adversaries. This motivates the synthetic scenario methodology in this thesis: suspicious mutations are injected by transforming real benign events in ways that remain operationally feasible and consistent with the registry schema, thereby preserving important statistical properties while enabling repeatable experiments.

3.5 Explainability and Operational Use in Critical Infrastructure

Explainability is increasingly regarded as a practical requirement for ML-based security controls, particularly in critical infrastructure environments where decisions must be verifiable, defensible and actionable. Even when predictive accuracy is high, opaque or black-box decisions are difficult to integrate into production workflows because security analysts must understand which feature deviations triggered an alert and how strong the anomaly signal is so they can decide what response is appropriate. As a result, many deployed security systems favor inherently interpretable models (e.g., tree-based detectors, rule-augmented anomaly models) or complement more complex models with post-hoc explanation techniques.

In this thesis, explainability is treated as a first-class design requirement: the detection layer produces not only an anomaly score but also a structured explanation grounded in delegation and glue semantics, enabling operators to assess risk and take informed action. This aligns with the feature design and modeling choices in this thesis, where explanations are derived from interpretable structural and behavioral features of registry mutations.

3.6 DNS and Domain Infrastructure Security

Research on DNS security largely focuses on traffic-based attacks (e.g., DDoS, cache poisoning, malicious domain classification) rather than authoritative registry mutation integrity. For example, ML has been applied to protect authoritative name servers against DDoS using traffic features [69] and to classify malicious domains using curated DNS datasets [43]. More closely related to registry operations, recent work explores anomaly detection in domain ecosystems using behavioral signals at the registrant level [14]. Nonetheless, comparatively little work directly addresses the integrity of authoritative registry mutations, particularly in hybrid architectures where cloud-hosted systems propagate changes into a sovereign authoritative database.

3.7 Positioning of This Work

Existing literature provides strong foundations in anomaly detection, log-based sequence modeling and cloud security monitoring, but does not directly address the combination of structured registry mutation data, delegation and glue integrity risks, and hybrid cloud trust boundaries where replication pipelines prioritize consistency over semantic legitimacy.

This thesis addresses that intersection by proposing an anomaly detection and explanation layer operating on registry mutations and aligned with the risks identified in Section 4.1.1. While delegation and glue integrity constitute the primary detection focus due to their direct impact on DNS resolution and infrastructure control, the framework additionally considers DNSSEC manipulation, coordinated infrastructure reuse, registrar abuse patterns and multi-domain mutation activity.

4 Threat Models and Detection Scope

This section defines the threat models considered in this thesis and clarifies which threats can be studied using the available registry mutation data. The scope is guided by the architectural trust boundary in a hybrid DNS registry deployment: registrar-facing and operational components may run in a public cloud, while the authoritative registry database and zone generation pipeline remain in a more trusted environment. In this setting, the primary security concern is not only unauthorized access, but the possibility that formally valid changes originating from a less trusted environment become authoritative through synchronization and replication.

A key conceptual distinction is therefore between structural validity and contextual legitimacy. Structural validity refers to whether a change satisfies schema constraints, transaction semantics and authorization requirements at the operational layer. Contextual legitimacy concerns whether the same change is expected, behaviorally consistent with historical patterns, and plausible given the registrar, reseller, and domain context. This thesis positions detection in this gap: it assumes the attacker can produce structurally valid mutations and seeks to identify those that are contextually suspicious, high-risk or campaign-like [51], [57].

4.1 System Context, Trust Boundary and Detection Goal

Hybrid registry deployments typically require that systems and the registry database remain synchronized continuously, using replication or change-data-capture mechanisms [46], [47]. These pipelines are engineered for correctness and availability, not for interpreting the semantic legitimacy of changes. As a result, if an attacker can influence cloud-hosted systems or interfaces, malicious changes may flow into the authoritative state without violating schema constraints or authentication checks.

An important constraint follows from this synchronization requirement: the detection mechanism proposed is designed as a monitoring and alerting detection layer rather than as an inline prevention gate that blocks or delays replication. Replication mechanisms are responsible for maintaining strict consistency between cloud-hosted systems and the sovereign registry database. Interrupting or delaying this process risks violating consistency guarantees and may introduce instability within a critical infrastructure environment.

Accordingly, the proposed system analyzes mutation events as they are replicated and identifies anomalous or suspicious patterns, particularly in delegation and glue changes, without interfering with the replication flow. Flagged events trigger alerts accompanied by structured explanations, enabling rapid investigation and response, while synchronization between environments is preserved. This design reflects the practical trade-off between security monitoring and continuity in hybrid registry architectures.

4.1.1 Threat Realization in Hybrid Registry Environments

The threat model considered in this thesis is informed by analyses conducted by SIDN, which highlight how operating registry services (DRS) in a public cloud introduce a complex threat landscape in which multiple independent causes can lead to compromised DNS output. These causes include direct manipulation of registry data, adversary-in-the-middle attacks, supply-chain compromise of container infrastructure, and platform-level breaches. Although technically distinct, these attack vectors converge toward a common operational effect: unintended or malicious modification of registry data used for zone generation. From the perspective of the authoritative database, these different attack paths collapse into a shared observable outcome, namely state transitions that are inconsistent with legitimate intent.

SIDN's analysis further emphasizes that mitigating such risks requires a combination of preventive, detec-

tive, and corrective controls. Preventive controls reduce the likelihood of compromise through mechanisms such as segmentation, encryption, and least-privilege access control, while corrective controls focus on recovery through rollback procedures and incident response. Within this spectrum, the approach in this thesis is positioned as a detective control. It leverages transaction-level mutation data to identify anomalous or high-risk changes as they propagate through the system, thereby reducing time-to-detection without interfering with the replication process described above.

Concretely, the threat model can be translated into observable detection hypotheses at the registry level. Rather than attempting to identify specific intrusion mechanisms or compromise vectors directly, the proposed detection approach focuses on outcome-relevant state transitions within authoritative registry data. These include anomalous delegation changes introducing previously unseen nameserver infrastructure, glue record modifications redirecting resolution endpoints, deviations in registrar or reseller behavior and unexpected DNSSEC-related configuration changes. Additionally, correlated mutations across multiple domains may indicate coordinated abuse or large-scale compromise activity.

By focusing on observable mutation effects rather than the initial attack vector itself, the framework remains largely agnostic to the underlying compromise mechanism while directly targeting the security-relevant consequences of unauthorized registry manipulation. This is particularly important in hybrid environments where the original point of compromise may occur outside the authoritative registry infrastructure and therefore remain unobservable within registry-level telemetry.

4.2 Why Delegation and Glue are Prioritized

This thesis prioritizes delegation and glue mutations because they are:

- among the most security-critical registry-controlled inputs to zone generation,
- directly aligned with the hazard pathway described in Section 4.1.1,
- high-impact even when changes remain formally valid and policy-compliant.

Delegation defines which authoritative infrastructure answers for a domain, and glue binds nameserver hostnames to routable IP endpoints. A successful attacker does not need to break DNS protocols, it is sufficient to introduce formally valid changes that redirect resolution to attacker-controlled infrastructure. Compared to many other registry objects, delegation and glue changes provide a direct mechanism for traffic redirection, interception or service disruption, and their impact can be immediate and global once incorporated into signed zone output [30], [67].

From a threat-modeling perspective, delegation and glue integrity are therefore treated as central security objectives within the proposed detection framework. However, the system is not limited exclusively to delegation and glue analysis. Real-world DNS attacks frequently involve combinations of multiple mutation classes, including registrar-level abuse, DNSSEC manipulation, infrastructure reuse, contact-information takeover and coordinated multi-domain activity. Consequently, the implemented detection framework incorporates additional contextual and relational signals beyond delegation and glue alone.

This broader perspective is particularly important in cloud-connected registry environments, where the introduction of public-cloud-hosted components expands the attack surface through exposed APIs, credential theft opportunities, automation abuse, supply-chain compromise and platform misconfiguration [23]. Even if some threats exist in both local and public-cloud deployments, the threat model described in Section 4.1.1 motivates focusing on mutation classes whose probability and operational impact increase under cloud-integrated operational workflows.

Accordingly, while delegation and glue transitions form the primary analytical focus of the detection design, the framework additionally incorporates DNSSEC configuration changes, registrar-originated mutations, identity and contact modifications and cross-domain infrastructure relationships as complementary security-relevant mutation classes.

4.3 Attacker Model and Assumptions

This thesis assumes that the sovereign environment hosting the authoritative registry database and the detection system is trusted, while cloud-hosted components may be compromised. Attackers are assumed capable of producing mutations that are syntactically valid and consistent with authorization boundaries at the cloud layer (e.g., through credential theft, session hijacking, compromised automation or a compromised service image). The detection system does not attempt to attribute root cause, it flags suspicious authoritative state transitions that may warrant response.

The attacker is assumed capable of influencing registry mutations at the operational layer while remaining within formally valid workflows. This includes the ability to trigger authorized-looking changes, abuse automation at scale or introduce mutations to avoid simplistic detection strategies.

The attacker is not assumed to have direct access to the sovereign authoritative database. The detection objective is therefore to identify anomalous or high-risk authoritative state transitions prior to their global propagation.

4.4 Threat Scenarios Expressed in Registry Mutations

This thesis considers threat scenarios that manifest as anomalous or high-risk mutations within authoritative registry data. The analysis emphasizes how attacks become observable through changes in delegation, glue records, DNSSEC configuration and related registry state transitions. In addition, contextual dimensions such as actor behavior and mutation volume are incorporated. Each scenario therefore characterizes both the adversarial objective and the observable effects in registry mutation data.

DNSSEC provides an important preventive security mechanism within several of the threat scenarios discussed below by enabling validating resolvers to detect inconsistent DNS data and by establishing a cryptographic chain of trust between parent and child zones [7]. However, the effectiveness of DNSSEC depends on the specific attack scenario. Some attacks, such as DNSSEC configuration manipulation, directly target the trust model itself, while others involve legitimate-looking registry mutations that may ultimately be incorporated into correctly signed zones. Consequently, DNSSEC alone may not be able to prevent all forms of registry abuse, particularly when malicious changes originate from authorized workflows or compromise of trusted operational systems. This motivates the need for complementary detective controls capable of assessing the legitimacy of registry mutations before they become authoritative.

4.4.1 T1: Unauthorized Delegation Change (Nameserver Set Manipulation)

In this scenario, an adversary modifies the set of authoritative nameservers associated with one or more domains. The primary objective is to obtain control over DNS resolution, thereby enabling phishing, credential interception or traffic manipulation [53].

Such manipulation may occur as a complete replacement of all nameservers, a partial substitution in which malicious nameservers are introduced alongside legitimate ones, or a gradual (low-and-slow) migration. From a data perspective, these changes are reflected as modifications to the relationship between domains and nameserver and to nameserver objects themselves.

Observable indicators include large set differences between successive configurations, the introduction

of previously unseen nameserver identifiers, and deviations from historical domain or registrar behavior. A particularly subtle variant is selective insertion, where a single attacker-controlled nameserver is added, enabling intermittent or targeted interception while preserving apparent normality.

DNSSEC may limit the impact of delegation attacks if the attacker is unable to establish a valid chain of trust for the delegated infrastructure. However, delegation changes accompanied by legitimate or compromised DNSSEC updates may still result in successful redirection.

4.4.2 T2: Glue Record Manipulation (Nameserver IP Endpoint Changes)

Glue record manipulation enables traffic redirection without altering nameserver hostnames. Instead, the adversary modifies the IP addresses associated with existing nameservers, causing resolvers to connect to attacker-controlled endpoints while delegation appears unchanged.

This attack is especially stealthy, as many validation mechanisms primarily evaluate hostnames rather than underlying IP mappings. In-bailiwick nameservers are particularly sensitive, since their glue records are essential for successful resolution.

Observable signals include sudden changes in IP addresses, transitions to previously unseen address space, elevated churn in glue records, and inconsistencies with historical redundancy or geographic distribution patterns. Despite structural compliance, such changes can effectively redirect traffic.

Recent work has demonstrated that stale and manipulable DNS glue records can enable large-scale domain hijacking and denial-of-service attacks, highlighting the security risks associated with glue record inconsistencies and resolver behavior [78].

DNSSEC can prevent successful impersonation when attackers cannot produce valid DNSSEC signatures for the redirected infrastructure. Nevertheless, glue manipulation may still cause disruption, misrouting or denial-of-service effects, and may become more impactful when combined with DNSSEC-related attacks.

4.4.3 T3: Registrar/Reseller Compromise and Authorized Mutation Abuse

Registrars and resellers act as trusted intermediaries in domain management. If compromised, for example through credential theft, weak authentication controls, API abuse or manipulation of account recovery mechanisms, an adversary can gain persistent control over domain operations.

Typically, the attacker first establishes persistence by modifying account-level attributes or authentication settings, after which authorized-looking mutations are performed, including delegation changes, glue updates, DNSSEC modifications and domain transfers. Because these actions are executed through legitimate registrar interfaces, they are largely indistinguishable from benign administrative operations at the protocol level.

Recent research by van Hove et al. [31] demonstrates that registrar compromise represents a realistic and significant threat vector. Their study empirically analyzed the security posture of major domain registrars and resellers, highlighting weaknesses in authentication workflows, insufficient rate limiting for multi-factor authentication mechanisms and the risks associated with domain takeover scenarios. The authors emphasize that successful compromise of registrar accounts enables attackers to directly modify nameserver configurations, DNSSEC settings and transfer controls through legitimate administrative interfaces, thereby providing extensive control over critical domain infrastructure and associated services.

The study further demonstrates that domain takeover attacks may have operational consequences comparable to major cyber incidents such as ransomware or distributed denial-of-service attacks, due to the attacker's ability to redirect DNS traffic, intercept email, manipulate authentication flows and obtain access to connected cloud services. Observable indicators of this threat scenario include abnormal mutation

volumes, unusual temporal mutation patterns, deviations from historical registrar behavior and coordinated reuse of attacker-controlled infrastructure across multiple domains. Because registrar-originated mutations appear legitimate, detection relies heavily on identifying anomalous behavioral and relational patterns rather than protocol violations alone.

This scenario therefore represents both an empirically supported and operationally critical attack vector within modern DNS ecosystems.

Because attackers operate through legitimate administrative workflows, DNSSEC alone may be insufficient to prevent abuse if corresponding DNSSEC-related changes can also be performed through the compromised account.

4.4.4 T4: Coordinated Multi-Domain Takeover (Bulk Mutation Campaign)

In this scenario, an adversary modifies a large number of domains within a single transaction or a short time window. The objective is to maximize impact by simultaneously redirecting multiple domains to shared attacker-controlled infrastructure, thereby enabling large-scale phishing campaigns or service disruption [39].

Such patterns may arise from compromised registrar automation, shared infrastructure breaches, or coordinated attacks across multiple accounts.

Observable signals include unusually high numbers of affected domains per transaction, uniform mutation patterns across domains, and convergence toward a small set of shared nameserver or glue resources. Such bulk behavior is rare in benign operations and therefore highly indicative of compromise.

Although DNSSEC may continue to provide integrity guarantees for individual domains, it does not inherently detect or prevent coordinated abuse campaigns affecting large numbers of domains simultaneously.

4.4.5 T5: High-Impact Domain Targeting Based on Query Volume

Adversaries may prioritize domains with high query volumes, such as widely used services, financial platforms, or critical public infrastructure. These domains are attractive targets because even short-lived manipulation can affect a large number of users.

While the observable mutation patterns are similar to those in delegation or glue manipulation, this scenario introduces an additional contextual dimension: domain importance. Incorporating query volume or domain popularity enables impact-aware detection, whereby anomalies affecting high-value domains are prioritized for analysis and response.

DNSSEC can help protect the integrity of affected domains, but it does not reduce the attractiveness of high-value targets nor provide prioritization based on operational impact.

4.4.6 T6: Direct Registry or Database Compromise

In this high-severity scenario, an adversary gains direct access to registry infrastructure, including backend databases, cloud environments, or deployment pipelines. This bypasses registrar-level controls entirely and enables large-scale manipulation of authoritative data.

The attacker may modify multiple domains simultaneously, introduce new nameservers or glue records, and alter DNSSEC configurations without generating corresponding registrar interaction logs [75].

Observable indicators include high-impact mutation bursts, simultaneous changes across unrelated domains, and inconsistencies with expected registrar-driven workflows. The scale and lack of conventional access traces make this scenario particularly critical.

This scenario is particularly severe because an attacker with registry-level access may be capable of modifying both delegation information and the DNSSEC trust configuration itself, thereby bypassing protections that would otherwise be provided by DNSSEC.

4.4.7 T7: Cross-Domain Infrastructure Reuse (Fingerprint Reuse)

An adversary may reuse the same infrastructure, such as nameservers, IP ranges, or DNSSEC key material, across multiple domains. While individual mutations may appear benign, such reuse introduces structural correlations across otherwise unrelated domains [54].

This behavior reflects operational efficiency from the attacker's perspective, but also creates detectable patterns within registry data.

Observable signals include repeated use of identical nameserver identifiers, shared glue IP ranges, and identical DNSSEC key fingerprints across unrelated domains. This scenario is particularly relevant for graph-based detection approaches, where shared infrastructure manifests as anomalous connectivity patterns.

Because DNSSEC validates individual trust relationships rather than infrastructure usage patterns, it provides limited visibility into coordinated infrastructure reuse across otherwise unrelated domains.

4.4.8 T8: Suspicious DNSSEC Configuration Manipulation

DNSSEC mutations affect the cryptographic trust model underlying DNS resolution [76]. An adversary may manipulate DS records to weaken validation guarantees, induce denial-of-service conditions, or enable traffic interception.

Relevant attack patterns include:

- **DS record insertion:** adding new DS records alongside existing ones, thereby introducing attacker-controlled key material and enabling staged takeover or redirection,
- **DS record replacement:** replacing DS records with attacker-controlled keys to bypass validation,
- **DS record removal:** breaking the chain of trust, causing validating resolvers to reject responses and rendering domains inaccessible.

DS insertion is particularly subtle, as it allows the adversary to introduce a new trust anchor without immediately disrupting resolution. Observable signals include unexpected DS changes, unusual key churn, deviations in cryptographic algorithms, and correlation with delegation or glue modifications.

This scenario is very significant because it directly targets the DNSSEC trust model itself, potentially weakening the protections that DNSSEC is intended to provide.

4.4.9 T9: TTL Manipulation and Persistence Amplification

An adversary may manipulate Time-To-Live (TTL) values to increase the persistence and impact of malicious changes. By increasing TTL values prior to introducing malicious delegation, glue, or DNSSEC modifications, incorrect data may remain cached for extended periods.

This technique amplifies the duration of attacks, allowing their effects to persist even after remediation.

Observable signals include sudden TTL increases relative to historical baselines and temporal correlation with other high-risk mutations.

Since TTL values influence caching behavior rather than data authenticity, DNSSEC generally does not mitigate the persistence effects introduced by TTL manipulation.

4.4.10 T10: Long-Term Low-and-Slow Manipulation

In low-and-slow scenarios, adversaries introduce incremental changes over extended periods in order to avoid detection mechanisms that rely on burst-based anomalies. The objective is to remain within expected variance while gradually shifting control.

Detection in this setting relies on identifying long-term drift, novelty, and deviations from domain or registrar specific baselines, underscoring the importance of context-aware and temporal modeling.

DNSSEC validates the integrity of published DNS data but does not detect gradual behavioral drift or long-term operational changes that can remain cryptographically valid.

4.4.11 T11: Domain Contact and Identity Hijacking

In this scenario, an adversary modifies the administrative or technical contact information associated with a domain in order to gain control over domain management processes [31]. Rather than immediately altering delegation or DNSSEC configuration, the attacker first attempts to compromise the administrative identity layer surrounding the domain.

This may include changing the registered contact name, email address, telephone number or postal address, as well as assigning attacker-controlled contacts to administrative or technical roles. Such modifications can facilitate subsequent malicious actions, including registrar account recovery abuse, unauthorized transfer requests, interception of registrar communications or later delegation and DNSSEC manipulation.

From a registry mutation perspective, this attack manifests as abnormal changes in contact records and contact-role relationships associated with domains. Observable indicators include sudden changes to administrative contact information, introduction of previously unseen contact identities, modifications to communication details and unusual reassignment of administrative roles for high-value domains.

Although these mutations may appear administratively valid, they are operationally significant because control over contact information can enable long-term persistence and reduce visibility into subsequent malicious activities.

This scenario is closely related to T3 (Registrar/Reseller Compromise and Authorized Mutation Abuse), as both attacks target the administrative control plane of domain management. However, while T3 focuses on compromise of registrar infrastructure or registrar accounts themselves, T11 specifically emphasizes manipulation of domain ownership and identity-related metadata as a precursor or supporting mechanism for broader domain takeover operations.

Because this attack targets administrative ownership and identity information rather than DNS data itself, DNSSEC provides little direct protection against this form of abuse.

4.4.12 T12: Multi-Stage Attack Sequences

Sophisticated attacks may unfold as sequences of individually plausible mutations, such as account modifications, registrar transfers, and subsequent delegation or DNSSEC changes.

These attacks may not be detectable when considering individual mutations in isolation. Effective detection therefore requires correlating events across time windows and identifying structured sequences of related changes that collectively indicate malicious activity.

While DNSSEC may mitigate certain individual stages within such attack chains, it is not designed to detect the broader sequence of coordinated actions that collectively indicate malicious activity.

4.5 Positioning of the ML Model

Aligned with Section 4.1.1, the ML component operates as a detective control on registry mutation events [66], [70], producing anomaly scores and structured explanations from near-real-time transactional data.

Because synchronization between cloud-hosted systems and the registry database must remain continuous, the framework is primarily designed for alerting, analyst triage and investigation support rather than blocking replication. This design avoids introducing a single point of failure within the replication pipeline while reducing time-to-detection for suspicious mutations.

In high-confidence scenarios involving indicators such as coordinated multi-domain changes or correlated delegation and DNSSEC anomalies, downstream processes such as DNSSEC signing or zone publication could theoretically be temporarily paused for affected transactions. Such an approach preserves synchronization while selectively delaying publication, introducing a controlled trade-off between availability and security.

To support practical usability, detection outputs must remain interpretable and context-aware, including information about affected registrars, domains, infrastructure relationships and dominant anomaly-contributing features.

5 Dataset: Temporally Versioned Authoritative Registry State

5.1 Data Source and Temporal Modeling

The dataset consists of authoritative registry mutation data stored in a normalized relational database representing the state of registry-managed DNS objects and their evolution over time. Unlike traffic traces or resolver logs, the dataset captures authoritative configuration state transitions that directly influence zone publication.

The schema models domains, nameservers, delegation relationships, glue endpoints, DNSSEC-related configurations and organizational actors such as registrars and resellers. All entities follow a temporally versioned design in which logical objects are represented as sequences of time-bounded records. Validity intervals and globally ordered identifiers enable deterministic reconstruction of historical registry state and mutation sequences.

Changes belonging to the same action are linked through shared transaction identifiers, allowing related updates across multiple objects to be grouped into a single mutation event. Certain identity attributes may be partially masked within the research environment. However, the detection objectives primarily rely on structural and behavioral mutation patterns rather than identity-specific information.

5.2 Delegation and Glue Model

Delegation integrity is defined by the interaction between domains, nameservers, and the relationships that bind them. Delegation specifies which authoritative nameservers are responsible for answering queries for a domain.

Changes to delegation correspond to modifications in the set of nameservers associated with a domain, including additions, removals, or complete replacement of the delegation set.

Glue information is modeled as the association between nameservers and their corresponding IP addresses. These associations are necessary for resolving nameservers that reside within the domain they serve. Changes in glue represent modifications to the infrastructure endpoints through which authoritative services are reached.

Together, delegation and glue define the effective resolution path of a domain at any point in time and therefore constitute a critical part of the authoritative control plane.

5.3 Mutation Event Representation

A mutation event is defined as a set of related changes executed together as part of a single transaction. Such events may involve simultaneous updates across domains, nameservers, infrastructure endpoints and DNSSEC configurations.

For analytical purposes, authoritative state is reconstructed immediately before and after each mutation event using temporal validity intervals, while event timestamps are derived from transaction commit times. This formulation enables logically related changes to be analyzed as a single unit, supporting consistent reasoning about multi-object and multi-step updates.

Each mutation event is transformed into a fixed-length feature representation for anomaly detection. Feature extraction encodes both structural and behavioral characteristics by comparing system state before

and after the mutation.

Structural features capture configuration-level changes such as delegation modifications, glue endpoint updates and DNSSEC-related transitions. Behavioral features model contextual properties including temporal dynamics, mutation burstiness and coordinated activity across multiple domains.

This combined representation captures both the structural impact of mutations and their operational context, enabling detection of events that remain syntactically valid but are anomalous in structure, scale or coordination.

5.4 Learning Paradigm and Evaluation

The dataset does not provide explicit ground-truth labels for malicious mutations. Structurally significant delegation, glue or DNSSEC changes may occur during legitimate processes such as provider migration, infrastructure consolidation or DNSSEC key rollover [34]. At the same time, attackers compromising registrar accounts or workflows may introduce malicious changes that appear syntactically and operationally valid.

This ambiguity makes direct classification difficult and motivates the use of unsupervised anomaly detection. Models are therefore trained on historical mutation events known to be benign, while events deviating significantly from learned structural or behavioral baselines are flagged as suspicious.

Evaluation is conducted using schema-valid synthetic attack scenarios derived from realistic mutation patterns, including delegation manipulation, glue redirection and DNSSEC-related configuration abuse while preserving temporal and relational consistency.

5.5 Privacy and Ethical Considerations

The dataset reflects registry activity and contains sensitive identifiers. This research focuses exclusively on structural mutation patterns and aggregated behavioral characteristics.

The research approach and data usage have been reviewed in consultation with the internal privacy committee at SIDN, ensuring alignment with organizational confidentiality requirements and responsible data handling practices.

6 Methodology and System Architecture

6.1 Framework Overview

The proposed system is designed as a modular, multi-stage pipeline that integrates data ingestion, feature engineering, graph construction, representation learning, unsupervised learning, anomaly scoring and explainability into a cohesive anomaly detection framework. At a high level, the system transforms raw DNS registry data into structured representations, learns normal behaviour patterns through machine learning models and produces anomaly alerts enriched with explanatory signals. Each stage of the pipeline is implemented as an independent component, enabling flexible experimentation and future extensibility.

The modular structure is explicitly reflected in the implementation, where distinct classes encapsulate each stage of the pipeline. For example, data extraction is handled by a dedicated access layer, feature engineering is encapsulated in a feature builder module, graph construction is managed by a graph builder and model training is performed by specialized learning components. This separation of concerns ensures that individual components can be modified or placed without affecting the overall system.

Figure 1 presents an overview of the complete anomaly detection pipeline developed in this research. The framework transforms authoritative DNS registry mutation data into both graph-based and feature-based representations, which are subsequently analyzed using complementary anomaly detection models. The resulting anomaly signals are combined through an ensemble scoring mechanism before being passed to thresholding and explainability components.

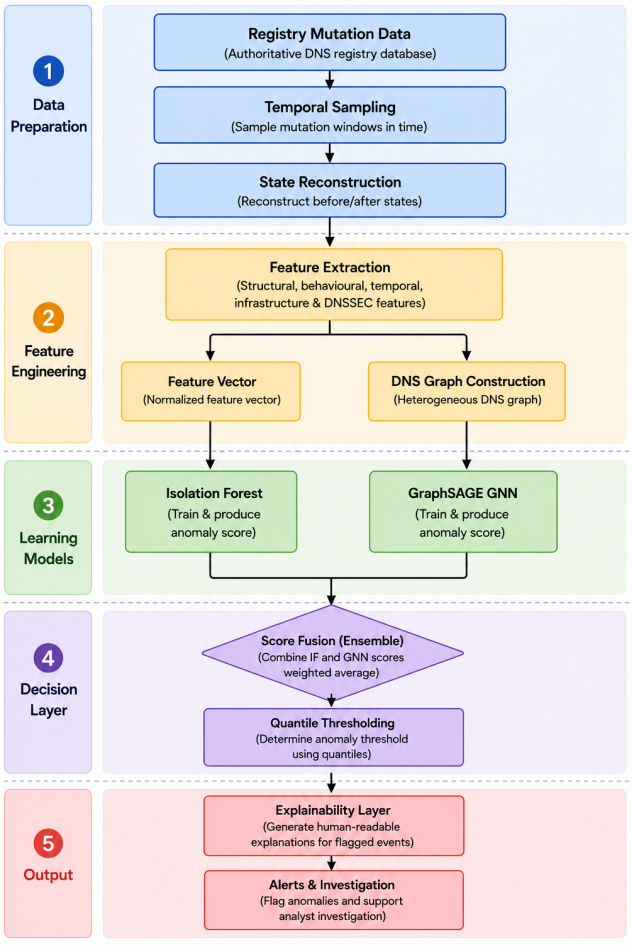


Figure 1: Overview of the proposed anomaly detection framework.

6.2 Data Processing and Temporal Sampling

As illustrated in Figure 1, temporal sampling constitutes the first stage of the framework and is responsible for constructing the historical observation windows used throughout the remainder of the anomaly detection pipeline.

The first stage of the framework involves extracting relevant data from multiple relational tables, including delegation records, glue records, domain metadata, registrar information and DNSSEC-related data. Together, these sources capture multiple aspects of DNS operations, ranging from infrastructure configuration to cryptographic state transitions.

To improve temporal generalization and reduce sensitivity to short-term fluctuations, the framework adopts a multi-window temporal sampling strategy during dataset construction. Rather than training on a single adjacent historical dataset, multiple overlapping temporal windows with varying durations and offsets are sampled independently and subsequently combined.

Long-duration windows capture stable registry behavior and long-term infrastructure patterns, while shorter windows capture evolving trends and recent mutation dynamics. Additional temporal offsets are introduced to reduce direct temporal leakage between training, validation and evaluation datasets.

This strategy exposes the learning algorithms to heterogeneous conditions across time, improving robustness against temporal drift and reducing overfitting to narrow periods or short-lived activity bursts. The resulting training dataset therefore reflects a combination of long-term baseline behavior and recent dynamics.

6.3 Feature Engineering and Signal Extraction

Following state reconstruction, the feature engineering stage shown in Figure 1 extracts behavioral, structural, temporal and DNSSEC-related signals that form the basis of both the feature-based and graph-based anomaly detection branches.

6.3.1 Behavioral and Structural Features

A comprehensive feature engineering pipeline was developed to transform raw registry mutation data into structured representations suitable for anomaly detection. The extracted features capture multiple dimensions of DNS behavior, including statistical, temporal, structural and DNS-specific characteristics. Statistical features include counts, ratios and entropy-based measures, such as nameserver entropy and subnet entropy, which quantify infrastructure diversity and distributional behavior. Novelty-based features are derived using historical baseline statistics, enabling computation of unseen ratios, inverse frequency measures and rarity indicators that highlight unusual entities or transitions. Temporal features capture mutation timing and activity dynamics through cyclic time encodings and burst-detection metrics that identify sudden increases in operational activity. Sequence-based features additionally model mutation ordering behavior, including patterns such as add-before-delete sequences, bigram rarity [10] and domain churn ratios. The pipeline further incorporates DNSSEC-specific features, including key-tag diversity, algorithm entropy and key reuse across domains. These signals are particularly relevant for identifying anomalous cryptographic behavior and suspicious trust-relationship transitions. To ensure consistency across training and inference, the feature pipeline maintains internal normalization parameters and baseline statistics that are reused during evaluation. All features are normalized using median and median absolute deviation (MAD) based robust scaling [16], [58] in order to reduce sensitivity to extreme values and stabilize feature distributions during training.

6.3.2 Dataset Balancing

An important consideration during dataset construction was the imbalance between different mutation categories present within registry data. Delegation-related mutations occur substantially more frequently than glue-related modifications, introducing the risk that dominant mutation categories disproportionately influence both feature learning and graph representation learning. To mitigate this effect, upper bounds were imposed on the number of sampled delegation and glue mutation tickets within each temporal window. Equal sampling limits were intentionally used for both categories to improve representation diversity and preserve sensitivity to infrastructure-level anomalies involving nameserver IP transitions and subnet manipulation. This balancing strategy reduces bias toward dominant mutation patterns while improving the ability of the framework to identify subtle glue-related anomalies that might otherwise become underrepresented within the training data.

6.3.3 Domain Importance Weighting

In addition to behavioral and structural signals, the framework incorporates contextual domain importance information derived from Entrada query statistics [72]. Query frequency is used as a proxy for operational importance and potential attack impact. Domains with consistently high query volumes are treated as higher-priority assets within the anomaly scoring process because disruptions or malicious modifications affecting widely queried domains may impact large numbers of users and services before detection or mitigation occurs. Government-related domains additionally receive elevated contextual weighting due to their sensitivity, potential national security relevance and the broader societal consequences that may arise if such domains are compromised.

6.4 Exploratory Modeling and Limitations of Classical Methods

Prior to adopting graph-based representation learning and ensemble anomaly detection, several classical unsupervised learning techniques were evaluated in order to better understand the structure of the dataset and assess whether anomalous behavior could be identified directly within the feature space.

6.4.1 PCA-Based Analysis

Principal Component Analysis (PCA) [2] was initially applied to the engineered feature space for dimensionality reduction and exploratory visualization. However, the resulting projections revealed a highly homogeneous structure with no clear cluster separation or visually distinguishable outliers.

This behavior is consistent with the nature of the dataset, which consists of benign operational registry mutations. Because PCA captures directions of maximum variance, subtle anomalies embedded within relational, temporal or behavioral dependencies remain difficult to distinguish in low-dimensional projection space.

6.4.2 Clustering Approaches

Clustering algorithms, including K-means and K-medoids, were subsequently evaluated in order to identify potential outliers through cluster distance and density characteristics. However, these methods proved insufficient for the registry mutation dataset.

The primary limitation is that clustering approaches implicitly assume the existence of meaningful separable cluster structures within the feature space. In practice, the observed clusters primarily reflected variations in benign operational behavior rather than meaningful separation between normal and anomalous

activity. Additionally, all observations are forcibly assigned to clusters, making it difficult to distinguish between legitimate operational variation and genuinely suspicious mutations.

6.4.3 Standalone Isolation Forest

Isolation Forest [40] was then evaluated as a dedicated anomaly detection method. Unlike clustering approaches, Isolation Forest identifies anomalies through recursive random partitioning, making it more effective for detecting feature-level outliers.

Although the model demonstrated some ability to identify statistically unusual mutations, it remained limited by its lack of relational awareness. Many DNS-related attacks emerge through coordinated infrastructure relationships across domains, nameservers, registrars and IP subnets rather than through extreme feature values in isolation.

These observations motivated the transition toward graph-based representation learning and ultimately the ensemble-based anomaly detection framework proposed in this research.

6.5 Heterogeneous Graph Representation Learning

As shown in Figure 1, one branch of the framework transforms the reconstructed DNS state into a heterogeneous graph representation that captures relationships between domains, nameservers, registrars, subnets and DNSSEC-related entities.

To effectively model the relational structure inherent in DNS registry data, this research employs a heterogeneous Graph Neural Network (GNN) architecture as the primary representation learning mechanism. Unlike traditional machine learning models that operate on independent samples, GNNs explicitly model dependencies between entities, making them particularly suitable for cybersecurity applications where interactions between domains, nameservers, registrars and infrastructure components are central to identifying anomalous behavior [71].

6.5.1 Heterogeneous Graph Construction

The DNS ecosystem is represented as a heterogeneous graph consisting of multiple node types, including tickets, domains, nameservers, IP subnets and registrars. Relationships between these entities are encoded as typed edges, such as ticket-to-domain, ticket-to-nameserver, domain-to-registrar and nameserver-to-subnet relationships. This design reflects the real-world structure of DNS systems, where operational behavior emerges through interactions between multiple interconnected entities.

Ticket nodes are associated with rich feature vectors derived from the feature engineering pipeline, while additional entity types are initialized using simpler statistical representations such as frequency-based features and infrastructure occurrence statistics.

The graph is implemented using the `HeteroData` structure with explicitly defined edge types in order to preserve semantic distinctions between different operational relationships.

6.5.2 GraphSAGE Architecture

The Graph Neural Network architecture is based on a two-layer heterogeneous GraphSAGE model implemented using `HeteroConv` and `SAGEConv` layers. Each node type is first projected into a shared latent representation space using type-specific linear transformations before message passing is performed across all edge types.

The first convolutional layer captures immediate neighborhood information, while the second layer enables higher-order relational interactions by propagating information across multiple graph hops. Non-linear

activation functions and dropout are applied between layers to improve representational expressiveness and reduce overfitting.

GraphSAGE was selected because of its inductive learning capabilities, allowing the model to generalize to previously unseen nodes and subgraphs without requiring complete graph reconstruction [28]. This property is particularly important in DNS environments where new domains, nameservers and infrastructure relationships continuously emerge.

6.5.3 Reverse Relations and Bidirectional Message Passing

To improve relational representation learning, reverse edges were explicitly added for every edge type within the graph. This enables bidirectional message propagation during training, allowing contextual information to flow both upstream and downstream across the graph structure.

Bidirectional propagation is particularly important in DNS ecosystems because anomalous behavior rarely originates from isolated entities. Instead, suspicious activity often emerges through coordinated relationships between domains, nameservers, registrars and shared infrastructure components.

For example, a ticket modifying a nameserver may appear benign when considered independently, while the associated nameserver may already be linked to suspicious domains, anomalous subnet allocations or previously observed malicious infrastructure behavior. Reverse message passing allows these contextual anomaly signals to propagate back toward the originating ticket node.

Similarly, reverse relations improve the ability of the model to detect coordinated attacks involving infrastructure reuse across multiple domains. When several domains share suspicious nameservers, IP subnets or DNSSEC material, anomaly signals can propagate through neighboring graph connections and strengthen the overall relational anomaly representation.

6.5.4 Self-Supervised Link Prediction Objective

The Graph Neural Network is trained using a self-supervised link prediction objective. During training, the model learns to reconstruct observed graph relationships while distinguishing them from synthetically generated edges produced through negative sampling.

This reconstruction objective encourages embeddings of structurally related entities to remain close within latent space, while unrelated entities become separated. As a result, tickets exhibiting unusual relational behavior become increasingly difficult for the model to reconstruct accurately.

From an anomaly detection perspective, this approach is particularly advantageous because it does not require labeled attack data and instead focuses on modeling normal relational behavior within the DNS ecosystem. Nodes and relationships that deviate significantly from learned graph structure naturally emerge as anomalies through elevated reconstruction difficulty.

6.5.5 Hyperparameter Configuration

The heterogeneous GraphSAGE model was implemented using hidden representations of 96 dimensions and final embedding vectors of 64 dimensions. The hidden dimensionality was selected to provide sufficient representational capacity for modeling complex relational dependencies while avoiding excessive model complexity and overfitting.

The model was trained for 40 epochs using the Adam optimizer with a learning rate of 2×10^{-3} and weight decay regularization of 10^{-4}. A dropout rate of 0.15 was applied between graph convolution layers to improve generalization across unseen graph regions.

Gradient clipping was additionally applied during optimization in order to stabilize training and prevent exploding gradients during message passing across large heterogeneous graph structures.

To evaluate optimization stability and convergence behavior, the reconstruction loss of the heterogeneous Graph Neural Network was monitored throughout training. Figure 2 presents the evolution of both the training reconstruction loss and the evaluation reconstruction loss across training epochs.

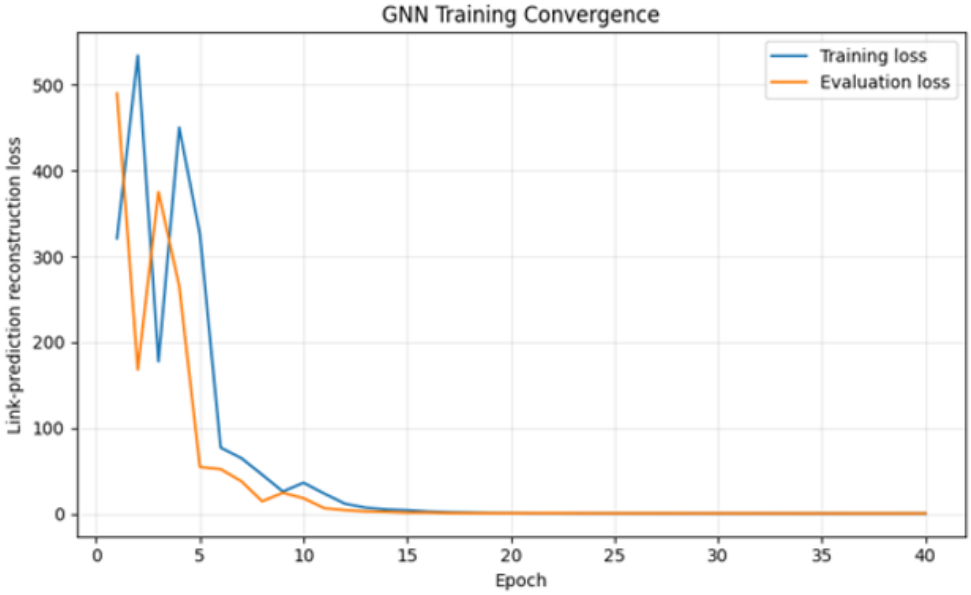


Figure 2: Training and evaluation reconstruction loss of the heterogeneous Graph Neural Network across epochs.

The model initially exhibited relatively high reconstruction loss values during the first training epochs. This behavior is expected due to random parameter initialization and the stochastic nature of negative edge sampling. As training progressed, both curves decreased rapidly, indicating that the model successfully learned increasingly accurate representations of the heterogeneous DNS graph structure.

After approximately 10 to 15 epochs, both the training and evaluation reconstruction losses stabilized near zero, demonstrating convergence of the optimization process. The close alignment between both curves suggests that the model generalized consistently across both training and evaluation data without signs of memorizing only the training graph structure. This indicates that the GraphSAGE-based heterogeneous GNN was able to effectively reconstruct normal relational patterns between tickets, domains, nameservers, IP subnets and registrars.

From an anomaly detection perspective, successful reconstruction of normal graph structure is particularly important, as anomalies correspond to nodes and relationships that deviate from these learned structural patterns. Consequently, the low reconstruction loss obtained during training supports the suitability of the learned embeddings for downstream anomaly scoring within the ensemble framework.

6.6 Isolation Forest Feature-Level Anomaly Detection

In parallel with graph construction, the feature-based branch illustrated in Figure 1 analyzes normalized feature vectors using Isolation Forest to identify statistical and behavioral anomalies that may not significantly alter graph structure.

While the Graph Neural Network effectively captures relational anomalies, it does not explicitly model feature-level outliers in isolation. To address this limitation, Isolation Forest is incorporated as a complementary anomaly detection component.

Isolation Forest isolates anomalies through recursive random partitioning of the feature space [40]. Because anomalous observations occupy sparse or statistically unusual regions of the feature distribution, they typically require fewer partitions to become isolated and therefore exhibit shorter average path lengths compared to normal samples.

6.6.1 Algorithmic Properties

Unlike distance-based or density-based anomaly detection methods, Isolation Forest does not explicitly model the full data distribution. Instead, anomalies are identified through randomized recursive partitioning, making the approach computationally efficient and well-suited for high-dimensional feature spaces [4], [40].

In this research, Isolation Forest is applied directly to the engineered feature vectors produced by the feature pipeline. These features capture statistical, temporal and behavioral characteristics of DNS operational activity, enabling the model to identify feature-level abnormalities such as unusual mutation frequencies, entropy shifts, rare infrastructure transitions and anomalous DNSSEC-related behavior.

6.6.2 Hyperparameter Configuration

The Isolation Forest component was configured using an ensemble of 500 isolation trees in order to improve anomaly score stability and reduce variance across heterogeneous feature distributions. The maximum sample ratio was set to 0.8, allowing each tree to observe a substantial subset of the training data while preserving the randomness necessary for effective anomaly isolation.

Robust scaling using median and median absolute deviation normalization was applied prior to training in order to reduce sensitivity to extreme feature values and stabilize feature distributions within the heavy-tailed registry mutation dataset.

The final configuration was selected empirically based on anomaly ranking stability, false positive rate reduction and consistency across multiple temporal evaluation windows.

6.7 Ensemble-Based Anomaly Scoring

The overall architecture presented in Figure 1 combines the outputs of the GraphSAGE and Isolation Forest components through an ensemble scoring mechanism. This design enables the framework to simultaneously capture relational anomalies within DNS infrastructure and statistical anomalies within mutation behavior.

The final anomaly detection framework adopts an ensemble-based detection strategy combining heterogeneous graph representation learning with feature-level anomaly detection. This design is motivated by the observation that no single model is sufficient to capture the full spectrum of anomalies that may occur within DNS registry mutation data.

The Graph Neural Network primarily captures relational and structural anomalies, including unusual interactions between domains, nameservers, registrars, IP subnets and DNSSEC-related infrastructure relationships. Through heterogeneous message passing, the model learns contextual dependencies between interconnected entities and is therefore particularly effective for identifying coordinated infrastructure reuse, anomalous delegation behavior and graph-structural inconsistencies.

In contrast, Isolation Forest operates directly on the engineered feature vectors and captures localized feature-level abnormalities that may not substantially alter graph topology. These include unusual mutation frequencies, entropy shifts, rare infrastructure transitions, temporal irregularities and anomalous DNSSEC-related characteristics.

Empirical evaluation demonstrated that the Graph Neural Network generally produced stronger anomaly separation for attacks involving coordinated infrastructure behavior and structural inconsistencies. However,

certain statistical outliers and localized deviations remained more visible within feature space. The complementary strengths of both approaches therefore motivated the use of an ensemble architecture rather than reliance on a single anomaly detection model.

In addition to the machine learning components, the framework incorporates a lightweight rule-based detection layer intended to improve robustness and interpretability. This layer encodes domain-specific DNS knowledge, including patterns related to delegation sequences, DNSSEC-related behavior and infrastructure transition logic. By combining multiple weak signals, the rule-based layer can help identify edge cases that may not independently produce sufficiently high anomaly scores while additionally providing human-readable contextual explanations.

6.7.1 Score Fusion and Thresholding

To combine both perspectives, normalized anomaly scores produced by the Graph Neural Network and Isolation Forest are aggregated using a weighted linear combination:

$$S_{ensemble} = w_{gnn} \cdot S_{gnn} + w_{if} \cdot S_{if}$$

where S_{gnn} represents the graph-based anomaly score, S_{if} represents the Isolation Forest anomaly score and w_{gnn} and w_{if} control the relative contribution of both models.

In the final implementation, the weighting was configured to emphasize the graph-based component while retaining a meaningful contribution from the Isolation Forest. This reflects the importance of relational anomalies within DNS ecosystems while still accounting for statistical feature-level deviations.

To improve stability and reduce sensitivity to extreme outliers, the framework adopts a quantile-based thresholding strategy based on the empirical anomaly score distribution observed during training rather than relying on fixed numerical cutoffs. In the final implementation, the anomaly threshold was calibrated near the 99.5th percentile of the training score distribution in order to maintain low false positive rates within highly imbalanced environments.

In real cybersecurity settings, excessive false positive rates significantly reduce the practical usability of anomaly detection systems by overwhelming analysts with low-quality alerts. Consequently, false positive reduction was treated as a primary design objective throughout the development process.

This thresholding strategy provides several advantages, including stable alert volumes, reduced sensitivity to score scale variations, adaptability across temporal windows and explicit control over expected false positive rates. The threshold calibration process therefore balances detection sensitivity with operational feasibility, ensuring that the resulting alert stream remains manageable for analysts while still capturing high-risk anomalies.

The ensemble approach offers several practical advantages. First, it improves robustness by reducing the likelihood of missed detections, since anomalies that are weakly represented within one model may still be captured by the other. Second, it enables simultaneous detection of both relational and feature-level anomalies rather than treating them as competing explanations. Third, it mitigates model-specific biases by balancing graph-structural and statistical perspectives.

Moreover, the ensemble design additionally provides a flexible framework that can be extended with additional detection signals, including rule-based heuristics, threat intelligence feeds or domain-specific indicators. The combined scoring strategy also enables more nuanced prioritization of suspicious mutations according to multiple contextual factors, which is particularly important in real-world cybersecurity environments where false positives must remain manageable while preserving high detection sensitivity.

6.8 Explainability and Operational Interpretation

As illustrated in Figure 1, explainability represents the final stage of the framework and operates as a post-hoc analysis layer that provides human-readable explanations for anomalies identified by the ensemble detector.

A central design objective of the proposed framework is the generation of interpretable explanations for anomalous registry mutations. Because the system is intended to support analysts within critical DNS infrastructure environments, explanations are necessary to bridge the gap between anomaly scores and human investigation workflows.

The explainability layer is implemented strictly as a post-hoc read-only analysis component. Explanations are generated only after anomaly scoring has completed and therefore do not influence anomaly scores, thresholds or ranking behavior. This separation preserves score stability and ensures that explainability mechanisms cannot alter model predictions or increase false positive rates.

The framework incorporates two complementary explainability mechanisms: feature-level explainability for the Isolation Forest component and graph-based relational explainability for the GNN component.

6.8.1 Feature-Level Explainability Using SHAP

For the Isolation Forest component, feature-level explanations are generated using SHAP (SHapley Additive exPlanations) [48]. SHAP estimates the contribution of individual features toward the anomaly score assigned to each mutation event.

The resulting explanations identify which engineered behavioral signals contributed most strongly to anomalous behavior. These signals include novelty ratios, rare nameserver usage, subnet churn, DNSSEC-related mutations, registrar anomalies and sequence-level mutation behavior. Typical high-contributing signals include unusually high nameserver novelty ratios, previously unseen glue subnets, abnormal DNSSEC key insertions, sudden mutation bursts, large-scale multi-domain mutation behavior and rare mutation sequences.

SHAP explanations provide localized mutation-level interpretability and allow analysts to understand why a mutation deviates from learned baseline behavior. This is particularly important in DNS registry environments where legitimate events, such as infrastructure migrations or DNSSEC maintenance operations, may initially appear anomalous despite being benign.

6.8.2 Graph-Based Relational Explainability

While SHAP explains feature-level anomalies, it cannot capture anomalies emerging from graph structure and relational dependencies. To address this limitation, the framework incorporates graph-based explainability mechanisms for the Graph Neural Network component.

The graph explainability layer decomposes anomaly behavior into relation-level and infrastructure-level signals by analyzing neighboring graph relationships and connectivity patterns. During inference, the system computes relation-specific anomaly statistics describing how strongly ticket nodes deviate from expected graph connectivity behavior.

These explanations include anomalous registrar relationships, unusual domain-to-nameserver mappings, suspicious nameserver-to-subnet relationships, infrastructure reuse across unrelated domains, abnormal concentrations of anomalous neighboring entities and coordinated multi-domain mutation behavior.

The system additionally computes anomaly propagation statistics across neighboring entities, enabling identification of the specific infrastructure components contributing most strongly to the anomaly signal.

Unlike traditional tabular explainability approaches, graph-based explainability provides structural context describing how entities interact abnormally within the DNS ecosystem. This is particularly important

for identifying coordinated infrastructure manipulation campaigns and cross-domain relationships that cannot be captured through feature-level analysis alone.

Together, feature-level SHAP explanations and graph-based relational analysis provide an interpretable anomaly detection framework that supports incident triage, registrar communication and infrastructure-level investigation within authoritative DNS registry environments.

6.9 Synthetic Attack Injection and Evaluation

Due to the absence of publicly available labeled datasets containing malicious DNS registry mutations, evaluation was conducted using threat-model-driven synthetic attack scenarios derived from realistic operational mutation behavior.

Because labeled DNS attack data is extremely limited, operationally sensitive and highly imbalanced, the proposed framework adopts an unsupervised anomaly detection approach rather than supervised classification. In operational registry environments, malicious behavior is rare and continuously evolving; supervised methods would therefore risk overfitting to previously observed attack patterns while failing to generalize to novel infrastructure abuse techniques.

The self-supervised graph reconstruction objective and Isolation Forest ensemble design instead allow the framework to model normal operational behavior directly, enabling detection of previously unseen anomalies without requiring extensive labeled attack datasets.

6.9.1 Threat Scenario Selection

Although 12 conceptual threat scenarios were identified in Section 4.4, only a subset 7 of was implemented as synthetic attack injections within the experimental framework. The implemented scenarios were selected based on three primary considerations: operational realism, relevance to observable registry mutation data and feasibility within the available dataset and system architecture.

The selected attacks primarily target delegation relationships, glue infrastructure, registrar behavior, DNSSEC trust configuration, temporal mutation behavior and coordinated multi-domain activity. These scenarios were chosen because they represent realistic and operationally relevant DNS abuse patterns that can be directly observed through registry mutations.

To further validate the operational plausibility of the selected attacks, discussions were conducted with domain experts from SIDN and researchers from NLnet Labs working on DNS infrastructure, registry operations and Internet security. During these discussions, delegation manipulation, glue record abuse, registrar compromise and DNSSEC misuse were consistently identified as some of the most critical and plausible attack vectors within registry environments. The selected attacks additionally align closely with previously documented DNS takeover incidents and operational threat scenarios.

Another important motivation for selecting these scenarios was attack diversity. The implemented attacks intentionally cover multiple layers of DNS operations, including delegation relationships, infrastructure-level glue mappings, registrar behavior, DNSSEC trust configuration, identity and contact metadata, temporal mutation behavior and coordinated multi-domain activity. This diversity ensures that the evaluation does not focus on a single anomaly type, but instead assesses the ability of the proposed framework to detect both structural and feature-level anomalies across different operational contexts.

Furthermore, the selected attacks are particularly suitable for evaluating the complementary strengths of the ensemble architecture. Delegation manipulation, infrastructure reuse and coordinated takeover attacks strongly exercise the relational capabilities of the GNN, while DNSSEC anomalies, mutation bursts, registrar-level deviations, gradual behavioral drift and contact-level modifications generate feature-space ab-

normalities that are effectively captured by the Isolation Forest component. Consequently, the implemented attack set provides a comprehensive evaluation scenario for the combined graph-based and feature-based anomaly detection framework.

6.9.2 Implemented Attack Scenarios

The implemented synthetic attacks correspond primarily to the following threat scenarios:

- **T1: Unauthorized Delegation Change** through malicious nameserver replacement, selective nameserver insertion and delegation hijacking,
- **T2: Glue Record Manipulation** through unauthorized changes to nameserver IP address mappings and subnet transitions,
- **T3: Registrar/Reseller Compromise and Authorized Mutation Abuse** through simulations of authorized-looking large-scale mutations originating from compromised registrar infrastructure,
- **T4: Coordinated Multi-Domain Takeover** through bulk mutation campaigns affecting multiple domains simultaneously and redirecting them toward shared infrastructure,
- **T8: Suspicious DNSSEC Configuration Manipulation** through DNSSEC key insertion, DS-like replacement behavior and anomalous cryptographic configuration changes,
- **T10: Long-Term Low-and-Slow Manipulation** through weaker and less aggressive mutation behavior designed to remain closer to normal operational variance,
- **T11: Domain Contact and Identity Hijacking** through anomalous administrative contact changes, attacker-controlled contact reassignment and operationally plausible identity-related mutations.

Several implemented attacks additionally reflect characteristics of other conceptual threat scenarios. Attacks targeting government domains and highly queried domains partially correspond to **T5** (high-impact domain targeting), since the framework explicitly prioritizes operationally important domains using Entrada query statistics and contextual weighting.

Similarly, infrastructure reuse across domains contributes to patterns associated with **T7** (cross-domain infrastructure reuse), particularly through shared nameserver, subnet and DNSSEC relationships represented within the heterogeneous graph structure.

Some scenarios additionally combine multiple mutation types simultaneously, partially reflecting the sequential and composite behavior described in **T12** (multi-stage attack sequences). For example, registrar-oriented attacks may combine delegation modifications, glue updates and DNSSEC-related changes within a single coordinated workflow.

All synthetic mutations were implemented while preserving schema validity, temporal consistency and operational plausibility in order to approximate realistic registry behavior as closely as possible.

6.9.3 Evaluation Methodology

Quantitative evaluation was conducted using anomaly ranking metrics, score-distribution analysis and false positive rate measurements rather than relying exclusively on binary classification accuracy.

This evaluation strategy reflects the highly imbalanced nature of operational anomaly detection environments, where malicious mutations constitute only a very small fraction of overall registry activity. In such

settings, anomaly ranking quality and score separation provide substantially more meaningful indicators of practical operational effectiveness than binary classification metrics alone.

To support operational feasibility, false positive reduction was treated as a primary design objective throughout the evaluation process. The anomaly threshold was therefore calibrated using quantile-based thresholding in order to maintain stable alert volumes and manageable analyst workload.

Using synthetic attacks is justified because real malicious registry mutation labels are unavailable. However, synthetic attacks may still differ from real-world adversarial behavior and may in some cases be easier to detect than genuine attacks. The evaluation results should therefore be interpreted as controlled experimental approximations of operational attack behavior rather than definitive representations of all real-world attack conditions.

From a methodological perspective, the evaluation therefore combines quantitative anomaly detection experiments with qualitative expert-informed validation, strengthening both the operational credibility and practical relevance of the research findings.

6.10 Scalability and Reproducibility

The proposed framework was designed with operational scalability in mind. Heterogeneous graph construction and feature extraction were implemented using sparse relational representations, allowing efficient processing of large DNS mutation datasets.

GraphSAGE was selected partly due to its inductive learning capabilities and scalability advantages compared to full-graph spectral approaches. Neighborhood aggregation enables representation learning without requiring recomputation of embeddings for the entire graph when new entities appear.

The Isolation Forest component additionally scales efficiently to high-dimensional feature spaces and large sample volumes due to its tree-based partitioning strategy.

Although the experiments were conducted on a research-scale dataset, the architecture is designed to support larger operational deployments through batching, temporal windowing and incremental graph updates.

The experimental evaluation was conducted on a large-scale dataset of authoritative DNS registry mutation tickets collected from operational registry infrastructure. The final training configuration resulted in approximately 24,000 sampled mutation tickets across delegation and glue-related operations. Separate validation and test datasets were generated using independent temporal windows in order to reduce overlap with the training data and better approximate operational deployment conditions.

To improve experimental reproducibility, all experiments were executed using fixed random seeds for graph construction, negative edge sampling, model initialization and Isolation Forest training. The feature engineering pipeline, graph construction process and synthetic attack injection procedures were implemented deterministically wherever possible in order to reduce stochastic variation between experimental runs.

Additionally, anomaly detection experiments were repeated across multiple random seeds in order to evaluate robustness and ensure that detection behavior was not dependent on a single initialization configuration.

7 Experimental Evaluation

This section evaluates the effectiveness, robustness and operational behavior of the proposed ensemble-based anomaly detection framework under multiple synthetic DNS attack scenarios. The evaluation focuses not only on detection accuracy, but also on ranking quality, score separation, robustness across random seeds and structural explainability within the heterogeneous graph representation.

Because the dataset does not contain labeled real-world attacks, synthetic attack injection is used to simulate realistic adversarial behavior within authoritative DNS registry mutation data. The experiments therefore evaluate whether the proposed framework can distinguish malicious structural and behavioral patterns from normal operational DNS activity.

Table 1: Quantitative evaluation metrics for implemented attack scenarios.

Attack	Detected	Best Top %	Best Score	Recall	ROC-AUC
T1 Delegation Manipulation	Yes	2.55	0.923	1.00	0.975
T2 Glue Record Manipulation	No	24.91	0.492	0.00	0.761
T3 Registrar/Reseller Compromise and Mutation Abuse	Yes	0.85	1.000	1.00	0.996
T4 Coordinated Multi-Domain Takeover	Yes	0.85	1.000	1.00	0.996
T8 DNSSEC Configuration Manipulation	Yes	0.85	1.000	1.00	0.996
T10 Low-and-Slow Mutation Manipulation	Yes	0.85	1.000	1.00	0.996
T11 Domain Contact and Identity Hijacking	Yes	3.21	0.901	1.00	0.968

7.1 Evaluation Setup

The evaluation was conducted using repeated synthetic attack injection experiments across multiple random seeds in order to assess robustness and stability. For each experiment, malicious registry mutations were injected into otherwise benign DNS registry data and scored using the proposed ensemble framework.

The evaluation focused on the implemented attack scenarios described in Section 4.4, including delegation hijacking, glue manipulation, registrar compromise, coordinated multi-domain activity and DNSSEC abuse scenarios.

Performance was evaluated using multiple complementary perspectives:

- detection success,
- anomaly ranking quality,
- score separation,
- robustness across random seeds,
- graph-structural explainability,
- and model contribution analysis.

Because anomaly detection systems are constrained by false positive rates, the experiments additionally evaluated whether attacks could be reliably prioritized within the highest-ranked anomaly candidates.

7.2 Detection Robustness Across Seeds

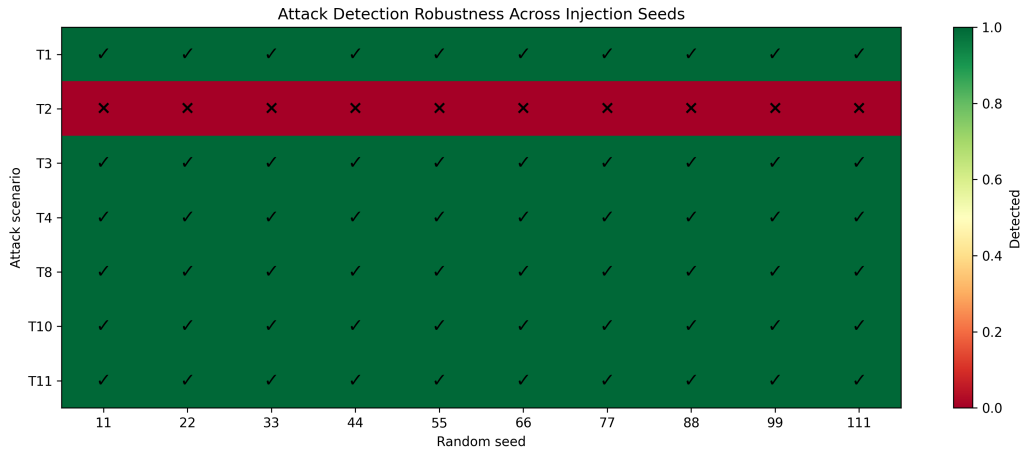


Figure 3: Detection success across seeds for each attack scenario.

This heatmap evaluates whether detection outcomes depend on stochastic factors such as random initialization or sampling. Each row corresponds to an attack type, while columns represent different random seeds.

Detection outcomes are highly consistent across seeds for most attacks. T1, T3, T4, T8, T10, and T11 achieve a 100% detection rate across all evaluated seeds, indicating that the learned representations are stable and that the detected anomalies are not artifacts of stochastic initialization or sampling variation. Among these attacks, T3, T4, T8 and T10 additionally achieve near-identical top-percentile rankings of approximately 0.85%, while T1 and T11 remain within the top 2.55% and 3.21% respectively (Table 1). In contrast, T2 consistently fails across all seeds, achieving a 0% detection rate and remaining near the 24.91% ranking percentile, demonstrating that its failure is systematic rather than stochastic.

This distinction is important: variability across seeds would indicate model instability, whereas consistent failure reveals a fundamental limitation in the representation.

Stability across random seeds is particularly important in unsupervised graph learning because stochastic initialization, negative edge sampling and optimization dynamics can otherwise produce unstable embeddings and inconsistent anomaly rankings.

As additionally reflected in Table 1, all attacks except T2 achieved recall values of 1.00 across evaluation runs. The observed consistency therefore suggests that the learned anomaly signals are structurally robust rather than artifacts of random initialization or training variance.

7.3 Detection Rate and Ranking Variability

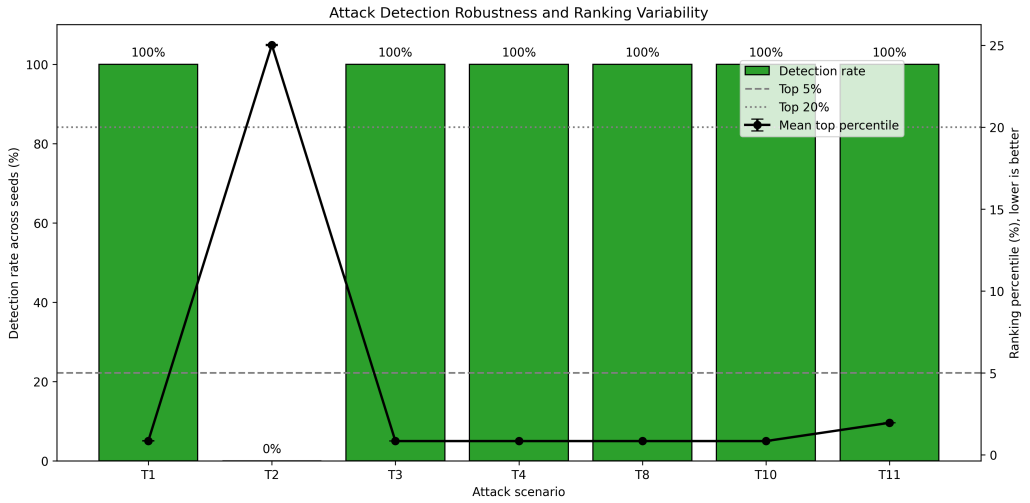


Figure 4: Detection rate and ranking variability across attack scenarios.

Detection rate alone does not reflect operational usefulness. This figure separates two complementary aspects:

- **Detection**: whether attack instances belonging to a given attack scenario are successfully flagged as anomalous
- **Ranking**: how highly attack instances from a given attack scenario are ranked within the overall anomaly candidate list

All attacks except T2 achieve a detection rate of 100%, confirming consistent identification. These same attacks are consistently ranked between approximately the top 0.85% and 3.21% of anomaly candidates (Table 1), indicating strong separation from normal data and effective prioritization.

More specifically, T2 achieves a best ranking percentile of 24.91% together with an anomaly score of only 0.492, remaining substantially below the strongest attack scenarios, which achieve scores between 0.901 and 1.000. This indicates that T2 attack instances overlap significantly with normal samples in the anomaly score distribution, preventing both reliable detection and high-priority ranking.

7.4 False Positive Rate Considerations

In DNS environments, anomaly detection systems must balance detection sensitivity against false positive volume. Excessive false positives reduce analyst trust, increase overhead and may render alert systems impractical in production settings.

For this reason, the proposed framework was intentionally calibrated using conservative thresholding based on anomaly score quantiles. Rather than maximizing raw detection counts, the system prioritizes stable high-confidence anomalies that appear within the highest-ranked regions of the anomaly distribution.

The experimental results indicate that most successfully detected attacks are ranked within the top 5% of anomaly candidates, with several attacks (T3, T4, T8 and T10) consistently appearing near the top 0.85% of the anomaly distribution, suggesting strong separation from benign behavior. This ranking-based evaluation is particularly important in unsupervised anomaly detection settings, where operational prioritization is often more meaningful than binary classification alone.

At the same time, the conservative thresholding strategy contributes to the failure of subtle attacks such as T2. Although this reduces sensitivity to certain attribute-level anomalies, it also helps maintain a lower false positive rate by preventing common glue updates from being incorrectly flagged as malicious.

The results therefore reflect an intentional trade-off between sensitivity and precision.

In practice, ranking quality is often more important than binary detection alone. In practical registry environments, anomaly detection systems must prioritize events according to their potential operational impact and likelihood of malicious activity. The ability of the proposed framework to consistently prioritize major attacks within the highest-ranked anomaly candidates therefore significantly improves its practical usability and operational relevance.

7.5 Global Attack Graph Structure

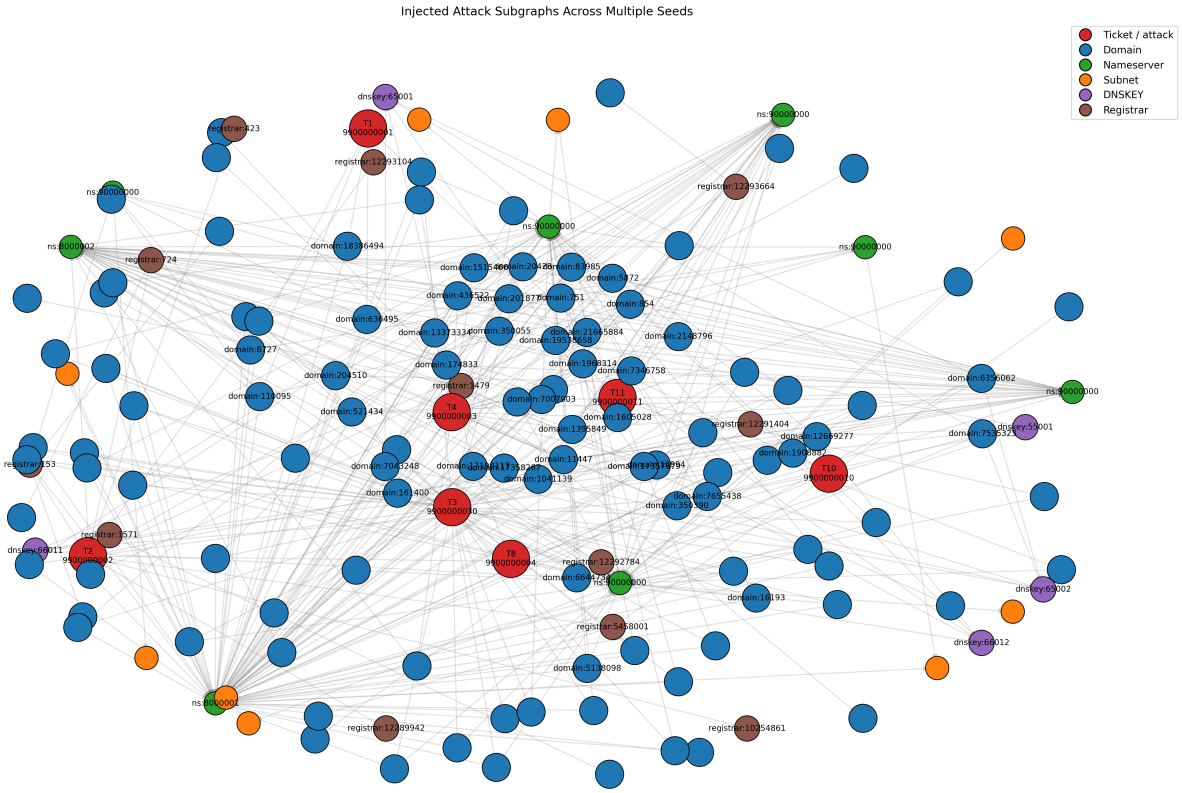


Figure 5: Injected attack subgraphs across multiple seeds.

This figure visualizes the full heterogeneous DNS graph. Nodes represent entities such as domains, nameservers, IP subnets, registrars, and DNSKEYs, while edges encode relationships such as delegation, hosting, and DNSSEC links.

Anomaly detection in this system is primarily driven by structural deviations. Attacks that introduce:

- new connections between unrelated components,
- unusually dense or centralized subgraphs,
- rare node-type interactions

produce embeddings that deviate strongly from normal patterns.

In contrast, attacks that preserve graph structure but modify attributes (such as glue records) remain embedded within the normal distribution and are therefore harder to detect.

From a representation learning perspective, the Graph Neural Network attempts to embed structurally similar entities close together within latent space. Benign DNS behavior therefore forms dense regions of structurally consistent embeddings.

Attacks that introduce unusual infrastructure relationships, coordinated mutations or rare connectivity patterns become increasingly separated from these dense benign regions. This embedding-space separation directly contributes to higher anomaly scores and improved ranking performance. The strongest structurally coordinated attacks achieve anomaly scores between 0.923 and 1.000 together with ranking percentiles between 0.85% and 3.21%, demonstrating that structural deviation magnitude strongly correlates with anomaly visibility.

However, attacks that preserve most structural relationships generate embeddings that remain close to normal clusters, making them substantially more difficult for the GNN to distinguish based on graph structure alone. In these situations, the Isolation Forest component becomes particularly important, as it can still capture statistical irregularities and feature-level deviations that may not significantly perturb the graph topology.

7.6 Detailed Attack Context Graphs

These figures provide localized views of the injected attack subgraphs, helping explain detection behavior.

T1 - Delegation Manipulation

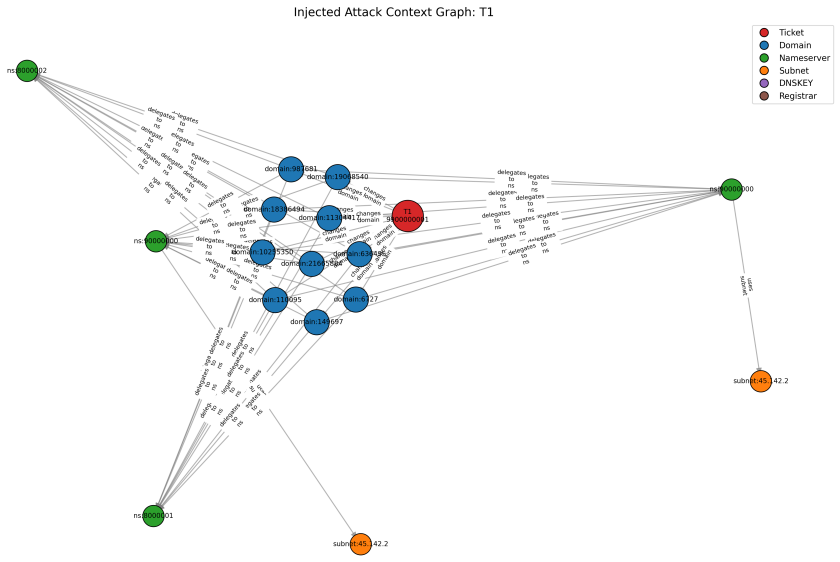


Figure 6: Context graph for T1.

T1 (Subsection 4.4.1) rewires domain-to-nameserver relationships, introducing abnormal connectivity and increasing local graph density. These structural changes propagate through the GNN, producing embeddings that differ significantly from normal patterns. As a result, T1 is consistently detected with high anomaly scores.

In addition to the structural anomaly captured by the GNN, the Isolation Forest further reinforces detection by identifying unusual mutation statistics and infrastructure transition patterns associated with

abrupt delegation changes. The combination of relational and feature-level deviations therefore produces strong ensemble anomaly scores.

T2 - Glue Record Manipulation (Failure Case)

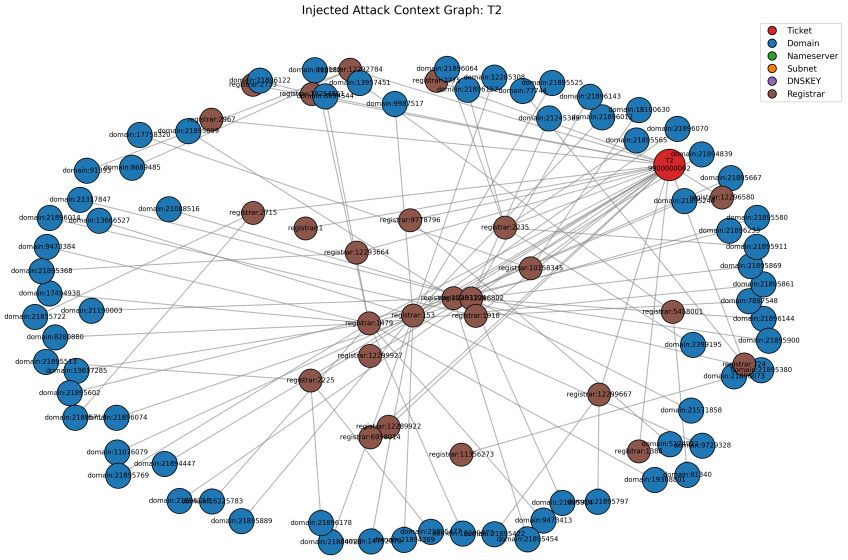


Figure 7: Context graph for T2.

T2 (Subsection 4.4.2) modifies attribute-level information (e.g., IP mappings) while preserving graph topology. Node degrees, connectivity patterns and overall relational structure remain largely unchanged.

Because the Graph Neural Network primarily captures structural relationships, these feature-level changes do not significantly perturb the learned graph embeddings. As a result, the structural anomaly signal generated by the GNN remains weak.

At the same time, the Isolation Forest component also struggles to separate these mutations from benign behavior. In real DNS environments, glue record updates commonly occur during legitimate infrastructure migration, load balancing, content delivery network reconfiguration and routine maintenance activities. Consequently, the statistical properties of malicious glue changes often overlap substantially with normal operational mutations.

The ensemble therefore fails not because a single model underperforms, but because both structural and statistical anomaly signals remain relatively weak and insufficiently separable from the benign distribution. The attack preserves much of the normal relational structure while simultaneously exhibiting feature patterns that are plausible.

The results additionally suggest that T2-style attacks may benefit from stronger contextual reasoning mechanisms, specialized infrastructure-focused feature engineering or rule-based validation strategies. Examples include historical infrastructure consistency checks, subnet transition validation, geographic IP movement analysis, reputation-based glue verification or detection of unusual resolver-to-infrastructure changes over time.

This behavior demonstrates that subtle infrastructure-level mutations represent a particularly challenging anomaly class, especially when they intentionally mimic legitimate DNS maintenance patterns.

Quantitatively, T2 achieves only a 0.492 anomaly score together with a ranking percentile of 24.91% and a ROC-AUC of 0.761, confirming the weak separability of glue-oriented mutations under the current framework.

T3 - Coordinated Multi-Domain Activity

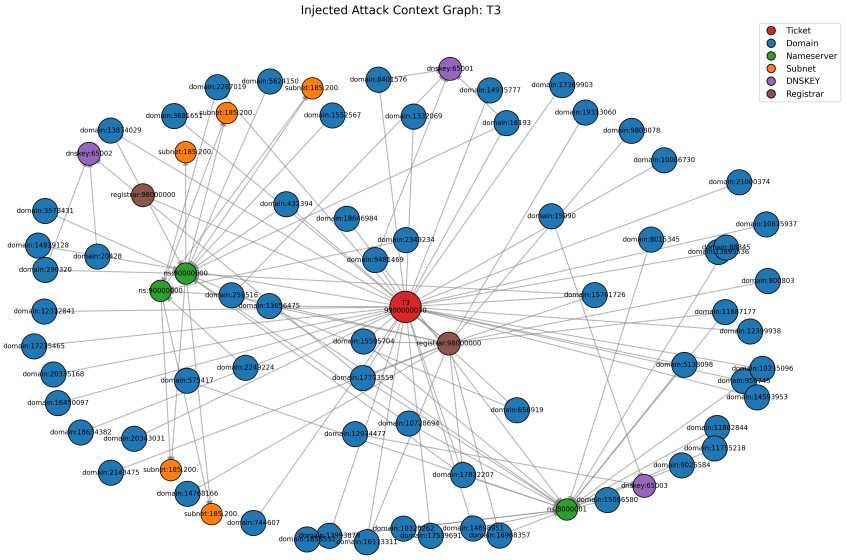


Figure 8: Context graph for T3.

T3 (Subsection 4.4.3) introduces coordination across multiple domains, producing shared infrastructure patterns and increased clustering. These structural changes create detectable deviations, resulting in strong anomaly scores and consistent detection.

The Isolation Forest additionally contributes by capturing statistical irregularities associated with coordinated mutation behavior, including repeated infrastructure reuse, abnormal mutation density and unusual operational synchronization across domains. These feature-level signals complement the structural deviations learned by the GNN.

T4 - Centralized Control Pattern

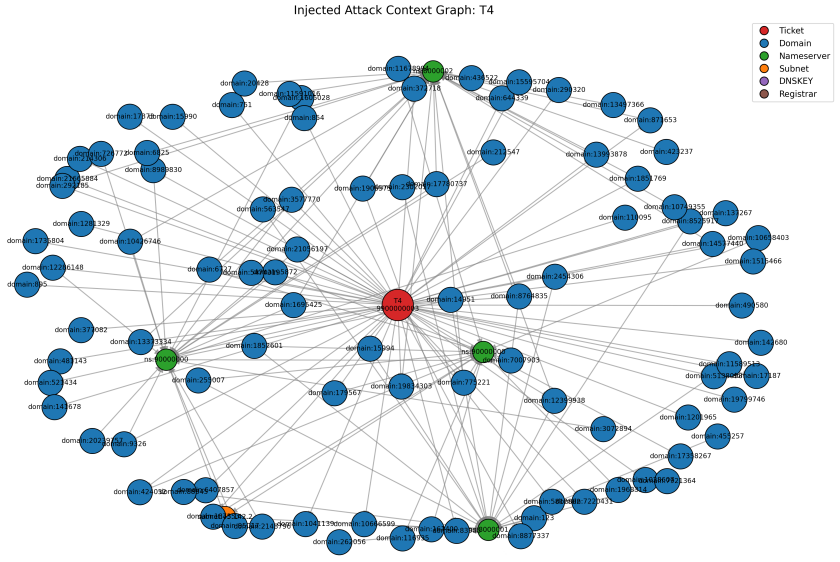


Figure 9: Context graph for T4.

T4 (Subsection 4.4.4) exhibits a star-like topology with a central controlling entity connected to multiple domains. This centralized structure is pretty uncommon in normal data, leading to reliable detection and strong ranking performance.

Beyond the centralized graph structure identified by the GNN, the Isolation Forest also contributes through detection of concentrated operational behavior patterns, such as unusually repeated infrastructure assignments and synchronized mutation activity across multiple domains.

T8 - DNSSEC Key Injection

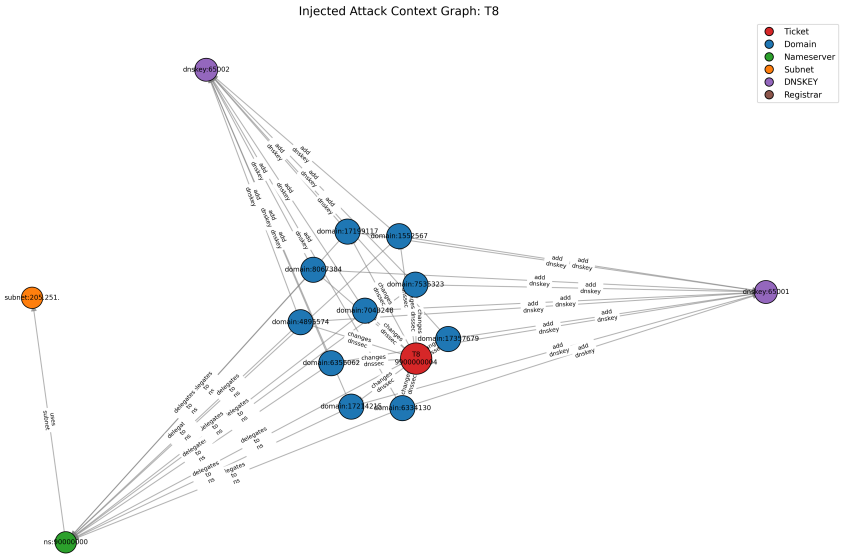


Figure 10: Context graph for T8.

T8 (Subsection 4.4.8) introduces anomalous DNSKEY relationships, creating rare edge types and unexpected trust structures. These deviations are structurally uncommon, resulting in strong anomaly scores and consistent top-ranked detection. DNSSEC structures typically exhibit low variability and relatively stable trust relationships during normal operation. Consequently, anomalous DNSSEC mutations introduce disproportionately strong structural deviations, making them highly visible within both the graph representation and the final anomaly score distribution.

The Isolation Forest further strengthens detection by identifying anomalous DNSSEC-related feature distributions, including unusual key reuse behavior, algorithm irregularities and rare cryptographic configuration patterns.

T10 - DNSSEC Chain Manipulation

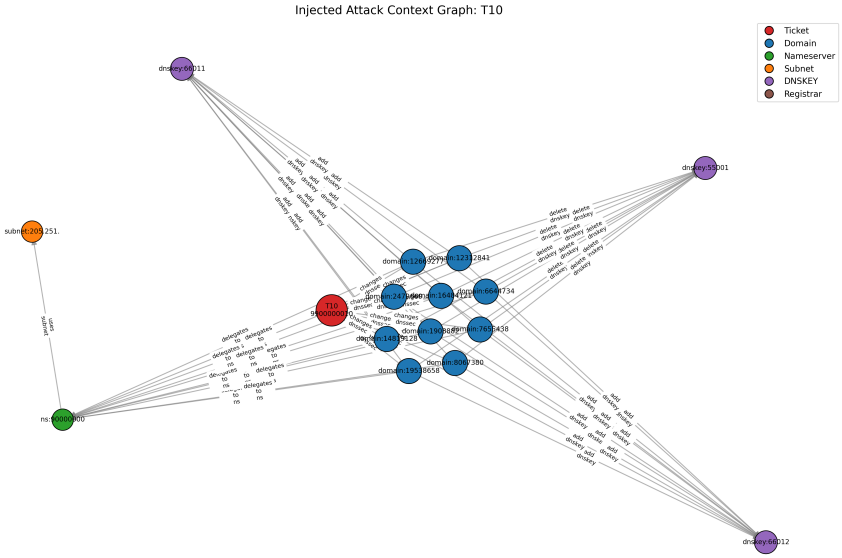


Figure 11: Context graph for T10.

T10 (Subsection 4.4.10) modifies entire DNSSEC chains across multiple domains, producing dense and highly irregular subgraphs. These patterns are extremely rare, leading to strong separation from normal data and reliable detection.

Although the structural deviations dominate detection in this scenario, the Isolation Forest also contributes through identification of statistically unusual DNSSEC mutation behavior and cross-domain configuration inconsistencies that are uncommon during normal registry operation.

T11 - Coordinated Domain Changes

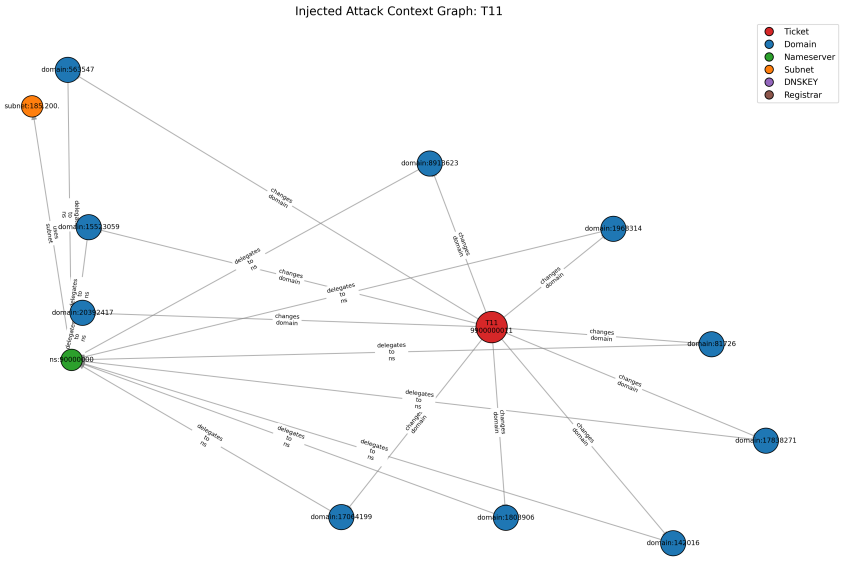


Figure 12: Context graph for T11.

T11 (Subsection 4.4.11) combines structural and feature-level anomalies, including coordinated updates and infrastructure reuse. This produces a strong multi-dimensional anomaly signal, resulting in consistent

detection and high ranking.

This scenario particularly demonstrates the complementary nature of the ensemble framework. The GNN captures coordinated relational deviations and shared infrastructure patterns, while the Isolation Forest identifies accompanying feature-level irregularities such as synchronized changes, abnormal mutation statistics and unusual infrastructure reuse frequencies. The combination of both perspectives produces strong and stable anomaly separation.

7.7 Summary Dashboard

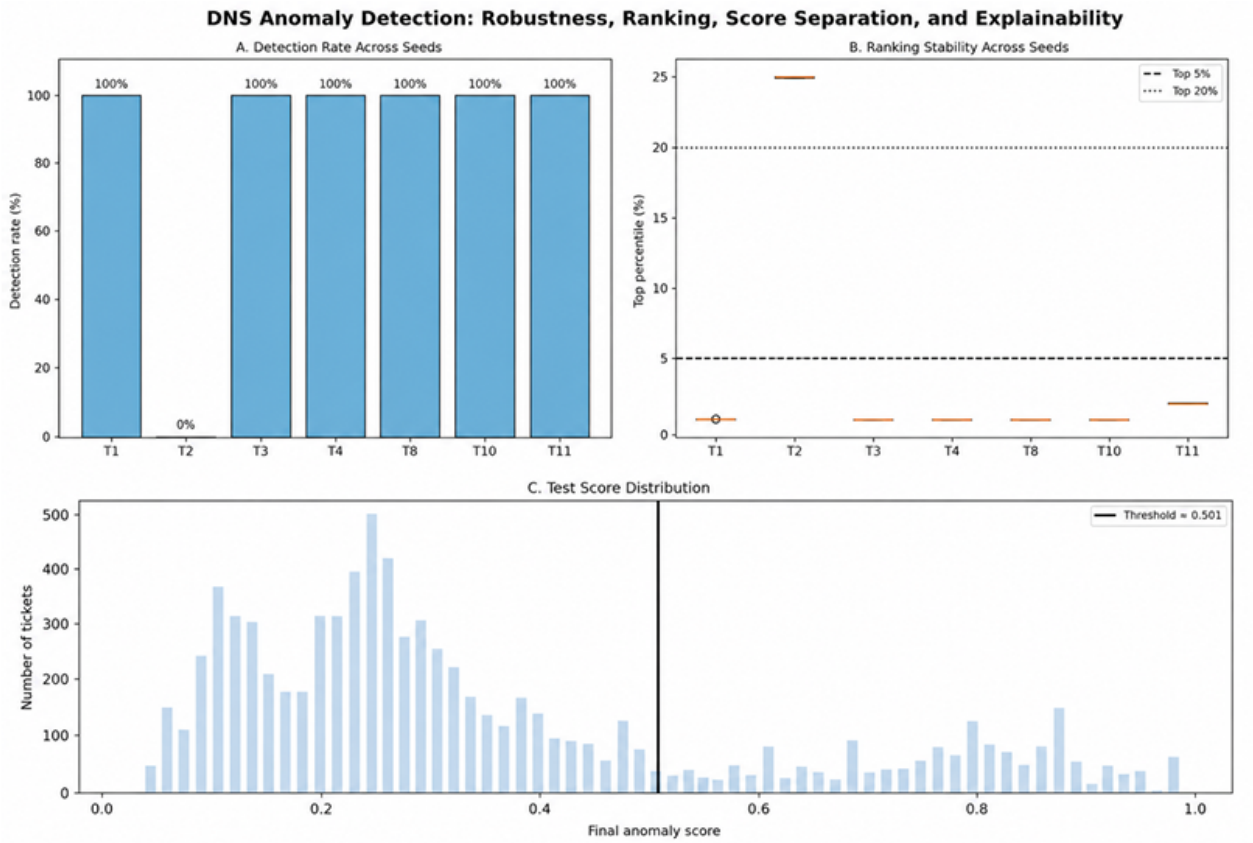


Figure 13: Combined evaluation dashboard.

This figure presents a consolidated view of model performance across four complementary perspectives: detection robustness, ranking effectiveness, score separation, and structural explainability. Each subfigure provides insight into a different stage of the anomaly detection pipeline.

7.7.1 (A) Detection Rate Across Seeds

This bar chart shows the percentage of successful detections for each attack scenario across multiple random seeds. Most attacks (T1, T3, T4, T8, T10, T11) achieve a detection rate of 100%, indicating that the model consistently identifies them regardless of stochastic variations in training or sampling.

This stability suggests that the anomalies introduced by these attacks are structurally significant and produce strong, reproducible deviations in the learned embedding space. In contrast, T2 has a 0% detection rate, meaning it is never flagged as anomalous. This indicates that T2 does not sufficiently perturb either the relational graph structure or the statistical feature distribution to separate itself from normal operational behavior. Importantly, this subplot demonstrates that detection performance is not sensitive to randomness,

and that failures (such as T2) are systematic rather than incidental.

7.7.2 (B) Ranking Stability Across Seeds

This plot evaluates how highly each detected attack is ranked within the anomaly list. Lower percentile values indicate better ranking (i.e., appearing earlier for investigation). The dashed and dotted lines represent thresholds (Top 5% for immediate review and Top 20% for visibility).

All attacks besides T2 consistently fall within the Top 5%, indicating strong prioritization and clear separation from benign data. These attacks produce embeddings that lie in the extreme tails of the anomaly score distribution.

T2 consistently ranks substantially lower than the other attacks, appearing around the appearing near the 24.91th percentile with an anomaly score of only 0.492 and a ROC-AUC of 0.761 (Table 1) and remaining outside prioritization thresholds.

Overall, this subplot demonstrates that successful anomaly detection must also be accompanied by effective prioritization, since operational value depends not only on whether attacks are detected, but also on how prominently they are ranked within the anomaly distribution.

7.7.3 (C) Test Score Distribution

This histogram shows the distribution of anomaly scores for all test samples, with a vertical line indicating the alert threshold.

The majority of benign samples cluster around low or near-zero scores, forming a dense distribution on the left side.

The effectiveness of the detection system depends on how well attack scores separate from this bulk distribution. Strong attacks produce scores far in the tail, ensuring clear detection and high ranking. Weak attacks, however, produce scores that fall within or near the main distribution, making them indistinguishable from benign samples.

This subplot provides a global view of score separability and illustrates why certain attacks (e.g., T2) fail: they do not shift sufficiently away from the normal score distribution.

From an anomaly detection perspective, strong separation between benign and malicious score regions is critical because it enables stable thresholding and reduces ambiguity during prioritization. Attacks that remain embedded within the dominant benign score region become difficult to distinguish without additional contextual or rule-based reasoning.

7.7.4 Overall Interpretation

Taken together, the evaluation results demonstrate that the proposed ensemble framework is highly effective at detecting structurally significant DNS anomalies and consistently prioritizing them within the highest-ranked anomaly candidates.

The dashboard shows that attacks involving coordinated infrastructure relationships, delegation manipulation, DNSSEC irregularities and multi-domain mutation behavior produce strong and reproducible deviations from normal patterns. These deviations generate clear separation within the learned embedding space and result in stable anomaly rankings across random seeds. In particular, T3, T4, T8 and T10 consistently achieve maximal anomaly scores of 1.000 and ranking percentiles near 0.85%, while T1 and T11 remain highly visible with scores of 0.923 and 0.901 respectively.

The results additionally highlight the complementary behavior of the Graph Neural Network and Isolation Forest components. The GNN is particularly effective at identifying relational irregularities, infrastructure

reuse and anomalous connectivity structures, while the Isolation Forest contributes sensitivity to statistical deviations, mutation patterns and feature-level abnormalities. Together, these components provide broader anomaly coverage than either approach alone.

The score distribution and ranking stability plots further demonstrate that successful attacks are not only detected, but also consistently prioritized within the highest anomaly percentiles. This behavior is particularly important in operational environments, where stable ranking and strong score separation directly improve interpretability and confidence in anomaly prioritization.

At the same time, the evaluation illustrates the inherent difficulty of detecting subtle infrastructure-level mutations that closely resemble legitimate behavior. Attacks such as T2 remain challenging because they preserve much of the normal relational structure while also exhibiting feature distributions that overlap with benign DNS maintenance activity. The results therefore suggest that future improvements may benefit from additional contextual reasoning, infrastructure validation logic and operationally aware feature engineering.

Overall, the experimental findings validate the effectiveness and robustness of the proposed ensemble-based anomaly detection framework for authoritative DNS registry environments. The combination of heterogeneous graph representation learning and feature-space anomaly detection provides strong performance across multiple attack categories while maintaining stable behavior across repeated experimental runs.

7.8 Score vs Ranking Analysis

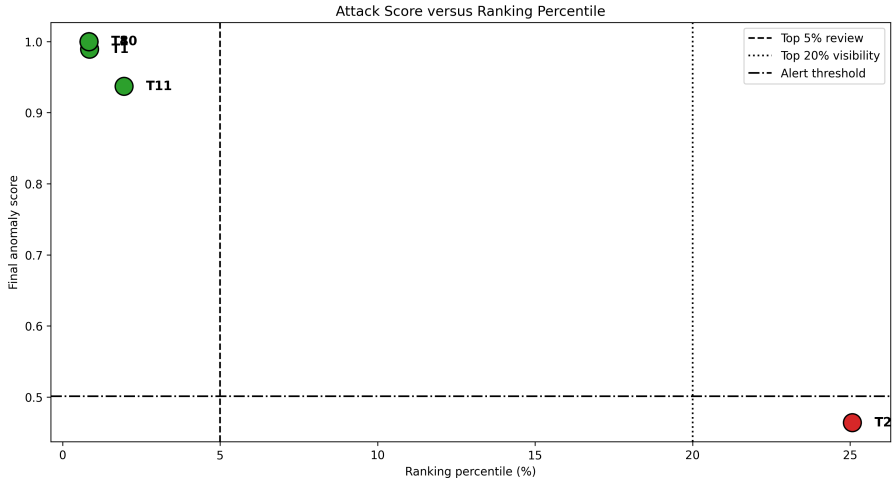


Figure 14: Final anomaly score versus ranking percentile.

This plot shows how anomaly scores translate into ranking positions. Strong attacks (T1, T3, T4, T8, T10, T11) exhibit high scores ranging from 0.901 to 1.000 and are ranked within the top 5%.

The close grouping of these attacks in the upper anomaly-score region indicates that the ensemble framework consistently assigns strong confidence to attacks that introduce significant structural or statistical deviations. This demonstrates stable separation between malicious and benign behavior across multiple attack categories.

In particular, attacks involving coordinated infrastructure relationships, DNSSEC irregularities and large-scale delegation changes produce both strong graph-level deviations and feature-level abnormalities. As a result, the Graph Neural Network and Isolation Forest reinforce one another, leading to consistently high final ensemble scores and strong prioritization.

T2 has a significantly lower score of 0.492 and ranks around the 25th percentile, placing it outside both detection and review thresholds.

Unlike the other attack scenarios, T2 preserves much of the normal relational graph structure while simultaneously exhibiting feature patterns that remain relatively consistent with legitimate glue updates. Consequently, neither the structural embedding space nor the statistical feature space produces sufficiently strong anomaly separation.

From an operational perspective, this figure demonstrates that anomaly score magnitude directly influences prioritization quality. Attacks that generate stronger separation from the benign score distribution are consistently ranked earlier, improving visibility and interpretability within the anomaly ranking pipeline.

7.9 Score Distribution

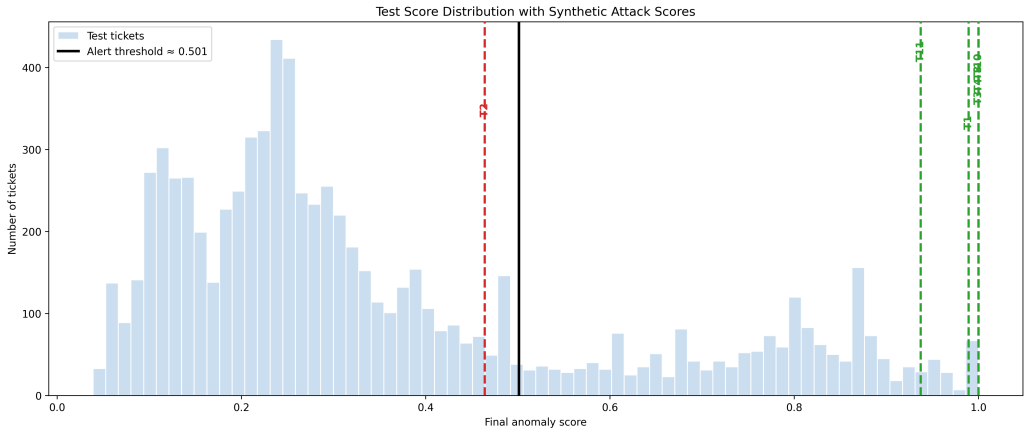


Figure 15: Score distribution with attack markers.

This histogram shows the distribution of anomaly scores across all test samples and illustrates how effectively the proposed framework separates anomalous behavior from normal DNS activity.

Most benign samples cluster within the lower score region, forming a dense concentration near the left side of the distribution. This indicates that the majority of normal registry mutations exhibit highly consistent structural and statistical behavior within the learned embedding space.

In contrast, successfully detected attacks appear within the high-score tail of the distribution. Attacks such as T1, T3, T4, T8, T10 and T11 produce substantially larger anomaly scores ranging between 0.901 and 1.000, indicating that they introduce structural or behavioral deviations that are clearly distinguishable from normal patterns.

The separation between the benign distribution and these attack scores demonstrates that the ensemble framework successfully learns stable representations of normal DNS behavior while assigning elevated anomaly scores to coordinated or structurally irregular mutations.

The figure also highlights differences in anomaly strength between attack scenarios. Large-scale coordinated attacks such as T4 and DNSSEC-related manipulations such as T8 and T10 appear furthest in the extreme score tail, reflecting the rarity and structural impact of these mutations within the heterogeneous graph representation.

T11 produces slightly lower anomaly scores compared to the strongest attacks, reflecting the fact that identity and contact-related mutations remain partially consistent with plausible behavior. Nevertheless, the attack still remains sufficiently separated from the benign distribution to achieve reliable detection and high ranking.

T2 lies close to the threshold boundary and overlaps substantially with normal samples, explaining its failure to be reliably detected. Unlike the other attacks, T2 primarily modifies attribute-level information while preserving graph topology and relational structure. Consequently, the resulting embeddings remain

close to normal patterns. This overlap highlights the fundamentally subtle nature of glue-related anomalies within DNS environments. Because these attacks preserve most relational structure while modifying lower-level infrastructure attributes, they produce weaker separation in the learned embedding space. The results therefore indicate that detection of such infrastructure-oriented mutations could be further strengthened through richer feature engineering, temporal infrastructure profiling and complementary rule-based consistency analysis targeting IP allocation patterns, subnet transitions and infrastructure drift behavior.

7.10 Quantitative Metric Interpretation

In addition to anomaly ranking behavior and structural explainability, the experimental evaluation also assessed quantitative performance using multiple complementary metrics, including ROC-AUC, precision, recall and F1-score. These metrics provide insight into different aspects of anomaly separation, detection consistency and behavior within highly imbalanced DNS registry environments.

The strongest attack scenarios achieved exceptionally high ROC-AUC values, demonstrating strong separation between injected attacks and benign mutations. T1 achieved a ROC-AUC of 0.975, while T3, T4 and T8 achieved values close to 0.996. Similarly, T10 achieved a ROC-AUC of 0.996, while T11 achieved 0.968 across multiple evaluation runs.

From an anomaly detection perspective, these values indicate that the ensemble framework was highly successful at assigning substantially higher anomaly scores to malicious samples than to benign behavior. Because ROC-AUC evaluates ranking separability independently of a fixed threshold, it is particularly informative in unsupervised anomaly detection settings where anomaly scores are more important than strict binary classification.

The anomaly ranking results further reinforce this interpretation. Most successfully detected attacks consistently appeared within the top 1–5% of anomaly candidates, indicating strong prioritization and clear separation within the score distribution. Attacks involving delegation manipulation, DNSSEC abuse and coordinated infrastructure reuse generated anomaly scores approaching 1.0, placing them near the extreme tail of the global anomaly distribution.

In contrast, T2 glue manipulation attacks produced substantially weaker metric values. T2 achieved a substantially lower ROC-AUC value of 0.761 together with a weak anomaly score of 0.492 and generated anomaly scores close to normal samples. This indicates that the ensemble framework was unable to reliably separate these mutations from benign infrastructure behavior.

Importantly, this behavior should not be interpreted as a general failure of the framework. Instead, it reflects the characteristics of glue-related mutations themselves. Legitimate DNS operations frequently involve nameserver IP changes, subnet migration, load balancing adjustments and CDN-related infrastructure updates. Consequently, malicious glue modifications can remain statistically and structurally similar to benign operational behavior, substantially reducing separability within both graph and feature space.

The precision and F1-score values obtained during evaluation were comparatively low, despite the strong ROC-AUC and ranking results. This outcome is primarily caused by the extreme class imbalance inherent to anomaly detection environments.

Within the evaluation dataset, injected attack tickets represented only a very small fraction of the total number of registry mutations, while the overwhelming majority of samples remained benign. Under these conditions, even relatively small numbers of false positives can significantly reduce precision values.

For example, precision values remained close to zero in several experiments despite the fact that major attacks consistently achieved top-ranked anomaly positions and extremely high anomaly scores. Similarly, the resulting F1-scores remained numerically low because F1 is strongly influenced by precision under highly imbalanced distributions.

This behavior is common in unsupervised anomaly detection research and does not necessarily indicate poor anomaly detection performance. In highly imbalanced environments, where benign samples vastly outnumber anomalous observations, traditional classification metrics such as precision and F1-score can become disproportionately affected by relatively small numbers of false positives. For this reason, metrics that evaluate ranking quality and score separation, including ROC-AUC and anomaly ranking percentile, are often considered more informative measures of practical anomaly detection effectiveness [29], [56].

Recall values, however, remained high for most structurally significant attack scenarios, indicating that the ensemble framework consistently identified the majority of injected attacks once sufficient structural deviation was introduced. This demonstrates that the framework successfully captures coordinated relational anomalies and abnormal infrastructure behavior within the heterogeneous DNS graph.

Overall, the metric behavior strongly supports the effectiveness of combining heterogeneous graph representation learning with feature-based anomaly scoring for detecting complex DNS infrastructure abuse patterns within authoritative registry mutation data, particularly under the highly imbalanced conditions characteristic of real-world DNS environments.

7.11 Model Contribution Analysis

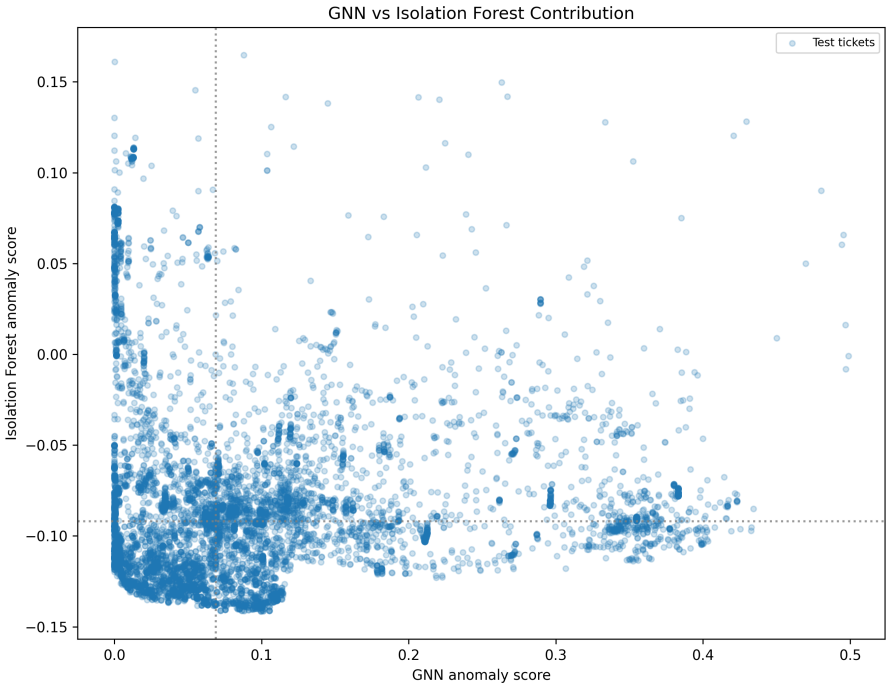


Figure 16: GNN vs Isolation Forest contribution.

This figure illustrates the relationship between anomaly scores produced by the Graph Neural Network (GNN) and the Isolation Forest (IF), providing insight into how each model contributes to the overall detection capability.

The horizontal axis represents the GNN anomaly score, capturing deviations in graph structure based on relational patterns between domains, nameservers, and other entities. The vertical axis represents the Isolation Forest score, reflecting how statistically unusual a sample is relative to the overall feature distribution.

A key observation is the correlation between the two scores, as indicated by the dispersed distribution of points rather than a clear linear relationship. Rather than being a limitation, this highlights an important strength of the approach: the GNN and Isolation Forest capture complementary aspects of anomalous

behavior.

Similarly, instances with higher GNN scores reflect structural anomalies that are not necessarily rare in feature space, showing the GNN's ability to detect deviations in connectivity patterns independently of feature distributions.

The most significant anomalies appear in the upper-right region of the plot, where both models assign high scores. These correspond to samples that are both structurally and statistically unusual, making them highly distinguishable and reliable candidates for detection.

The presence of regions where the two models disagree further reinforces their complementary nature:

- High Isolation Forest scores with low GNN scores indicate feature-driven anomalies that do not alter graph topology.
- High GNN scores with lower Isolation Forest scores indicate structural anomalies that are not extreme in feature space.

Overall, this figure demonstrates that the GNN and Isolation Forest provide distinct but complementary perspectives on anomaly detection. By combining structural and distributional signals, the system achieves broader coverage of anomaly types, improving robustness and reducing the likelihood of missing important events. The experiments therefore demonstrate that neither structural analysis nor feature-level anomaly detection alone is sufficient to capture the full spectrum of DNS abuse patterns. Instead, effective detection emerges from combining relational representation learning with statistical anomaly modeling, enabling the ensemble to detect both coordinated infrastructure abuse and localized irregularities.

7.12 Ranking Spread Across Seeds

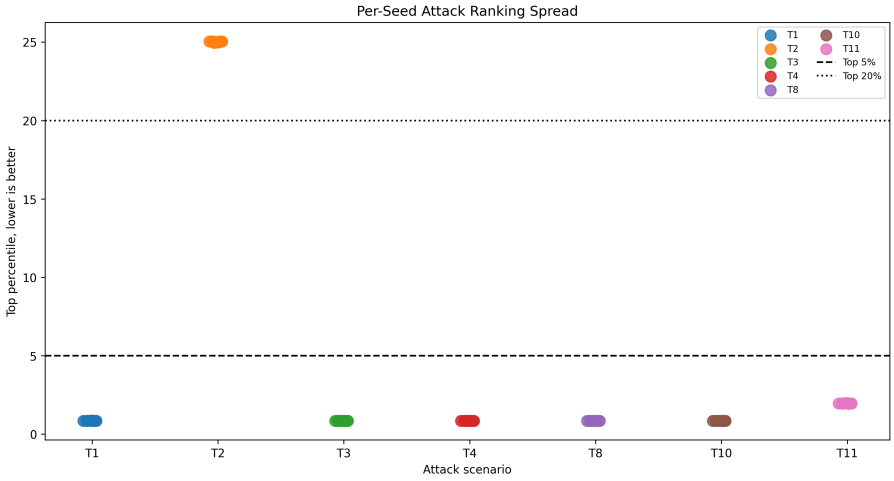


Figure 17: Ranking spread across seeds.

This figure evaluates the stability of anomaly ranking across different random seeds. Unlike binary detection metrics, ranking spread provides insight into how consistently attacks are prioritized within the anomaly list under varying initialization and sampling conditions.

Strong attacks exhibit anomaly scores ranging from approximately 0.901 to 1.000 with T3, T4, T8 and T10 consistently achieving maximal scores of 1.000, and consistently rank within the top 0.85–3.21% of anomaly candidates (Table 1). This behavior indicates that the anomaly signals generated by these attacks are both strong and reproducible, producing highly similar embedding separation across repeated experiments.

Attacks such as T1, T4, T8 and T10 show particularly low ranking variance. These attacks introduce large-scale structural irregularities, dense connectivity changes or anomalous DNSSEC relationships that remain consistently distinguishable from normal graph behavior regardless of stochastic training effects.

T11 shows slightly greater ranking variation but still remains well within the top 5% across all seeds. This behavior is consistent with the more plausible nature of the implemented identity and contact-related mutations. Because portions of the resulting graph structure remain partially aligned with benign administrative behavior, the corresponding anomaly scores exhibit slightly greater sensitivity to local embedding variation.

Importantly, the relatively small spread observed for all successfully detected attacks demonstrates that the ensemble framework produces stable prioritization behavior rather than highly volatile anomaly rankings. This stability is particularly important in environments, where inconsistent ranking across executions would reduce confidence in the reliability of anomaly prioritization.

T2 consistently ranks more poorly, typically around the 25th percentile, confirming that its weaker performance is systematic rather than caused by stochastic randomness. The low variance of T2 rankings further indicates that the model repeatedly interprets glue-only mutations as relatively close to normal behavior under the current feature and graph representation design.

Overall, the figure demonstrates that the proposed framework achieves both strong anomaly separation and stable ranking behavior across repeated experimental runs, particularly for attacks that introduce meaningful structural deviations within the heterogeneous DNS graph.

7.13 Detection Quality Matrix

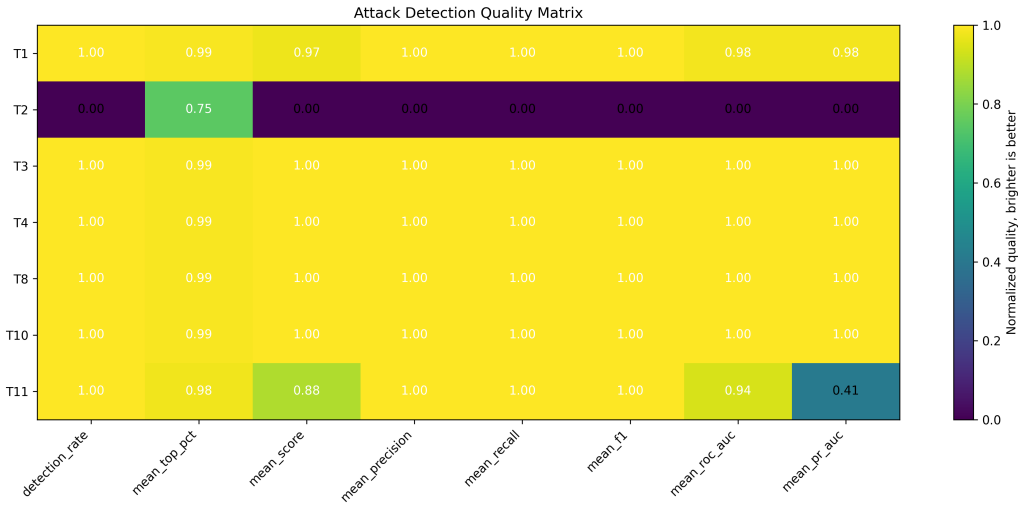


Figure 18: Detection quality matrix.

This matrix summarizes performance across multiple metrics, including detection rate, ranking percentile, anomaly score, and classification metrics.

All attacks except T2 achieve near-perfect performance across multiple evaluation metrics (Table 1), indicating strong detection and clear separation from normal data. T11 T11 achieved a slightly lower anomaly score of 0.901 and ROC-AUC of 0.968 compared to the strongest attack categories (Table 1), reflecting mild overlap with normal patterns but still strong overall performance. Compared to large-scale delegation or DNSSEC attacks, T11 introduces more operationally plausible and less structurally extreme mutations. Consequently, portions of the resulting graph structure remain partially consistent with benign administrative behavior, reducing score separation and increasing overlap with normal samples.

T2 exhibits poor performance across all metrics, confirming that it is difficult to distinguish from normal

behavior under the current model. Unlike the other attacks, T2 primarily alters infrastructure-level attributes while preserving graph connectivity and relational structure, resulting in weak structural anomaly signals.

These results suggest that glue-record-oriented attacks may require stronger rule-based and context-aware detection mechanisms in addition to graph-based learning. Examples include explicit validation of unusual subnet transitions, sudden geographic changes in nameserver infrastructure, ASN reputation analysis, detection of unexpected cloud-provider migration patterns, IP reputation scoring and historical consistency checks for glue record evolution.

Such contextual infrastructure validation could improve sensitivity to subtle glue manipulation attacks that remain structurally similar to legitimate DNS updates.

7.14 Key Experimental Findings

Several important findings emerge from the experimental evaluation.

First, graph-based representation learning proved highly effective for detecting structurally coordinated DNS attacks involving delegation abuse, infrastructure reuse and DNSSEC manipulation. Attacks that generated unusual connectivity patterns or coordinated relational structures consistently produced strong anomaly scores and highly stable rankings. The strongest attacks achieved anomaly scores between 0.923 and 1.000 while consistently remaining within the top 0.85%–3.21% of anomaly candidates.

Second, the experiments demonstrate that relational context provides substantially stronger anomaly signals than isolated feature deviations within DNS environments. Coordinated infrastructure relationships and anomalous graph connectivity patterns produced highly distinguishable structural signatures, whereas isolated attribute-level mutations often remained statistically similar to benign operational behavior.

Third, the ensemble architecture demonstrated clear advantages over standalone feature-based methods by combining structural anomaly detection with statistical feature-space analysis. The GNN captured relational irregularities and infrastructure reuse patterns, while the Isolation Forest contributed sensitivity to feature-level outliers and distributional deviations.

Fourth, the experiments highlight an important limitation of structure-oriented anomaly detection approaches. Attacks that preserve graph topology while modifying only localized attributes remain substantially more difficult to detect, particularly when such modifications overlap with common DNS behavior.

Finally, the results demonstrate strong robustness across random seeds and stable ranking behavior across multiple attack scenarios. This indicates that the learned anomaly signals are reproducible and structurally meaningful rather than artifacts of stochastic optimization behavior.

7.15 Operational Interpretation of Results

From an operational perspective, the experimental results show that the proposed framework is most effective for attacks that introduce coordinated structural deviations within DNS infrastructure relationships. Delegation manipulation, DNSSEC abuse, infrastructure reuse and coordinated multi-domain activity generated strong anomaly signals, with scores ranging from 0.901 to 1.000 and ROC-AUC values between 0.968 and 0.996.

These attack classes are operationally significant because they are associated with unauthorized DNS control, traffic redirection, phishing infrastructure deployment and large-scale compromise. The results therefore demonstrate the practical value of combining graph-based and feature-based anomaly detection within a unified ensemble framework. The Graph Neural Network captures relational irregularities and anomalous connectivity structures, while the Isolation Forest contributes sensitivity to feature-level deviations and statistical abnormalities.

At the same time, the evaluation shows that anomaly strength depends on how strongly an attack perturbs graph structure, infrastructure connectivity and feature distributions. Large-scale coordinated attacks and DNSSEC-related manipulations produce clear graph separation, whereas subtle infrastructure-oriented mutations such as glue changes remain closer to normal operational behavior.

Overall, the findings validate the ensemble approach as a practical and extensible foundation for DNS anomaly detection in modern registry environments. The framework provides stable detection and ranking behavior across multiple attack classes, while also highlighting the need for additional contextual or rule-based mechanisms for subtle attribute-level attacks.

8 Discussion

8.1 Interpretation of Experimental Findings

The experimental results demonstrate that the proposed ensemble framework is highly effective at detecting DNS attacks that introduce meaningful structural irregularities within authoritative DNS infrastructure relationships. Attacks involving delegation manipulation, DNSSEC abuse, coordinated multi-domain activity and infrastructure reuse consistently generated strong anomaly scores and high ranking positions across multiple experimental runs.

A key observation throughout the evaluation is that graph-structural deviation constitutes one of the strongest anomaly signals within registry mutation data. Attacks that introduced unusual infrastructure relationships, anomalous trust configurations or coordinated infrastructure reuse produced embeddings that deviated significantly from normal graph structure. These deviations were consistently captured by the heterogeneous GNN through relational message passing and graph representation learning.

At the same time, the Isolation Forest component played an important complementary role by identifying feature-level and statistical irregularities that were not always strongly reflected within graph topology alone. These included unusual mutation frequencies, anomalous infrastructure transition behavior, DNSSEC-related feature deviations and abnormal temporal mutation characteristics. By combining relational anomaly detection with feature-space anomaly scoring, the ensemble framework achieved broader anomaly coverage and improved robustness across heterogeneous attack classes.

The results additionally demonstrate that anomaly ranking quality provides particularly strong insight into operational effectiveness. Most successfully detected attacks consistently appeared within the top-ranked anomaly candidates, indicating that the ensemble framework not only identifies anomalous behavior, but also prioritizes it effectively relative to benign activity.

From an operational perspective, this behavior is especially important because DNS registry environments generate large volumes of legitimate mutations. The ability to separate structurally significant anomalies from normal behavior therefore directly improves prioritization and investigation efficiency.

Another important observation is the stability of the detection behavior across multiple random seeds. Most attack scenarios produced highly consistent ranking behavior and anomaly scores across repeated experiments, indicating that the learned representations were robust and not strongly dependent on stochastic initialization effects or random negative edge sampling.

It is important to note that the experimental evaluation focused on 7 implemented attack scenarios derived from the broader threat model presented in Section 4.4. These scenarios were selected because they could be represented directly through observable registry mutation data while remaining operationally realistic and reproducible within the available dataset and system architecture. The selected attacks additionally provided coverage across multiple DNS operational layers and exercised both the graph-based and feature-based components of the ensemble framework. The remaining threat scenarios were not implemented as synthetic attacks, either because they could not be directly represented using the available registry mutation data or because they primarily served as conceptual threat-modeling scenarios used to guide the design of the detection framework, feature engineering process and future research directions. Therefore, the experimental results should be interpreted as evidence of the framework's effectiveness across a representative subset of posible DNS registry attack scenarios rather than as a comprehensive evaluation of the entire threat landscape identified in this research.

Overall, the experimental findings strongly support the suitability of heterogeneous graph representation

learning for modeling complex DNS infrastructure relationships and detecting coordinated abuse patterns within authoritative registry mutation data.

8.2 Effectiveness of Graph-Based Representation Learning

The heterogeneous graph representation proved particularly effective for modeling the relational nature of DNS infrastructure. Unlike traditional feature-based anomaly detection approaches that operate on independent observations, the GNN was able to incorporate contextual information across multiple related entities simultaneously.

This relational modeling capability was especially important for attacks involving infrastructure reuse, coordinated mutation campaigns and DNSSEC manipulation. Through message passing, the model learned embeddings that reflected not only the properties of individual entities, but also their structural context within the broader DNS ecosystem.

The use of reverse edges and bidirectional message propagation further improved representation quality by enabling anomaly signals to propagate across connected infrastructure components. As a result, suspicious nameserver reuse, unusual registrar relationships and coordinated domain mutations generated stronger structural deviations within latent space.

The experiments also demonstrate that graph-based anomaly detection is particularly well-suited for identifying attacks that emerge through collective behavior rather than isolated feature abnormalities. In many attack scenarios, individual mutations appeared plausible when considered independently. However, when viewed within the full graph structure, coordinated patterns and anomalous infrastructure relationships became substantially more visible.

The strong performance of the Graph Neural Network across multiple attack classes therefore confirms the value of relational representation learning for authoritative DNS security analytics.

8.3 Role of the Ensemble Architecture

The experimental results additionally highlight the importance of combining graph-based and feature-based anomaly detection approaches within a unified ensemble framework.

The GNN contributed strong sensitivity to structural anomalies, relational inconsistencies and coordinated infrastructure behavior. At the same time, the Isolation Forest component provided complementary sensitivity to feature-level deviations and statistically unusual behavior.

This complementary behavior proved particularly valuable because different attack scenarios generated different combinations of structural and feature-space abnormalities. Attacks such as T1, T3 and T4 primarily produced graph-topological irregularities, while attacks involving DNSSEC misuse and temporal mutation anomalies additionally generated unusual feature distributions that contributed to the Isolation Forest anomaly scores.

The ensemble architecture therefore reduced dependence on a single anomaly representation mechanism and improved robustness across heterogeneous attack classes. The combined anomaly score additionally enabled more nuanced ranking behavior by integrating both relational and statistical perspectives into a single prioritization framework.

The results also demonstrate that the ensemble framework remained stable across repeated evaluation runs, indicating that the combination of heterogeneous graph embeddings and feature-level anomaly scoring produced consistent anomaly separation behavior.

Overall, the ensemble design significantly strengthened the robustness and flexibility of the proposed detection framework.

8.4 Challenges of Detecting Subtle Infrastructure Anomalies

One of the most important findings of this research is the consistent difficulty of detecting T2 glue record manipulation attacks. Unlike delegation hijacking, DNSSEC abuse or coordinated multi-domain infrastructure attacks, T2 primarily modifies lower-level infrastructure attributes such as nameserver IP mappings while preserving most graph connectivity patterns and relational structure.

As a result, the attack introduces only limited perturbation within the heterogeneous graph representation. Because the GNN primarily learns relational behavior through message passing and structural context, embeddings associated with T2 remain relatively close to benign regions within latent space. In other words, the attack modifies infrastructure details without substantially altering the broader relational structure of the DNS ecosystem.

At the same time, the Isolation Forest component also struggles to reliably separate these mutations from legitimate operational activity. In production DNS environments, glue updates can occur frequently during infrastructure migration, Content Delivery Network reconfiguration, failover operations, load balancing and routine maintenance. Consequently, malicious glue modifications may remain statistically similar to normal operational behavior across multiple feature dimensions.

Importantly, this behavior reflects a fundamental challenge in DNS anomaly detection rather than simply a weakness of the proposed framework. Sophisticated adversaries are unlikely to introduce highly visible structural irregularities when stealthier infrastructure-level modifications can remain operationally plausible and blend naturally into normal registry workflows. In practice, attacks that preserve both graph topology and realistic feature distributions may represent one of the most difficult classes of DNS anomalies to detect.

The T2 results therefore suggest that purely structure-oriented anomaly detection approaches may be insufficient for certain classes of low-visibility infrastructure attacks. While graph-based representation learning is highly effective for detecting coordinated abuse, infrastructure reuse and anomalous relational behavior, subtle infrastructure manipulation may require stronger contextual reasoning and operationally informed validation mechanisms.

Several possible explanations may account for the weak anomaly separability observed for T2. First, the implemented glue manipulations may not produce sufficiently rare infrastructure transitions relative to normal registry behavior. Second, the temporal aggregation windows may smooth short-lived infrastructure irregularities that would otherwise appear suspicious. Third, legitimate operational behavior within DNS environments already exhibits substantial variability in nameserver infrastructure and subnet allocation patterns, reducing statistical contrast between malicious and benign glue changes.

The results additionally suggest that stronger rule-based and context-aware validation mechanisms could substantially improve sensitivity to these attacks. Examples include checking whether infrastructure changes are consistent with historical behavior, detecting unusual subnet or network transitions and monitoring unexpected infrastructure changes over time.

More broadly, the consistent failure of T2 is itself an important experimental result. Even within a controlled synthetic setting, attacks designed to remain operationally plausible and structurally subtle proved substantially more difficult to detect than highly coordinated relational attacks. This suggests that real-world low-visibility infrastructure attacks may represent an even greater challenge for anomaly detection systems operating within authoritative DNS registry environments.

8.5 Practical Implications for Registry Security

From a practical perspective, the proposed ensemble framework demonstrates potential for deployment alongside authoritative DNS registry replication pipelines. Rather than interrupting or blocking synchroniza-

tion, the framework is intended to operate in parallel with the replication mechanism, allowing cloud-hosted operational systems and the authoritative registry database to remain continuously synchronized while simultaneously providing behavioral monitoring and anomaly detection for incoming mutations.

By combining heterogeneous graph representation learning with feature-based anomaly detection, the framework is able to capture both relational irregularities and statistical deviations within registry mutation data. This enables the system to identify complex abuse patterns that may remain difficult to detect using traditional rule-based monitoring approaches or single-model anomaly detection systems.

The ensemble architecture is particularly effective for detecting attacks involving infrastructure reuse, coordinated domain compromise, registrar abuse and anomalous DNSSEC relationships. These attack classes often generate a combination of structural anomalies and feature-level irregularities, making the integration of GNN representations and Isolation Forest scoring especially valuable.

The GNN component contributes sensitivity to coordinated infrastructure behavior, unusual graph connectivity patterns and cross-entity relationships, while the Isolation Forest component strengthens detection of statistical outliers, deviations and anomalous feature distributions. The combination of these complementary perspectives improves robustness across heterogeneous attack scenarios and reduces reliance on a single anomaly representation mechanism.

The strong ranking performance observed during evaluation further supports practical deployment feasibility. Rather than producing only binary anomaly decisions, the framework generates continuous anomaly scores that enable prioritization of suspicious mutations according to estimated anomaly severity and contextual abnormality.

An additional insight highlighted by this research concerns a potential security gap that may emerge within hybrid or public-cloud-integrated DNS registry infrastructures prior to final zone generation.

Operational publication pipelines commonly include integrity validation mechanisms that verify consistency between the generated zone file and the final signed zone prior to publication. These safeguards are highly effective for detecting corruption, signing failures or unauthorized modification occurring after zone generation.

However, these mechanisms do not evaluate whether the registry mutations used to generate the zone were themselves malicious or anomalous. If malicious delegation changes, glue modifications or DNSSEC-related mutations are introduced earlier within the registry workflow, they may become incorporated directly into the generated zone and subsequently signed as part of the normal publication process without triggering integrity validation failures.

This consideration becomes particularly important in hybrid and public-cloud-integrated registry environments, where mutation processing, replication pipelines or intermediate components may exist outside traditional tightly controlled on-premise infrastructure boundaries. Under these conditions, malicious or unauthorized mutations introduced prior to zone generation may propagate through otherwise valid workflows while remaining fully consistent with downstream signing and publication checks.

As a result, semantically malicious but technically valid mutations may become embedded within the authoritative zone while still satisfying existing integrity verification mechanisms. This creates a potential detection gap between mutation processing and final zone publication.

The proposed ensemble framework was specifically designed to address this gap by performing anomaly detection directly on registry mutations before zone generation occurs. Rather than validating only the integrity of the final generated zone, the framework analyzes the structural and behavior of mutations themselves, enabling detection of suspicious infrastructure changes before they become incorporated into the signed authoritative DNS zone.

Additionally, the ensemble-based design provides a flexible foundation for incorporating future sig-

nals, including threat intelligence feeds, registrar reputation data, DNS telemetry, historical infrastructure observations and specialized rule-based validation mechanisms.

The results therefore demonstrate how ensemble-based anomaly detection can provide meaningful value for monitoring security-critical authoritative DNS infrastructure and identifying complex registry abuse patterns across multiple layers of DNS operations.

8.6 Limitations

Although the proposed ensemble framework demonstrates strong performance across most implemented attack scenarios, several important limitations should be considered when interpreting the results.

The most important limitation involves the use of synthetic attack injection rather than real-world labeled attack datasets. Because publicly available labeled datasets for malicious authoritative DNS registry mutations do not exist, synthetic attack generation was necessary in order to evaluate detection behavior under controlled and reproducible conditions.

Although the implemented attack scenarios were designed to reflect realistic DNS abuse patterns and were refined using expert feedback, synthetic attacks cannot fully reproduce the stealth characteristics, adaptive behavior and operational diversity of real adversaries. In particular, synthetic attacks may unintentionally introduce stronger structural irregularities or cleaner anomaly patterns than would occur in practice, potentially making them easier to detect than real-world attacks.

Importantly, the consistent difficulty observed for T2 glue manipulation attacks partially supports this concern. Even within a controlled synthetic setting, attacks designed to remain operationally plausible and structurally subtle proved substantially more difficult to detect. This suggests that real-world low-visibility infrastructure attacks may represent an even greater challenge for anomaly detection systems operating within authoritative DNS environments.

Another important limitation involves the highly imbalanced nature of DNS registry mutation data. Malicious behavior represents only a very small fraction of total mutations, while the overwhelming majority of operational activity remains benign. Under these conditions, even relatively small numbers of false positives can significantly affect traditional classification metrics such as precision and F1-score. For this reason, this research focused more strongly on anomaly ranking quality, score separation and operational prioritization behavior rather than strict binary classification performance alone.

Additionally, the experiments were conducted using mutation data originating from a specific registry environment and operational context. Different registries may exhibit different mutation characteristics, registrar ecosystems and operational workflows, potentially influencing anomaly distributions and graph topology behavior.

Finally, the current framework operates over fixed temporal aggregation windows. While this design improves graph stability and computational tractability, it may reduce sensitivity to extremely short-lived attacks or long-term evolving abuse patterns spanning multiple temporal windows.

9 Conclusion and Future Work

9.1 Conclusion

This thesis investigated the problem of detecting malicious or suspicious DNS registry mutations at the authoritative replication boundary within modern hybrid registry infrastructures. As registry operators increasingly adopt hybrid and cloud-integrated architectures, the replication boundary between cloud-hosted operational systems and sovereign authoritative infrastructure becomes an important security control point. Mutations that successfully traverse this boundary may ultimately become incorporated into signed authoritative DNS zones and propagated throughout the global DNS ecosystem.

To address this problem, this research proposed an explainable ensemble-based anomaly detection framework combining heterogeneous graph representation learning with feature-level anomaly detection. The framework models registry mutations as relational infrastructure behavior involving domains, name-servers, registrars, IP subnets and DNSSEC-related entities, while simultaneously incorporating behavioral and statistical mutation features through an Isolation Forest component.

The experimental evaluation demonstrated that the proposed framework is highly effective for detecting structurally significant attacks involving delegation manipulation, coordinated infrastructure abuse, DNSSEC-related anomalies and multi-domain compromise behavior. Most implemented attack scenarios achieved anomaly scores between 0.901 and 1.000 together with ROC-AUC values ranging from 0.968 to 0.996, while also maintaining highly stable ranking behavior across multiple random seeds.

At the same time, the results revealed an important limitation involving subtle infrastructure-level attacks such as T2 glue record manipulation. Unlike the other attack scenarios, T2 preserved most graph connectivity patterns and remained statistically similar to legitimate operational behavior. Consequently, the framework struggled to reliably separate these mutations from benign infrastructure activity.

The consistent inability of detecting T2 glue record manipulation represents one of the most important insights of the research. It demonstrates that attacks designed to remain operationally plausible and structurally subtle may constitute one of the most difficult classes of DNS anomalies to detect, even when using advanced graph-based anomaly detection techniques. The results therefore highlight both the strengths and limitations of structure-oriented anomaly detection approaches within authoritative DNS registry environments.

Beyond its anomaly detection capabilities, the proposed framework addresses an important operational security gap that may emerge prior to authoritative zone generation and publication, particularly within hybrid and public-cloud-integrated registry infrastructures. In this sense, the system can be viewed as a pre-zone behavioral risk filter that evaluates the legitimacy and operational characteristics of registry mutations before they become incorporated into the signed authoritative DNS zone.

Overall, the results demonstrate that ensemble-based graph anomaly detection constitutes a promising and operationally relevant approach for strengthening security monitoring at authoritative DNS registry boundaries while also highlighting important challenges associated with detecting low-visibility infrastructure abuse.

9.2 Future Work

Although the proposed ensemble framework demonstrates strong performance across multiple DNS attack scenarios, several promising directions exist for extending and improving the system in future research.

One important direction involves improving sensitivity to subtle infrastructure-level anomalies such as

T2 glue manipulation attacks. The experimental results demonstrated that attacks preserving both graph topology and operationally realistic feature distributions remain substantially more difficult to distinguish from benign infrastructure behavior. Future systems may therefore benefit from stronger contextual and rule-based infrastructure validation mechanisms operating alongside machine learning models.

Examples include checking whether infrastructure changes are unusual compared to previous behavior, monitoring unexpected network or subnet changes and identifying suspicious infrastructure updates over time. Such mechanisms could provide additional operational context that may not be fully captured through graph structure and statistical feature analysis alone.

Another promising direction involves incorporating dynamic and temporal graph learning techniques. The current framework operates on temporally aggregated graph snapshots constructed over fixed windows. While this design provides stable graph representations and computational tractability, future systems could model continuously evolving DNS infrastructure behavior using dynamic graph neural networks or temporal message passing architectures. Such approaches may improve sensitivity to slow-moving infrastructure compromise, long-term attack progression and sequential multi-stage abuse campaigns.

Future work could additionally investigate alternative GNN architectures and anomaly detection models. Although GraphSAGE demonstrated strong performance within this research, attention-based architectures such as Graph Attention Networks (GATs), relational attention mechanisms and transformer-inspired graph models may improve the ability of the system to selectively focus on anomalous infrastructure relationships during message passing.

Additional graph anomaly detection approaches, including contrastive representation learning, graph autoencoders and dynamic graph anomaly detection frameworks, may further improve representation quality and anomaly sensitivity. Similarly, more advanced ensemble strategies involving temporal sequence models, statistical drift detectors or specialized DNSSEC validation mechanisms may strengthen operational robustness across heterogeneous attack classes.

Another important research direction involves improving evaluation realism beyond synthetic attack injection. Because publicly available labeled datasets for malicious authoritative DNS registry mutations are effectively unavailable, this research relied on threat-model-driven synthetic attacks in order to evaluate anomaly detection behavior under controlled conditions. However, synthetic attacks may unintentionally introduce clearer anomaly patterns or stronger structural irregularities than would occur in practice.

Future work could therefore investigate alternative evaluation methodologies that reduce dependence on purely synthetic attack generation. Possible approaches include expert-assisted labeling, red-team simulation exercises, adversarial attack emulation and semi-supervised learning techniques. Such approaches could improve understanding of how anomaly detection systems perform against stealthier and more operationally realistic adversaries.

Moreover, future systems may also benefit from integrating additional contextual information into the anomaly scoring pipeline. Passive DNS telemetry, BGP data and threat intelligence feeds may provide valuable supplementary context for identifying sophisticated infrastructure abuse patterns.

Another promising direction involves complementing mutation-level anomaly detection with post-generation zone validation mechanisms. While the proposed framework focuses on detecting suspicious mutations before zone publication, additional assurance could be obtained by validating the consistency of generated DNS zones after publication. Because a large portion of zone content remains relatively stable between generation cycles, future systems could perform periodic DNSSEC and delegation consistency checks on newly generated zones and their associated authoritative infrastructure.

Such a validation layer could verify that delegated nameservers remain correctly configured, that DNSSEC trust chains are intact and that authoritative servers return responses consistent with the published delega-

tion state. Additional cross-validation against delegated domains, child zones and authoritative nameserver infrastructure may further help identify inconsistencies introduced through infrastructure-level manipulation. This direction may be particularly valuable for detecting subtle glue-related attacks, where changes to nameserver IP mappings preserve both graph structure and operational plausibility while potentially introducing inconsistencies that only become observable during DNS resolution or DNSSEC validation.

Rather than replacing anomaly detection, such a mechanism would provide an additional independent validation layer capable of detecting classes of infrastructure abuse that remain difficult to identify through mutation behavior alone. The combination of behavioral monitoring before zone publication and consistency validation after zone generation may therefore provide stronger defense-in-depth guarantees for future DNS registry security architectures.

Finally, the current research evaluated only a subset of the broader DNS threat landscape. Future work could therefore incorporate more diverse attack scenarios, including cross-registry abuse, long-term stealthy compromise strategies and more advanced DNSSEC manipulation campaigns.

Overall, the results of this research demonstrate that ensemble-based graph anomaly detection constitutes a highly promising direction for operational DNS security analytics, while also highlighting important opportunities for improving detection sensitivity, evaluation realism and operational robustness within future DNS registry security systems.

References

[1] *International Journal of Latest Technology in Engineering Management & Applied Science*, vol. 14, no. 8, pp. 1039–1045, Sep. 2025. DOI: 10.51583/IJLTEMAS.2025.1408000133 [Online]. Available: https://www.ijltemas.in/submission/online/article/view/2813

[2] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[3] G. Akiwate, S. Savage, G. M. Voelker, and K. C. Claffy, "Risky bizness: Risks derived from registrar name management," in *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 673–686.

[4] W. S. Al Farizi, I. Hidayah, and M. N. Rizal, "Isolation forest based anomaly detection: A systematic literature review," in *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, IEEE, 2021, pp. 118–122.

[5] J. Alonso et al., "Understanding the challenges and novel architectural models of multi-cloud native applications–a systematic literature review," *Journal of Cloud Computing*, vol. 12, no. 1, p. 6, 2023.

[6] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. "Protocol modifications for the dns security extensions. "[Online]. Available: https://datatracker.ietf.org/doc/html/rfc4035

[7] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "Dns security introduction and requirements," Tech. Rep., 2005. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc4033

[8] M. I. Ashiq, O. Hureau, C. Deccio, and T. Chung, "Decoding dnssec errors at scale: An automated dnssec error resolution framework using insights from dnsviz logs," in *Proceedings of the 2025 ACM Internet Measurement Conference*, 2025, pp. 242–257.

[9] B. B R and D. P, "Modern cloud security threats and vulnerabilities: A comprehensive review," *IARJSET*, vol. 12, Sep. 2025. DOI: 10.17148/IARJSET.2025.12846

[10] R. Bekkerman and J. Allan, "Using bigrams in text categorization," Technical Report IR-408, Center of Intelligent Information Retrieval, UMass . . ., Tech. Rep., 2004.

[11] T. Callahan, M. Allman, and M. Rabinovich, "On modern dns behavior and properties," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 3, pp. 7–15, 2013.

[12] Z. Chen, J. Liu, W. Gu, Y. Su, and M. R. Lyu, "Experience report: Deep learning-based system log analysis for anomaly detection," *arXiv preprint arXiv:2107.05908*, 2021.

[13] Cloudflare. "What is dns?" Accessed: 2026-06-01. [Online]. Available: https://www.cloudflare.com/learning/dns/what-is-dns/

[14] A. Del Soldato, D. Sartiano, and M. Martinelli, "Reads: Enhancing. it domains integrity with registrant's anomalies detection system," in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, IEEE, 2024, pp. 1–7.

[15] U. Demirbaga, "Advancing anomaly detection in cloud environments with cutting-edge generative ai for expert systems," *Expert systems*, vol. 42, no. 2, e13722, 2025.

[16] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, 2021, pp. 4027–4035.

[17] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *Proceedings of the 2019 SIAM international conference on data mining*, SIAM, 2019, pp. 594–602.

[18] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, 2017, pp. 1285–1298.

[19] O. A. Ekle and W. Eberle, "Anomaly detection in dynamic graphs: A comprehensive survey," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 8, pp. 1–44, 2024.

[20] ENISA, *Critical cloud computing*, https://www.enisa.europa.eu/publications/critical-cloud-computing, Accessed: 2026-03-30, 2013.

[21] ENISA, *Enisa emerging threat landscape 2020*, https://securitydelta.nl/media/com_hsd/report/338/document/ETL2020-Emerging-Trends-A4-1-.pdf, Accessed: 2026-03-30, 2020.

[22] G. Fernandes Jr, J. J. Rodrigues, L. F. Carvalho, J. F. Al-Muhtadi, and M. L. Proença Jr, "A comprehensive survey on network anomaly detection," *Telecommunication Systems*, vol. 70, no. 3, pp. 447–489, 2019.

[23] N. Gruschka and M. Jensen, "Attack surfaces: A taxonomy for attacks on cloud services," in *2010 IEEE 3rd international conference on cloud computing*, IEEE, 2010, pp. 276–279.

[24] H. Guo et al., "Logformer: A pre-train and tuning pipeline for log anomaly detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, 2024, pp. 135–143.

[25] F. Hadadi, J. H. Dawes, D. Shin, D. Bianculli, and L. Briand, "Systematic evaluation of deep learning models for log-based failure prediction," *Empirical Software Engineering*, vol. 29, no. 5, p. 105, 2024.

[26] T. Hagemann and K. Katsarou, "A systematic review on anomaly detection for cloud computing environments," in *Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference*, 2020, pp. 83–96.

[27] T. Halvorson, M. F. Der, I. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From. academy to. zone: An analysis of the new tld land rush," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 381–394.

[28] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[29] S. Han, X. Hu, H. Huang, M. Jiang, and Y. Zhao, "Adbench: Anomaly detection benchmark," *Advances in neural information processing systems*, vol. 35, pp. 32 142–32 159, 2022.

[30] R. Houser, S. Hao, Z. Li, D. Liu, C. Cotton, and H. Wang, "A comprehensive measurement-based investigation of dns hijacking," in *2021 40th International Symposium on Reliable Distributed Systems (SRDS)*, IEEE, 2021, pp. 210–221.

[31] K. van Hove, J. van der Ham-de Vos, and R. van Rijswijk-Deij, "Domijn: The security of domain registrars and the risk of a domain name takeover," [Online]. Available: https://weis2026.econinfosec.org/wp-content/uploads/sites/13/2026/05/WEIS2026_paper_9.pdf

[32] A. Hrusto, N. B. Ali, E. Engström, and Y. Wang, "Monitoring data for anomaly detection in cloud-based systems: A systematic mapping study," *ACM Transactions on Software Engineering and Methodology*, 2025.

[33] ICANN, *2013 registrar accreditation agreement*, 2013. [Online]. Available: https://itp.cdn.icann.org/en/files/accredited-registrars/registrar-accreditation-agreement-21jan24-en.pdf

[34] Internet Corporation for Assigned Names and Numbers (ICANN), *Root zone algorithm rollover study*, ICANN public report, Accessed: 2026-02-20, May 2024. [Online]. Available: https://www.icann.org/en/system/files/files/root-zone-algorithm-rollover-study-23may24-en.pdf

[35] W. Jia, R. M. Shukla, and S. Sengupta, "Anomaly detection using supervised learning and multiple statistical methods," in *2019 18th IEEE international conference on machine learning and applications (ICMLA)*, IEEE, 2019, pp. 1291–1297.

[36] H. Kim, B. S. Lee, W.-Y. Shin, and S. Lim, "Graph anomaly detection with graph neural networks: Current status and challenges," *IEEe Access*, vol. 10, pp. 111 820–111 829, 2022.

[37] A. Kumar et al., "Revolutionising anomaly detection: A hybrid framework for anomaly detection integrating isolation forest, autoencoder, and conv. lstm," *Knowledge and Information Systems*, vol. 67, no. 12, pp. 11 903–11 953, 2025.

[38] Z. Li, Y. Zhu, and M. Van Leeuwen, "A survey on explainable anomaly detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 1, pp. 1–54, 2023.

[39] K. Lim, R. Sommese, M. Jonker, R. Mok, D. Kim, et al., "Registration, detection, and deregistration: Analyzing dns abuse for phishing attacks," *arXiv preprint arXiv:2502.09549*, 2025.

[40] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*, IEEE, 2008, pp. 413–422.

[41] I. A. O. Maldonado, E. Meeuwissen, P. de Haan, and R. van der Mei, "Telosian: Reducing false positives in real-time cyber anomaly detection by fast adaptation to concept drift.," in *ICISSP (2)*, 2025, pp. 84–97.

[42] J. V. Mamidala, A. Attipalli, S. J. Enokkaren, V. Bitkuri, R. Kendyala, and J. Kurma, "A survey on hybrid and multi-cloud environments: Integration strategies, challenges, and future directions," *International Journal of Humanities and Information Technology*, vol. 5, no. 02, pp. 53–65, 2023.

[43] C. Marques, S. Malta, and J. P. Magalhães, "Dns dataset for malicious domains detection," *Data in brief*, vol. 38, p. 107 342, 2021.

[44] K. Md Imran, A. Mohammad Kowshik, and M. MD Asief, "Ai-based anomaly detection in cloud databases for insider threats," *Journal of Adaptive Learning Technologies*, vol. 2, no. 6, pp. 8–29, 2025.

[45] Microsoft. "Validate dnssec responses." Accessed: 2026-03-30. [Online]. Available: https://learn.microsoft.com/en-us/windows-server/networking/dns/validate-dnssec-responses

[46] B. A. Milani and N. J. Navimipour, "A systematic literature review of the data replication techniques in the cloud environments," *Big Data Research*, vol. 10, pp. 1–7, 2017.

[47] R. Mokadem, J. Martinez-Gil, A. Hameurlain, and J. Kueng, "A review on data replication strategies in cloud systems," *International Journal of Grid and Utility Computing*, vol. 13, no. 4, pp. 347–362, 2022.

[48] E. Mosca, F. Szigeti, S. Tragianni, D. Gallagher, and G. Groh, "Shap-based explanation methods: A review for nlp interpretability," in *Proceedings of the 29th international conference on computational linguistics*, 2022, pp. 4593–4603.

[49] P. Nama, M. Bhoyar, S. Chinta, and P. Reddy, "Optimizing database replication strategies through machine learning for enhanced fault tolerance in cloud-based environments," *Machine learning (ML)*, vol. 63, no. 3, 2023.

[50] S. Ness, V. Eswarakrishnan, H. Sridharan, V. Shinde, N. V. P. Janapareddy, and V. Dhanawat, "Anomaly detection in network traffic using advanced machine learning techniques," *IEEE Access*, vol. 13, pp. 16 133–16 149, 2025.

[51] T. Q. Nguyen, R. Laborde, A. Benzekri, and B. Qu'hen, "Detecting abnormal dns traffic using unsupervised machine learning," in *2020 4th Cyber Security in Networking Conference (CSNet)*, IEEE, 2020, pp. 1–8.

[52] S. Norozpour Sıgaroodı, “Risk assessment for identifying threats, vulnerabilities and countermeasures in cloud computing,” *Jordanian Journal of Informatics and Computing*, vol. 1, no. 1, 2025.

[53] Y. Nosyk, M. Korczyński, C. H. Gañán, M. Król, Q. Lone, and A. Duda, “Don’t get hijacked: Prevalence, mitigation, and impact of non-secure dns dynamic updates,” in *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2023, pp. 1480–1489. DOI: 10.1109/TrustCom60117.2023.00202

[54] Y. Nosyk et al., “Exposing the roots of dns abuse: A data-driven analysis of key factors behind phishing domain registrations,” in *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security*, 2025, pp. 618–632.

[55] C. Nwachukwu, K. Durodola-Tunde, and C. Akwiwu-Uzoma, “Ai-driven anomaly detection in cloud computing environments,” *International Journal of Science and Research Archive*, vol. 13, no. 2, pp. 692–710, 2024.

[56] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, “Deep learning for anomaly detection: A review,” *ACM computing surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.

[57] A. Pekar and R. Jozsa, “Evaluating ml-based anomaly detection across datasets of varied integrity: A case study,” *Computer Networks*, vol. 251, p. 110 617, 2024.

[58] T. Pham-Gia and T. Hung, “The mean and median absolute deviations,” *Mathematical and Computer Modelling*, vol. 34, no. 7, pp. 921–936, 2001, ISSN: 0895-7177. DOI: https://doi.org/10.1016/S0895-7177(01)00109-1 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895717701001091

[59] V. Priya and S. Soumya, “Hybrid anomaly detection in cloud environments: A survey of machine learning and deep learning frameworks,” *Journal of Computational Analysis and Applications*, vol. 33, no. 8, pp. 5632–5648, 2024. DOI: 10.48047/jocaaa.33.8.5632

[60] A. Sadeghi Jahromi, A. Abdou, and P. Oorschot, “A survey and evaluation framework for secure dns resolution,” *IEEE Communications Surveys & Tutorials*, vol. 28, pp. 5532–5560, Jan. 2026. DOI: 10.1109/COMST.2026.3678148

[61] M. Sakurada and T. Yairi, “Anomaly detection using autoencoders with nonlinear dimensionality reduction,” in *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, 2014, pp. 4–11.

[62] G. Schmid, “Thirty years of dns insecurity: Current issues and perspectives,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2429–2459, 2021.

[63] M. Seifert, S. Kuehnel, and S. Sackmann, “Hybrid clouds arising from software as a service adoption: Challenges, solutions, and future research directions,” *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–35, 2023.

[64] SIDN. “Constructive talks with dutch cloud service providers,” SIDN News and Blogs, Accessed: Jan. 26, 2026. [Online]. Available: https://www.sidn.nl/en/news-and-blogs/constructive-talks-with-dutch-cloud-service-providers

[65] B. Singh, “Enhancing real-time database security monitoring capabilities using artificial intelligence,” *International Journal of Current Engineering and Scientific Research (IJCESR)*, 2017.

[66] K. Sivaprasad Yerneni, A. Ravi Teja, K. Sri Harsha, and Y. Naresh Kiran Kumar Reddy, “Towards proactive cloud security: A survey on ml and deep learning-based intrusion detection systems,” *J Contemp Edu Theo Artific Intel: JCETAI-116*, 2025.

[67] R. Sommese et al., “When parents and children disagree: Diving into dns delegation inconsistency,” in *International Conference on Passive and Active Network Measurement*, Springer, 2020, pp. 175–189.

[68] K. Spohn, P. Sattler, and C. Dietze, “The state of DNS delegation: Technical challenges and solution approaches,” in *Proceedings of the Seminar Innovative Internet Technologies and Mobile Communications (IITM), Winter Semester 2024/2025*, ser. Network Architectures and Services (NET), vol. NET-2025-05-1, Munich, Germany: Chair of Network Architectures, Services, School of Computation, Information, and Technology, Technical University of Munich, May 2025, pp. 125–129. DOI: 10.2313/NET-2025-05-1_22

[69] L. A. Trejo, V. Ferman, M. A. Medina-Pérez, F. M. A. Giacinti, R. Monroy, and J. E. Ramirez-Marquez, “Dns-advp: A machine learning anomaly detection and visual platform to protect top-level domain name servers against ddos attacks,” *IEEE Access*, vol. 7, pp. 116 358–116 369, 2019.

[70] Y. Wang and X. Yang, “Machine learning-based cloud computing compliance process automation,” *arXiv preprint arXiv:2502.16344*, 2025.

[71] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[72] M. Wullink, G. C. Moura, M. Müller, and C. Hesselman, “Entrada: A high-performance network traffic data streaming warehouse,” in *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*, IEEE, 2016, pp. 913–918.

[73] Z. Yuan, Q. Sun, H. Zhou, M. Shao, and X. Fu, “A comprehensive survey on gnn-based anomaly detection: Taxonomy, methods, and the role of large language models,” *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 7, pp. 4407–4432, 2025.

[74] M. Zhang et al., “Misty registry: An empirical study of flawed domain registry operation,” in *34th USENIX Security Symposium (USENIX Security 25)*, Seattle, WA: USENIX Association, Aug. 2025, pp. 7117–7134, ISBN: 978-1-939133-52-6. [Online]. Available: https://www.usenix.org/conference/usenixsecurity25/presentation/zhang-mingming

[75] M. Zhang et al., “Misty registry: An empirical study of flawed domain registry operation,” in *34th USENIX Security Symposium (USENIX Security 25)*, 2025, pp. 7117–7134.

[76] Q. Zhang, Z. Shen, I. Karim, E. Bertino, and Z. Li, “Proving dnssec correctness: A formal approach to secure domain name resolution,” *arXiv preprint arXiv:2512.11431*, 2025.

[77] Y. Zhang et al., “Cross the zone: Toward a covert domain hijacking via shared {dns} infrastructure,” in *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 5751–5768.

[78] Y. Zhang et al., “Rethinking the security threats of stale {dns} glue records,” in *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 1261–1277.

[79] Y. Zhauniarovich, I. Khalil, T. Yu, and M. Dacier, “A survey on malicious domains detection through dns data analysis,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[80] L. Zheng, Z. Li, J. Li, Z. Li, and J. Gao, “Addgraph: Anomaly detection in dynamic graph using attention-based temporal gcn.,” in *IJCAI*, vol. 3, 2019, p. 7.

[81] S. Zhou, Q. Tan, Z. Xu, X. Huang, and F.-l. Chung, “Subtractive aggregation for attributed network anomaly detection,” in *Proceedings of the 30th ACM international conference on information & knowledge management*, 2021, pp. 3672–3676.