CONTENTS

LIST OF FIGURES

LIST OF TABLES

BGP-Sentry: Expanding BGP Trust to Non-RPKI ASes via RPKI-enabled ASes as Observers

Anonymous authors

***Abstract*—The Border Gateway Protocol (BGP) lacks trust mechanisms for Autonomous Systems (ASes) not enrolled in the Resource Public Key Infrastructure (RPKI). While RPKI provides cryptographic validation, systemic barriers limits that universal adoption is unachievable—currently only 37% of ASes enforce RPKI, leaving 63% without verifiable trust. We propose BGP-Sentry, a blockchain-based framework that extends trust coverage to the entire BGP ecosystem without requiring universal RPKI adoption. BGP-Sentry transforms RPKI-enabled ASes into behavioral observers that monitor and rate non-RPKI neighbors. Through continuous assessment, non-RPKI ASes accumulate blockchain-recorded trust scores, classifying them as Highly Trusted (≥90), Trusted (70–89), Neutral (50–69), Suspicious (30–49), or Malicious (<30). Network operators can then configure routing policies based on these behavioral ratings, enabling informed decisions about previously unverifiable routes. BGP Coin incentives ensure observer honesty and participation quality. Experimental evaluation demonstrates 96.5% coverage of non-RPKI ASes within 30 days, with sub-linear scaling across 50-500 ASes and consensus times under 3.2 seconds. BGP-Sentry provides a practical pathway to universal BGP trust coverage by complementing RPKI's cryptographic foundation with behavioral trust assessment.**

***Index Terms*—Border Gateway Protocol (BGP), Trustworthy Routing, Blockchain, RPKI, Economic Penalty**

I. Introduction

The Border Gateway Protocol (BGP) is the foundational protocol enabling inter-domain routing across the global Internet, connecting thousands of Autonomous Systems (ASes) [1]. However, the absence of a native authentication mechanism exposes BGP to a wide range of attacks, hijacked [2] and leaked [3], resulting in severe disruptions such as traffic blackholing, interception, or global outages [4]. Recent work demonstrates that even RPKI-protected networks remain vulnerable to update-message flood attacks that create route oscillations capable of destabilizing large portions of the Internet [5].

Since 1995, BGP validation and filtering have relied on the Internet Routing Registry (IRR). It is essentially a public database where operators publish IP-prefix/ASN pairs authorized to originate each prefix. However, IRRs provide neither strong publisher authentication nor cryptographic guarantees of prefix ownership, leading to persistent risks of misinformation or misconfiguration [6]–[8]. Moreover, conflicting data across multiple IRRs creates loopholes that adversaries can exploit to bypass route-monitoring filters [9].

To achieve cryptographic trust, the Resource Public Key Infrastructure (RPKI) [10] was standardized in 2012, to eliminate many BGP attack surfaces and reduce operational errors. Unlike IRRs, RPKI guarantees integrity through certificates and Route Origin Authorizations (ROAs), and today is enforced by approximately 27% of ASes. As a result of major ISPs' participation, 47.7% of IPv4 and 50.45% of IPv6 prefixes are covered by valid ROAs [11].

Although existing research has criticized RPKI's centralized trust model and its complex fetching and caching mechanisms [11], and recent systematization reveals that 56% of RPKI validators contain implementation vulnerabilities exploitable for cache poisoning and denial of service [12], we advocate for its continued evolution and identify two critical challenges that threaten robust ROA enforcement in real-world operations:

Stagnation of ROA Coverage. Although nearly half of all global prefixes are protected by ROAs, only 27% of ASes enforce RPKI-based validation on their BGP sessions, leaving the remainder of the Internet unable to cryptographically verify incoming announcements. Hybrid approaches that combine RPKI with IRR filtering [7] still inherit IRR's weak authentication model, opening additional attack vectors.

Absence of Audit Trail. Existing blockchain-based approaches to BGP security, such as RouteChain [13] and ISRchain [14], focus primarily on passive logging and post-hoc validation. RouteChain groups ASes by geographical proximity rather than actual BGP peering relationships, misaligning with real-world routing topology. ISRchain proposes smart-contract-based route validation but lacks real-time enforcement mechanisms or behavioral trust assessment. Neither system provides reputation-based consequences for misbehavior or incentives for honest observer participation. The RPKI framework (RFC 6480 [10]) specifies how to issue ROAs and perform on-the-fly origin validation, but it does not define any standardized mechanism for retaining or querying historical validation results. As a result, operators cannot retrospectively identify ASes that repeatedly fail RPKI checks nor perform systematic forensic analysis of past invalid announcements.

As such, these observations demonstrate a critical need for an RPKI-based BGP network to 1) incorporate persistent, verifiable auditing for behavior monitoring and 2) expand visibility and accountability (i.e., trust) into non-RPKI networks for informed routing decision-making.

To address these protection gaps, we introduce BGP-Sentry, a blockchain-based trust-expansion framework that transforms RPKI-enabled ASes into distributed observers of their neighbors (including non-RPKI ASes) and records all observed BGP behaviors on a shared ledger. These events are then used to compute a trust score for each AS (RPKI-enabled and Non-RPKI ASes), using protocol-embedded scoring algorithms executed during blockchain consensus. BGP-Sentry employs various behavioral metrics, such as announcement consistency, reverse-path verification, registry cross-checking, and endorsement alignment, to update trust scores dynamically. Finally, ASes are classified into five tiers—Highly Trusted (≥90), Trusted (70–89), Neutral (50–69), Suspicious (30–49), and Malicious (<30)—providing both a historical audit trail among RPKI ASes and informed routing decisions for non-RPKI ASes. Additionally, BGP Coins incentivize RPKI-enabled observers to maintain honest and accurate monitoring of their non-RPKI neighbors.

Our contributions are as follows:

(1) Expanding Trust-based Routing to Non-RPKI ASes. By turning RPKI-enabled ASes into voluntary, distributed BGP observers, we mitigate the real-world challenge of deploying RPKI across the entire network. The proposed blockchain with protocol-embedded trust score computing expands the RPKI root trust to non-RPKI networks and allows informed routing decision-making for RPKI-enabled ASes interfacing unknown ASes, enhancing overall network robustness.

(2) Enabling Behavioral Monitoring and Observer Incentives and Accountability. The RPKI-based blockchain onboarding process preserves the root of trust; thus, BGP-Sentry minimizes the risk of misbehavior of blockchain participants (RPKI-enabled ASes) and achieves reliable trust score assessment for non-RPKI ASes. The BGP Coin incentive system motivates RPKI observers to provide honest and accurate monitoring, while trust scores create reputation-based consequences for non-RPKI ASes, encouraging improved routing behavior.

(3) Establishing Auditable Trust Rating through Scalable Consensus BGP-Sentry employs a Proof of Population (PoP) consensus mechanism—one verified node, one vote—enabling all RPKI observers to agree on behavioral trust ratings for non-RPKI ASes. These consensus-driven ratings are recorded on the blockchain, providing immutable and auditable trust assessments. Evaluation demonstrates scalable operation across large networks with negligible overhead and sub-linear growth.

The rest of this paper is organized as follows: we first outline the motivation for our system design in section II. Then, in section III, we describe system architecture and workflow, followed by a comprehensive discussion of design rationales using security use cases, in section IV. Next, in section V-A4, we evaluate the proposed system in terms of overhead and security enforcement, with extended discussions located in section VI. Lastly, we discuss related works and conclude our work in sections VII and VIII, respectively.

II. Background and Motivation

A. Background

The Border Gateway Protocol (BGP) is the core interdomain routing protocol that enables Autonomous Systems (ASes) to exchange reachability information across the Internet [15]. Despite its critical role, BGP operates largely on implicit trust: route announcements are accepted without strong validation of origin or propagation paths. While extensions like BGPsec aim to provide path security, they face limited deployment. This lack of universal trust enforcement leaves BGP vulnerable to prefix hijacking, route leaks, and other malicious manipulations.

The Resource Public Key Infrastructure (RPKI) offers cryptographic validation by binding IP prefixes to authorized origin ASes, reducing certain hijacking risks [16]. However, RPKI adoption remains incomplete—covering roughly 37% of the global routing space—and focuses only on origin validation [7], [17]. Non-RPKI ASes can still advertise unverified or bogus routes, leaving a significant portion of the Internet exposed to manipulation and policy violations.

Blockchain provides immutability, distributed consensus, and tamper-evident logging, properties well-suited to BGP's trust problem [18]. By enabling ASes to collaboratively record and verify routing events, blockchain ensures that misbehavior is permanently auditable and subject to consensus-based validation. Prior work, such as RouteChain, demonstrated how blockchain can support collective anomaly detection and route decision-making, extending security beyond RPKI's partial coverage.

B. Motivation

Securing trust in the BGP protocol presents multifaceted challenges that existing solutions fail to adequately address. Although RPKI has been proposed as a potential solution, it alone is not sufficient to achieve robust BGP security. This limitation is particularly alarming given that BGP vulnerabilities now threaten not only Internet availability but also financial assets in decentralized environments. For instance, the April 2018 MyEtherWallet attack exploited BGP by hijacking Amazon Route 53 prefixes, enabling attackers to steal over $150,000 worth of Ether in under two hours [19], underscoring the urgent need for stronger protections. We identify three critical challenges that explain why current approaches, including RPKI and existing blockchain-based solutions, cannot fully secure BGP.

Challenge 1: Incomplete Global Adoption of RPKI: RPKI deployment covers only about 37% of Autonomous Systems [7], [17], leaving a significant portion of the Internet without cryptographic validation. This incomplete coverage stems from economic incentive misalignment where RPKI-adopting ASes bear deployment and maintenance costs while providing security benefits to the entire ecosystem without recognition or rewards, creating a classic chicken-and-egg problem where ROA creation is only valuable if networks perform validation, while validation deployment is only worthwhile if sufficient ROAs exist. Non-RPKI ASes can "free-ride" on security improvements provided by RPKI adopters without contributing to the system or being held accountable, while detection and threat mitigation efforts by RPKI participants protect non-participants without compensation or responsibility sharing.

RPKI will never achieve 100% coverage due to various systemic barriers including technical complexity, financial constraints, policy disagreements, and political considerations that prevent universal adoption. In the recherche community the trend is more on enhencing the rpki technooogy, like keeping record in the blockchain [13], Even if we make RPKI more robust and efficient, a substantial portion of ASes will remain outside the RPKI ecosystem due to deployment challenges, resource limitations, or organizational resistance. We cannot simply ignore or abandon these non-RPKI ASes, as they represent critical parts of the Internet infrastructure and their vulnerabilities affect the entire routing ecosystem. Therefore, any comprehensive BGP security solution must find ways to extend trust and validation to the non-RPKI portion of the Internet rather than writing them off as perpetually insecure.

Challenge 2: Absence of Authenticated Routing Observer: Existing blockchain-based BGP security solutions face fundamental flaws in their trust establishment mechanisms, particularly in the initial selection and onboarding of trusted nodes for consensus and validation. Prior work like RouteChain establishes trust by randomly selecting nodes regardless of their RPKI status or routing credibility, creating a circular problem where the system relies on unverified entities to validate routing information, potentially allowing malicious or misconfigured nodes to participate in consensus decisions that affect global routing security. Current blockchain approaches fail to leverage existing cryptographic validation infrastructure when determining which nodes should participate in the trusted consensus network. By ignoring RPKI deployment status and ROA validation capabilities, these systems miss the opportunity to build upon proven verification mechanisms, instead treating all nodes as equally trustworthy despite varying levels of routing competence and security commitment. Without requiring demonstrated routing competence or cryptographic validation capabilities (such as RPKI deployment and ROA maintenance), blockchain-based BGP systems allow potentially unreliable nodes to join the trusted network, compromising the integrity of the entire validation framework.Therefore, an authenticated observer network—verifiable during onboarding and incentivized to operate honestly—is needed to monitor and rate the behavior of non-RPKI ASes, enabling comprehensive coverage of the entire BGP ecosystem.

Challenge 3: Limited Scalability of Security Intelligence Consensus: Current BGP security approaches, including RPKI, operate without the ability to accumulate and maintain comprehensive security intelligence over time. Each security decision is made in isolation without considering historical attack patterns, behavioral trends, or cumulative evidence of malicious activity. This intelligence gap prevents security systems from building a comprehensive understanding of threat landscapes and distinguishing between isolated incidents and coordinated attack campaigns.

The absence of persistent security intelligence severely limits the ability to perform proactive threat detection and establish context-aware security policies. Recent research highlights that many BGP incidents stem from repeated misconfigurations and systematic abuse patterns rather than isolated events [20], yet current systems cannot leverage this historical context for improved security decisions [21].While distributed ledger technology offers immutability for tracking behavioral history, achieving consensus among BGP observers presents unique challenges. Unlike traditional blockchain systems where all nodes observe a shared transaction mempool, BGP monitoring produces fundamentally asymmetric knowledge where each observer sees only announcements reaching its edge router, and ASes can announce different routes to different neighbors based on network configurations and policies. This means distributed observers may hold conflicting yet valid views of the same routing events. Existing solutions lack consensus algorithms designed for BGP's asymmetric, high-volume environment that can scale to Internet-wide deployment while enabling observers to reliably agree on behavioral assessments despite their differing vantage points.

Our Solution: Blockchain-based Trust Expansion Framework. To overcome the aforementioned challenges and provide a secure networking routing environment, we propose BGP-Sentry, a complementary blockchain-based framework that works alongside existing RPKI infrastructure and extends its capabilities. BGP-Sentry specifically addresses persistent origin validation gaps that affect all network participants, including those outside the RPKI-enabled ASes. Using RPKI-enabled ASes as a decentralized network of verified observers, BGP-Sentry enables scalable and continuous behavioral monitoring of routing announcements. It employs smart contract-based trust mechanisms on the blockchain to ensure transparency, accountability, and tamper-resistant validation, while implementing BGP Coin incentives to maintain observer honesty and participation quality.

Specifically, BGP-Sentry addresses the three challenges as follows: (1) *Extended Coverage and Trust Propagation:* BGP-Sentry extends cryptographic trust to networks that do not currently participate in RPKI by introducing a behavioral scoring system that evaluates the legitimacy of route announcements from non-RPKI ASes, effectively bridging the coverage gap and enabling trust propagation beyond the current 37% RPKI deployment while implementing BGP Coin rewards for RPKI adopters and reputation-based accountability measures for non-RPKI ASes; (2) *RPKI-Verified Observers with Honesty Incentives:* BGP-Sentry implements a verified onboarding process that leverages RPKI-enabled ASes as trusted observers based on their demonstrated cryptographic validation capabilities and ROA maintenance, ensuring that only verified, competent nodes participate in the blockchain consensus mechanism. BGP Coin incentives motivate active participation—observers earn coins for collecting signatures, proposing transactions to the blockchain, voting in consensus, and detecting attacks. This dual mechanism—verified identity through RPKI onboarding and sustained engagement through economic rewards—establishes a reliable observer network for monitoring non-RPKI ASes.; and (3) *Scalabel Concensus with Persistent Intelligence:* BGP-Sentry employs a Proof of Population (PoP) consensus mechanism that ensures democratic participation—one verified node, one vote—while preventing

Sybil attacks through mandatory RPKI verification during onboarding. This lightweight consensus scales efficiently to Internet-wide deployment without energy-intensive mining or stake-based plutocracy. Through blockchain's immutable ledger, the system maintains persistent security intelligence, tracking historical attack patterns, distinguishing isolated incidents from coordinated campaigns, and enabling context-aware security decisions based on cumulative threat intelligence.

Through this multi-faceted approach, BGP-Sentry enhances routing security by complementing and strengthening the current RPKI ecosystem without requiring replacement or overhaul of existing infrastructure, allowing easy adoption across the Internet routing landscape.

III. Proposed Architecture

A. System Overview

BGP-Sentry is a blockchain-based framework that transforms RPKI-enabled ASes into distributed observers for rating non-RPKI Autonomous Systems through behavioral trust assessment. The system addresses the fundamental challenge that 40% of Internet routing operates without RPKI validation by leveraging the 60% RPKI-enabled infrastructure as a trusted observation network.

The architecture consists of three core components: (1) **RPKI Observer Network** that monitors and reports non-RPKI routing behavior, (2) **Proof of Population Consensus Framework [22]** that processes observations, attack classifications, and penalties atomically, and (3) **Trust Rating and Incentive System** that maintains behavioral accountability through scoring and rewards. All interactions are recorded on a permissioned blockchain accessible only to RPKI-enabled participants.

Non-RPKI ASes receive **Trust Scores (0-100)** reflecting their routing behavior quality, affecting their network privileges. These scores remain non-tradable to prevent market manipulation of security assessments. **RPKI observers** earn **BGP Coins** as incentives to maintain honest, accurate monitoring of non-RPKI neighbors, with potential future financial value to enhance participation motivation. This dual assessment model extends cryptographic trust to non-participating networks while ensuring observer accountability.

B. RPKI Node Onboarding and Verification

Our proposed architecture restricts participation in the blockchain-based observation role to RPKI-verified Autonomous Systems through an onboarding and validation procedure. During the onboarding phase, a candidate node proves control of a valid RPKI certificate and associated ROAs, and existing members verify these credentials before voting on admission. Upon acceptance, the node is issued a blockchain identity that is cryptographically tied to the RPKI resources it provided and is allowed to join the peer-to-peer blockchain network. After onboarding, the node operates as a fully trusted participant, taking part in consensus, block production, and governance activities. Because blockchain identity is bound to RPKI identity at the time of onboarding, non-RPKI entities are unable to generate valid protocol messages and therefore cannot join or influence the network during this phase. Importantly, once onboarding is complete, the system does not require continuous RPKI verification, which allows future extensions of the model where the blockchain can operate independently of the RPKI infrastructure.

C. Proof of Population Consensus Process

The system employs a novel time-based Proof of Population consensus mechanism that leverages the existing population of RPKI-enabled nodes as validators, avoiding energy-intensive mining while ensuring democratic participation in security decisions.

1) Time-Based Competition Model: When RPKI nodes observe suspicious BGP events, they compete to create blockchain transactions by collecting signatures from other RPKI nodes within a 30-second window. The node collecting the most valid signatures becomes the transaction creator, while others provide endorsement signatures. This competitive model ensures distributed participation while preventing coordination deadlocks.

2) Combined Reporter/Merger Architecture: RPKI nodes serve as both reporters and mergers to maintain observation context and eliminate the coordination overhead of discovering which other nodes witnessed the same BGP events. This combined role reduces network complexity while preserving the contextual information necessary for accurate consensus decisions.

3) Winner Determination and Fork Prevention: Transaction creation rights are determined through hash-based selection: each node computes hash(prefix + AS + timestamp_window), and the node with the lowest hash value becomes the transaction creator. This deterministic algorithm prevents gaming while ensuring fair competition among observers. Hash-based winner determination and atomic transaction bundling prevent competing transactions for identical BGP events, while blockchain immutability ensures trust score adjustments cannot be reversed or duplicated.

4) Nested Decision Structure: Each blockchain transaction contains bundled decisions processed atomically:

- BGP observation details (announcement, timestamp, AS-PATH)
- Attack classification (determined by majority vote among observers)
- Penalty/reward amounts (predetermined by protocol rules)
- Trust score adjustments (calculated automatically)

This nested approach ensures data integrity by preventing partial consensus failures that could leave orphaned observations without security assessments.

5) Consensus Security Analysis: We analyze the security of our Proof of Population (PoP) consensus mechanism through formal threat modeling, safety and liveness guarantees, Byzantine fault tolerance bounds, and game-theoretic incentive analysis.

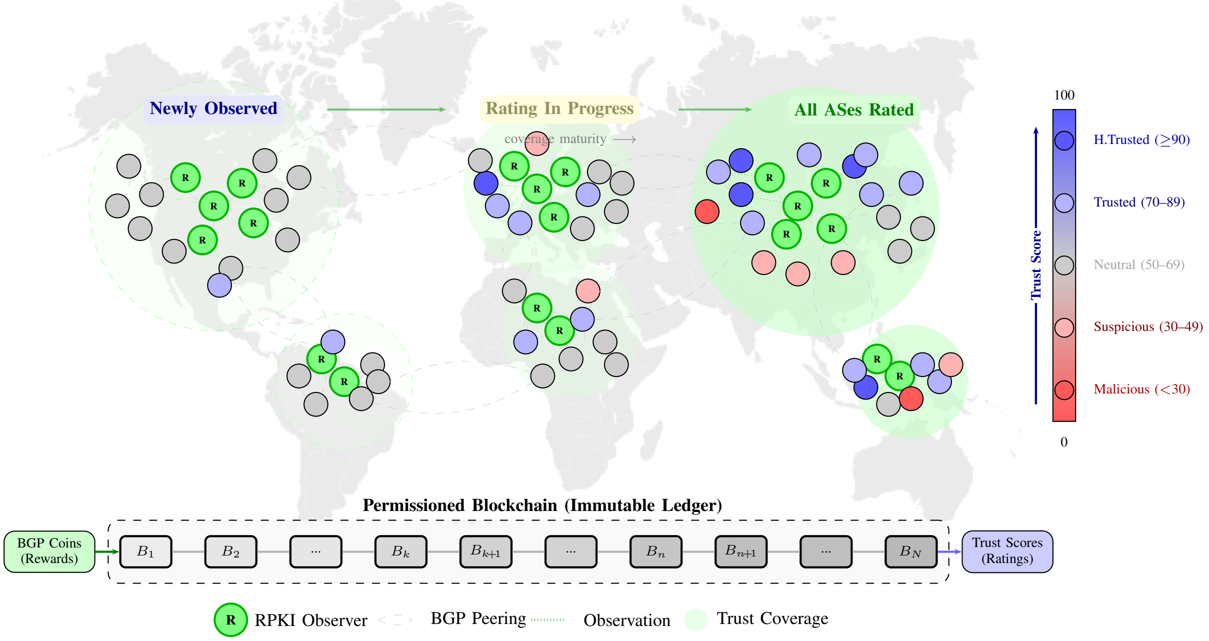

Fig. 1. BGP-Sentry trust coverage expansion reading left-to-right. **Left (Newly Observed):** Regions where RPKI observers (green, marked "R") have just begun monitoring—most non-RPKI ASes remain at default Neutral (gray). **Center (Rating In Progress):** Regions under active monitoring where behavioral data is accumulating; some ASes are assessed while others remain Neutral. **Right (All ASes Rated):** Mature regions where all non-RPKI ASes have been assessed—scores range from Highly Trusted (≥90, dark blue) to Malicious (<30, dark red). The permissioned blockchain (bottom) records observations as blocks accumulate, with BGP Coin rewards incentivizing observers and trust scores enforcing behavioral accountability. The vertical bar (right) maps node colors to the five-tier trust classification.

a) Threat Model: We consider an adversary $\mathcal{A}$ with the following capabilities:

- **Node Corruption:** $\mathcal{A}$ controls up to f out of n RPKI-enabled nodes, where $f < n/3$
- **Network Control:** $\mathcal{A}$ can delay (but not indefinitely block) messages between honest nodes
- **Computational Bounds:** $\mathcal{A}$ is computationally bounded and cannot forge RPKI certificates or digital signatures
- **Adaptive Corruption:** $\mathcal{A}$ can adaptively choose which nodes to corrupt, but corruption requires time $\Delta_{corrupt} > T_{window}$ (longer than the consensus window)

We explicitly exclude the following from our threat model: (1) compromise of Regional Internet Registry (RIR) infrastructure that issues RPKI certificates, (2) adversaries controlling $f \geq n/3$ RPKI-enabled nodes, and (3) complete network partitioning that indefinitely isolates honest nodes.

b) Sybil Attack Prevention: Unlike permissionless blockchains vulnerable to Sybil attacks, BGP-Sentry eliminates this attack vector through mandatory RPKI verification during onboarding:

Theorem 1 (Sybil Resistance)**.** *The probability of a non-RPKI entity successfully joining the consensus network is negligible in the security parameter λ.*

Proof Sketch. Onboarding requires presenting a valid RPKI certificate chain rooted at an RIR trust anchor and demonstrating control of the associated ROA private key. An adversary without legitimate AS ownership must either: (1) forge an RPKI certificate, which requires breaking the underlying cryptographic primitives (negligible probability under standard assumptions), or (2) compromise an RIR (excluded from threat model). Therefore, $Pr[\text{Sybil success}] \leq negl(\lambda)$. □

c) Safety Property: Safety ensures that honest nodes never finalize conflicting transactions for the same BGP event.

Theorem 2 (Consensus Safety)**.** *If two honest nodes finalize transactions tx and tx' for the same BGP event $(prefix, AS, timestamp\_window)$, then $tx = tx'$.*

Proof Sketch. Transaction creation rights are determined by the deterministic hash function $H(prefix \| AS \| timestamp\_window)$, where the node with the lowest hash value becomes the designated creator. For conflicting transactions to be finalized:

1) Two different nodes must claim creator rights for the same event, requiring $H_1 = H_2$ (hash collision, negligible probability), or
2) A malicious creator must produce two different transactions, but finalization requires $\geq 2f+1$ signatures from distinct RPKI nodes. With $f < n/3$ corrupted nodes, at least one honest node must sign both conflicting transactions, which honest nodes refuse by protocol.

Therefore, safety holds with overwhelming probability. □

d) Liveness Property: Liveness ensures that valid BGP observations eventually get recorded on the blockchain.

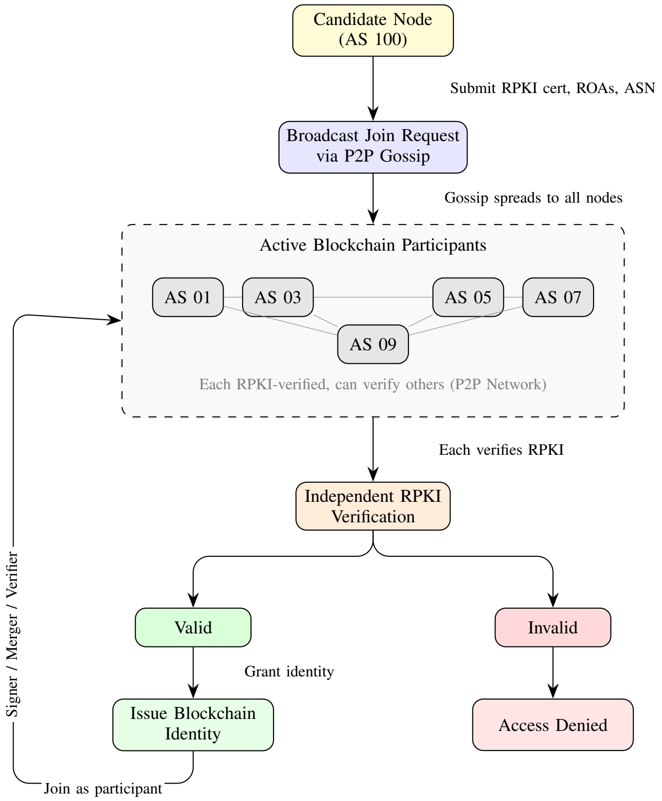

Fig. 2. RPKI-based node onboarding. (1) Candidate AS submits RPKI credentials. (2) Request spreads via P2P gossip. (3) Each existing node independently verifies RPKI (deterministic). (4) Valid credentials receive blockchain identity. (5) New member joins as Signer (vote in consensus), Merger (propose transactions), and Verifier (onboard future candidates).

Theorem 3 (Consensus Liveness)**.** *Under partial synchrony (messages delivered within bound Δ after GST), any BGP event observed by at least one honest RPKI node will be finalized within time $T_{finality} = T_{window} + R \cdot T_{block} + \Delta$.*

Proof Sketch. Given an observed BGP event: (1) the designated creator is determined within T_{window} by hash-based selection; (2) if the designated creator is honest, it initiates signature collection immediately; (3) if the designated creator is malicious and abstains, the next-lowest-hash honest node assumes creator responsibility after timeout $T_{timeout}$; (4) with $n - f > 2n/3$ honest nodes, sufficient signatures ($\geq 2f+1$) are always obtainable; (5) finality is achieved after R block confirmations. The honest majority ensures progress cannot be indefinitely blocked. □

e) Byzantine Fault Tolerance: BGP-Sentry tolerates up to $f < n/3$ Byzantine (arbitrarily malicious) RPKI nodes:

$$f_{max} = \left\lfloor \frac{n-1}{3} \right\rfloor \tag{1}$$

This bound arises from the signature threshold requirement: transactions require $2f+1$ signatures for finalization, ensuring that even if all f malicious nodes collude, they cannot finalize transactions without honest node participation. Conversely, $n - f \geq 2f + 1$ honest nodes can always achieve finalization without malicious cooperation.

f) Fork Prevention and Resolution: Secret forking is prevented through the public signature collection requirement:

Lemma 1 (Fork Detectability)**.** *Any attempt to create a secret fork requires obtaining $2f+1$ signatures without honest node awareness, which is impossible when $f < n/3$.*

When transient forks occur due to network delays, resolution follows the *heaviest signature chain* rule:

$$Chain_{canonical} = \arg \max_{c \in Forks} \sum_{b \in c} |Signatures(b)| \tag{2}$$

Finality is achieved after R confirmations, where R is configurable based on desired security-latency tradeoffs. With $R = 6$ (default), the probability of a finalized block being reverted is bounded by $(f/n)^R < 0.0014$ even under worst-case adversarial conditions.

g) Incentive Compatibility Analysis: We demonstrate that honest behavior constitutes a Nash equilibrium under the BGP Coin mechanism through game-theoretic analysis.

Let U_H denote the expected utility of honest behavior and U_D denote the expected utility of dishonest behavior for an RPKI observer:

$$U_H = C_{base} \cdot A_{max} \cdot P_{max} \cdot Q_{max} \cdot T \tag{3}$$

$$U_D = C_{base} \cdot A_{min} \cdot P_{min} \cdot Q_{min} \cdot T + G_{attack} - P_{detect} \cdot L_{penalty} \tag{4}$$

where:

- $A_{max} = 1.5$, $A_{min} = 0.5$ (accuracy multiplier bounds)
- $P_{max} = 1.2$, $P_{min} = 0.8$ (participation multiplier bounds)
- $Q_{max} = 1.3$, $Q_{min} = 0.9$ (quality multiplier bounds)
- G_{attack} = potential gain from facilitating an attack
- P_{detect} = probability of detection through post-hoc analysis
- $L_{penalty}$ = penalty for detected dishonesty (coin slashing + reputation loss)

Theorem 4 (Incentive Compatibility)**.** *Honest behavior is a dominant strategy when:*

$$\frac{U_H}{U_D} = \frac{2.34}{0.36 + \frac{G_{attack} - P_{detect} \times L_{penalty}}{C_{base} \times T}} > 1 \tag{5}$$

With system parameters calibrated such that $P_{detect} \geq 0.85$ (achievable through blockchain audit trails and cross-observer validation) and $L_{penalty} \geq 3 \times G_{attack}$, the inequality holds for all rational adversaries. The resulting $6.5\times$ earning differential between honest and dishonest behavior, combined with the immutable audit trail enabling post-hoc detection, ensures that the expected utility of honest participation strictly dominates dishonest strategies.

h) Post-Hoc Analysis and Accountability: The blockchain's immutable record enables retrospective detection of dishonest behavior through:

1) **Cross-validation audits:** Comparing an observer's reported assessments against other observers' records for the same BGP events
2) **Temporal consistency checks:** Detecting contradictory assessments from the same observer across time windows
3) **Ground truth reconciliation:** Correlating trust score assignments with subsequently confirmed attacks or legitimate behavior

Detected dishonesty triggers automatic penalties: BGP Coin slashing (minimum 50% of accumulated balance), accuracy multiplier reduction to $A_{min} = 0.5$ for subsequent 90 days, and public recording of the violation on the blockchain. This accountability mechanism ensures that even if dishonest behavior initially escapes real-time detection, the expected long-term cost exceeds any short-term gains.

6) Consensus Protocol Specification: We formalize the Proof of Population consensus mechanism through four interconnected algorithms. Algorithm 1 presents the main consensus protocol, Algorithm 2 describes AS-path-based observer discovery, Algorithm 3 details time-bounded signature gathering, and Algorithm 4 specifies atomic transaction creation.

The protocol achieves consensus correctness through the interplay of these algorithms: observation-driven consensus (Algorithm 1) ensures every BGP event is recorded with a three-tier consensus status, targeted observer discovery (Algorithm 2) reduces network overhead from $O(N)$ to $O(k)$ with random-peer fallback for coverage guarantees, time-bounded vote collection (Algorithm 3) enables timely finalization with early termination when the BFT threshold $\tau = \max(3, \min(\lfloor N/3 \rfloor + 1, 5))$ is met, and post-commit attack classification (Algorithm 4) uses independent per-node detection with majority-vote confirmation to update non-RPKI trust scores and distribute BGPCOIN incentives.

D. Knowledge Aggregation and Coordination

Unlike traditional blockchains where all nodes share symmetric knowledge of pending transactions, BGP observations are fundamentally asymmetric across network locations. Each RPKI node observes only BGP announcements that traverse its position in the network topology, creating scenarios where observers hold conflicting yet valid views of the same routing events. BGP-Sentry addresses this challenge through synchronized time-window processing, intelligent observer discovery, and confidence-weighted trust assessment based on observer coverage.

1) Asymmetric BGP Observation Problem: Each RPKI node observes only a subset of BGP announcements based on their network position and peering relationships. A prefix announcement from a non-RPKI AS may be visible to some RPKI observers but not others, depending on routing policies, AS-path propagation, and peering agreements. This asymmetry creates an inherent coordination challenge: nodes must achieve consensus about behavioral assessments for events that they may not all have witnessed directly. Unlike financial blockchains where double-spending is globally detectable, BGP misbehavior may only be observable from specific vantage points, requiring mechanisms to aggregate

Algorithm 1 Observation-Driven Consensus Protocol

```
Require: BGP event $e = (prefix, origin\_AS, timestamp, AS\_path)$
Require: Current node $N_i$ with RPKI identity $ID_i$
Require: Consensus threshold $\tau = \max(3, \min(\lfloor N/3 \rfloor + 1, 5))$
Require: Timeout $T_{reg} = 30s$, $T_{atk} = 60s$
Ensure: Blockchain transaction $TX$ with consensus status

$dedup\_key \leftarrow (prefix, origin\_AS)$
if $dedup\_key \in$ LastSeenCache $\wedge$ age $< T_{sampling}$ then
    return $\perp$    ▷ Skip duplicate within sampling window
end if

$TX \leftarrow$ CREATESIGNEDTX$(e, SK_i)$    ▷ RSA-2048 signature
$Obs \leftarrow$ DISCOVEROBSERVERS$(AS\_path)$    ▷ Alg. 2
BROADCASTVOTEREQUEST$(TX, Obs)$    ▷ Request peer votes
$Sigs \leftarrow$ COLLECTVOTES$(TX, \tau, T_{reg})$    ▷ Alg. 3

if $|Sigs| \geq \tau$ then    ▷ Full consensus
    $TX.status \leftarrow$ CONFIRMED
else if $|Sigs| \geq 1$ then    ▷ Partial consensus (timeout)
    $TX.status \leftarrow$ INSUFFICIENT_CONSENSUS
else    ▷ No peer corroboration
    $TX.status \leftarrow$ SINGLE_WITNESS
end if
APPENDTOCHAIN$(TX)$    ▷ All observations recorded
REPLICATEBLOCK$(TX$, peers$)$    ▷ P2P block replication
return $TX$
```

Algorithm 2 AS-Path Observer Discovery

```
Require: AS-path $P = [AS_1, AS_2, \ldots, AS_n]$
Require: RPKI registry $\mathcal{R}$, threshold $\tau$
Ensure: Set of relevant RPKI observers $Obs$

$Obs \leftarrow \emptyset$
for each $AS_j \in P$ do
    for each $AS_k \in$ NEIGHBORS$(AS_j)$ do
        if $AS_k \in \mathcal{R}$ then $Obs \leftarrow Obs \cup \{AS_k\}$
        end if
    end for
end for
if $|Obs| < \tau$ then    ▷ Fallback: random RPKI peers
    $Obs \leftarrow Obs \cup$ RANDOMSAMPLE$(\mathcal{R} \setminus Obs,\ \tau - |Obs|)$
end if
return $Obs$
```

Algorithm 3 Time-Bounded Vote Collection

```
Require: Transaction $TX$, observers $Obs$, timeout $T$
Require: Threshold $\tau = \max(3, \min(\lfloor N/3 \rfloor + 1, 5))$
Ensure: Set of valid votes $V$

$V \leftarrow \emptyset$, $t_0 \leftarrow$ NOW()
for each $N_j \in Obs$ do                                ▷ Async broadcast
    ASYNCSEND(VOTEREQUEST($TX$), $N_j$)
end for
while NOW() $- t_0 < T$ do
    if RECEIVE(vote $v_j$ from $N_j$) then
        if $N_j \in \mathcal{R} \wedge N_j \notin V \wedge$ VERIFYSIG($v_j$) then
            $V \leftarrow V \cup \{(N_j, v_j)\}$
        end if
    end if
    if $|V| \geq \tau$ then                              ▷ Early termination
        break
    end if
end while
return $V$
```

Algorithm 4 Attack Classification and Incentive Distribution

```
Require: Committed transaction $TX$, detector node $N_d$
Require: Non-RPKI rating table $\mathcal{T}$, BGPCOIN ledger $\mathcal{L}$
Ensure: Updated $\mathcal{T}$ and $\mathcal{L}$

// Step 1: Independent attack detection
$attacks \leftarrow$ DETECTATTACKS($TX$)                  ▷ 4 detectors
if $attacks = \emptyset$ then
    $\mathcal{T}[origin\_AS].legit\_count$ += 1
    if $\mathcal{T}[origin\_AS].legit\_count \bmod 100 = 0$ then
        $\mathcal{T}[origin\_AS].score$ += 1              ▷ Per 100 clean
    end if
    return
end if

// Step 2: Attack consensus via majority vote
BROADCASTATTACKPROPOSAL($TX$, $attacks$)
Collect votes; each peer runs own DETECTATTACKS
if $yes > no \wedge total \geq 3$ then                    ▷ ATTACK_CONFIRMED
    $P \leftarrow$ GETPENALTY($type$)                     ▷ −20/18/25/10/15
    $\mathcal{T}[origin\_AS].score$ += $P$
    $\mathcal{L}[N_d]$ += $C_{base} \times A \times P \times Q$      ▷ Detector reward
    for each correct voter $N_j$ do
        $\mathcal{L}[N_j]$ += 2                           ▷ Voter reward
    end for
else if $no > yes$ then                                  ▷ NOT_ATTACK
    $\mathcal{L}[N_d]$ −= 20                              ▷ False accusation penalty
end if
```

partial observations into coherent trust assessments.

2) Window-Based Synchronization Protocol: To coordinate distributed observations without centralized data collection, all RPKI nodes process events within synchronized time windows. Nodes read BGP observations from their local routing logs within agreed timestamp ranges (typically 30-second intervals), enabling coordinated processing with deterministic temporal boundaries. When a node detects suspicious behavior, it initiates signature collection within the current window, and all participating nodes independently verify whether they observed corroborating evidence within the same timeframe. Clock synchronization through standard NTP protocols ensures temporal consistency across geographically distributed observers, while the 30-second window accommodates typical BGP propagation delays without sacrificing detection responsiveness.

3) AS-Path Observer Discovery: Rather than flooding all RPKI nodes with signature requests, the system uses AS-path analysis to identify which nodes likely observed the same BGP announcement. When a node detects a suspicious announcement with AS-path $[AS_1, AS_2, ..., AS_n]$, it queries the blockchain registry to identify RPKI-enabled ASes that are direct neighbors of any AS in the path. This targeted approach reduces network overhead from $O(N)$ to $O(k)$ where k is the number of relevant observers, while maintaining consensus coverage by focusing signature collection on nodes with direct observation capability.

4) Signature Collection Optimization: The system prevents network saturation during high BGP activity periods by limiting signature collection to nodes identified through AS-path analysis. Each signature request includes the observation details (prefix, origin AS, timestamp, AS-path), allowing receiving nodes to verify against their local logs before signing. Nodes that observed corroborating evidence provide endorsement signatures, while nodes without relevant observations abstain rather than sign blindly. This selective participation ensures that consensus reflects actual multi-vantage-point verification rather than uninformed agreement.

5) Transaction Buffering and Deduplication: To manage memory under sustained BGP traffic, each RPKI node maintains a three-layer buffer architecture: (1) a *knowledge base* that stores recent BGP observations within a configurable time window (default 480 s) for knowledge-based vote validation, (2) a *pending-votes buffer* that holds transactions awaiting consensus signatures, and (3) a *last-seen cache* that records the most recent blockchain-committed timestamp for each (prefix, origin AS) pair, enabling O(1) duplicate detection without scanning the chain.

All three buffers enforce configurable capacity limits (Table VI footnote) to prevent unbounded memory growth: when a buffer reaches capacity, the oldest entries are evicted. Additionally, TTL-based garbage-collection threads periodically remove expired entries—knowledge-base observations older than the time window, and committed-transaction identifiers older than a cleanup interval. Multi-layer deduplication operates at three stages: RPKI-node-level filtering suppresses duplicate (prefix, origin) observations within a one-hour window, the last-seen cache prevents re-committing the same announcement to the blockchain, and vote-replay detection rejects duplicate signatures from the same AS on the same transaction. A final blockchain-level check ensures that even if a transaction bypasses earlier layers, it cannot be written to the chain twice. All parameters—buffer sizes, time windows, and cleanup intervals—are exposed as environment variables (Table VII) for tuning without code changes.

Consensus timeout handling ensures zero data loss: regular transactions time out after 30 s and attack proposals after 60 s. Timed-out transactions are still committed to the blockchain with a `consensus_status` field indicating "CONFIRMED" (3+ approvals), "INSUFFICIENT_CONSENSUS" (1–2 approvals), or "SINGLE_WITNESS" (0 approvals), preserving every observation as an immutable record regardless of vote coverage.

6) Observer Coverage and System Response: A core design principle of BGP-Sentry is that *every* BGP observation—whether it represents normal routing behavior or a detected attack—is processed through the blockchain pipeline. RPKI observers do not merely flag anomalies; they record all behavioral evidence as immutable blockchain transactions, building a complete audit trail for each non-RPKI AS. The number of observers determines the confidence of that record. Figure 3 illustrates four coverage scenarios and how BGP-Sentry handles each.

(a) No RPKI Observer—No Blockchain Record: When a non-RPKI AS (N1) has no RPKI-enabled neighbors, no observations can be collected and no transactions are written to the blockchain for N1. The system assigns a default Neutral score (50) but marks it as *unverified*. Notably, this scenario does not occur in any of our four CAIDA-derived datasets (100–1,000 ASes): because real-world RPKI adoption reaches $\sim$50–60% within BFS subgraphs, every non-RPKI AS has at least one RPKI neighbor. This confirms that even with partial RPKI deployment, 1st-hop observation achieves complete non-RPKI coverage in realistic topologies. In networks with lower RPKI density, routing policies can distinguish between verified and unverified default scores; as new RPKI-enabled peers appear, BGP-Sentry automatically begins recording behavior without reconfiguration.

(b) Single Observer—All Behavior Recorded, Low Confidence: When exactly one RPKI node (R1) peers with N1, R1 processes *every* BGP announcement from N1 through the four attack detectors. Both outcomes produce blockchain transactions: legitimate announcements are recorded as *clean observation* entries that maintain N1's trust score, while detected anomalies produce *attack detection* transactions with associated penalties. This means the blockchain accumulates positive evidence (clean behavior) alongside negative evidence (attacks), not just violations. However, because only one vantage point exists, the system assigns high uncertainty ($\sigma = \pm 20$ points), and operators can require multi-observer confirmation before acting on the score.

(c) Dual Observers—Cross-Validated Blockchain Evidence: With two RPKI observers (R1, R2), both independently record all of N1's behavior on the blockchain. When both report consistent clean behavior, the accumulated positive evidence strengthens N1's trust score with medium confidence ($\sigma = \pm 10$ points). When both detect the same attack, the cross-validated penalty carries stronger corroboration. Conflicting observations—e.g., R1 detects a prefix hijack but R2 sees a legitimate announcement—are both recorded, and the consensus mechanism weighs corroborating evidence before committing penalties.

(d) Multiple Observers—Comprehensive Record, Trust Promotion: When four or more RPKI nodes observe N1, the blockchain accumulates behavioral evidence from diverse vantage points with narrow confidence ($\sigma = \pm 3$ points). This comprehensive record enables *trust promotion*: if N1 demonstrates sustained clean behavior verified by multiple independent observers, its score rises above 70, promoting it from Neutral to Trusted. Conversely, if N1 launches an attack, the high-confidence multi-observer detection results in stronger penalties. The system will not promote an AS based on a single observer's reports alone—only comprehensive, corroborated blockchain evidence supports tier changes.

The trust scoring mechanism incorporates observer count as a confidence multiplier:

$$C_{confidence} = 1 - \frac{1}{\sqrt{N_{observers} + 1}} \tag{6}$$

where $N_{observers}$ is the number of RPKI nodes with direct peering to the assessed non-RPKI AS. This formula ensures diminishing returns from additional observers while maintaining meaningful confidence improvements as coverage increases.

E. Trust Rating and Incentive Framework

The system combines immediate behavioral penalties for non-RPKI nodes with long-term incentives for RPKI observers, creating a comprehensive accountability framework that addresses both security threats and participation motivation.

1) Non-RPKI Trust Score Mechanism: The trust rating engine operates as protocol-embedded algorithms that execute automatically during blockchain consensus, eliminating centralized trust authorities. The system employs dual engines for comprehensive behavioral assessment: immediate penalty enforcement and periodic compliance evaluation.

The **Reactive Trust Engine** provides sub-second response to detected violations through immediate trust score adjustments. Penalties follow the formula:

$$Penalty = P_{base} \times S_{weight} \times R_{factor} \tag{7}$$

where base penalties are $P_{prefix} = 20$, $P_{subprefix} = 18$, $P_{bogon} = 25$, $P_{flapping} = 10$, $P_{leak} = 15$, with severity weights and recency factors accounting for attack impact and violation patterns. Repeated attacks within 30 days incur an additional -30 penalty, and persistent attackers (3+ total attacks) receive a -50 penalty.

The **Adaptive Trust Engine** conducts monthly evaluations using behavioral metrics derived from 30-day historical analysis, representing the only mechanism capable of trust score improvement. The evaluation considers attack frequency,

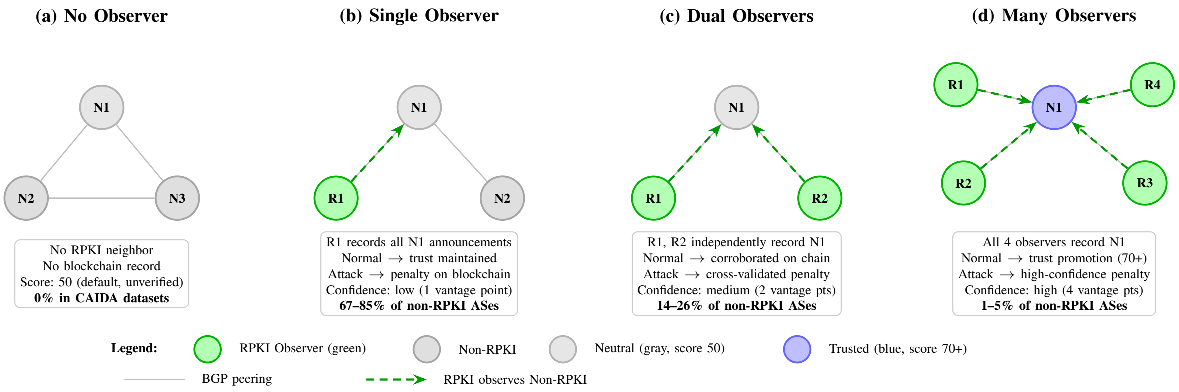
(a) No Observer
(b) Single Observer
(c) Dual Observers
(d) Many Observers
N1
N2
N3
No RPKI neighbor
No blockchain record
Score: 50 (default, unverified)
0% in CAIDA datasets
N1
R1
N2
R1 records all N1 announcements
Normal → trust maintained
Attack → penalty on blockchain
Confidence: low (1 vantage point)
67–85% of non-RPKI ASes
N1
R1
R2
R1, R2 independently record N1
Normal → corroborated on chain
Attack → cross-validated penalty
Confidence: medium (2 vantage pts)
14–26% of non-RPKI ASes
R1
N1
R4
R2
R3
All 4 observers record N1
Normal → trust promotion (70+)
Attack → high-confidence penalty
Confidence: high (4 vantage pts)
1–5% of non-RPKI ASes
Legend:
RPKI Observer (green)
Non-RPKI
Neutral (gray, score 50)
Trusted (blue, score 70+)
BGP peering
RPKI observes Non-RPKI

Fig. 3. Observer coverage scenarios and blockchain recording, with coverage percentages from our CAIDA-derived datasets (100–1,000 ASes). In every observed case, RPKI nodes record *all* BGP behavior—both legitimate announcements and detected attacks—as transactions on the blockchain. (a) No RPKI neighbor: theoretically possible but does not occur in our datasets (0%), since real-world RPKI adoption ($\sim$50–60% in BFS subgraphs) ensures every non-RPKI AS has at least one RPKI neighbor. (b) Single observer (67–85%): the dominant case; R1 records all of N1's announcements with low confidence. (c) Dual observers (14–26%): R1 and R2 independently record and cross-validate N1's behavior on-chain. (d) Multiple observers (1–5%): comprehensive recording enables high-confidence trust promotion or penalties.

announcement stability, prefix consistency, response time, and participation patterns to reward sustained compliance.

Trust scores classify non-RPKI ASes into five actionable tiers: Highly Trusted ($\geq$90) receive preferential routing, Trusted (70–89) receive normal routing privileges, Neutral (50–69) operate under standard monitoring, Suspicious (30–49) face enhanced monitoring with restricted access, and Malicious ($<$30) experience restricted routing access. All ASes start at 50 (Neutral).

2) RPKI BGP Coin System: BGP Coins serve as protocol-level incentive tokens designed to maintain RPKI observer honesty and participation quality. While initially internal-only and non-tradable, the system design allows for future financial value to enhance participation motivation and create sustainable economic incentives for network security monitoring.

RPKI observers earn coins through verified security contributions according to:

$$C_{earned} = C_{base} \times A_{accuracy} \times P_{participation} \times Q_{quality} \quad (8)$$

where base rewards range from 10 coins/day for consistent monitoring to 100 coins for accurate attack detection, with multipliers reflecting historical accuracy (0.5–1.5), participation consistency (0.8–1.2), and evidence quality (0.9–1.3).

3) Coin Economy and Circulation: The coin economy operates through a sustainable circular model. A protocol treasury initially containing 10 million coins distributes rewards to observers based on three primary activities: daily monitoring contributions (+10 coins), confirmed attack detection (+100 coins), and consensus participation (+1–5 coins per block signing). Observers accumulate coins in their wallets, with final earnings adjusted by the multiplicative factors above—honest, high-quality participants can achieve up to 2.34× base rewards, while dishonest behavior reduces earnings to 0.36×, creating a 6.5× earning gap that economically discourages manipulation.

Accumulated coins provide utility through network services: governance voting rights, access to premium analytics dashboards, and priority technical support. Spent coins undergo processing where 50% are permanently burned (preventing inflation and maintaining coin value) while 50% are recycled to the treasury (ensuring long-term reward sustainability). This dual mechanism balances deflationary pressure with operational continuity.

4) Governance and Democratic Control: Governance mechanisms enable democratic protocol evolution through BGP Coin-weighted voting. Different decision categories require varying consensus thresholds: trust scoring modifications require 75% agreement, reward parameter adjustments require 66%, and new threat detection integration requires 60%. This tiered approach ensures community control over critical system parameters while preventing centralized manipulation of economic incentives. The blockchain's immutable audit trail provides transparency, allowing any participant to verify that governance decisions and coin distributions follow protocol rules.

F. Design Rationale

In this subsection, we explain the key design decisions that shape BGP-Sentry architecture. Each decision is made after considering the practical constraints of BGP monitoring and the limitations of existing approaches.

Why Permissioned Blockchain: We choose permissioned blockchain over public blockchain like Ethereum for several practical reasons. First, RPKI infrastructure already gives us a natural way to control who can participate, so we do not need additional access control mechanism. Second, permissioned consensus can achieve finality in less than 3 seconds, whereas public blockchain takes several minutes which is not acceptable for attack detection scenarios. Third, since all participants are known RPKI-enabled ASes, we can hold them accountable for their actions which is difficult in anonymous public chains.

Fourth, the computational requirement is much lower so ASes can run the system on their existing infrastructure without buying specialized hardware for mining.

Why 30-Second Competition Window: The choice of 30-second window for signature collection is based on how BGP announcements propagate in real networks. From literature we know that BGP announcements take roughly 2 to 5 minutes to propagate globally across the Internet. So 30 seconds is enough time for observers in different geographic locations to receive the announcement and participate in consensus. If we use shorter window like 10 seconds, observers that are far away will be excluded because of network delay. On the other hand, if we use longer window like 60 seconds or more, the system response to attacks becomes too slow and malicious routes can cause damage before detection. We found 30 seconds to be a reasonable balance between inclusivity and responsiveness.

Why Time-Window Synchronization: Traditional blockchain systems assume that all nodes have same view of pending transactions, but this assumption does not hold for BGP observations. Each RPKI node can only see BGP announcements that pass through its location in the network topology. This creates a fundamental problem where nodes need to reach consensus about events that not everyone has witnessed. We solve this by using synchronized time windows where all nodes process observations within agreed timestamp boundaries. This way, nodes do not need to wait for acknowledgment from every other node because the time boundary itself acts as a deterministic cutoff point. All nodes can independently verify whether an observation falls within the window or not.

Why Combined Reporter and Merger Role: In many blockchain designs, the entity that submits transactions is different from the entity that creates blocks. However, we intentionally combine these two roles for BGP-Sentry. The reason is that BGP observations are local in nature. Only the node that actually received a BGP announcement has the full context including exact timing, complete AS-PATH, and relationship with the announcing peer. If we separate reporter and merger roles, then the merger would need to collect this contextual information from reporters which adds communication overhead and creates opportunity for information to be lost or delayed. By keeping both roles together, the observer maintains full context throughout the consensus process.

Why These Penalty Values: The trust penalty values we use are $P_{bogon} = 25$, $P_{prefix} = 20$, $P_{subprefix} = 18$, $P_{leak} = 15$, and $P_{flapping} = 10$. These numbers reflect the actual severity of each attack type in terms of traffic impact. Bogon injection receives the highest penalty because announcing reserved address space (e.g., RFC 1918) is always illegitimate and indicates clear malicious intent. Prefix hijacking receives a high penalty because attackers can redirect all traffic destined to a prefix. Subprefix hijacking receives a slightly lower penalty because, due to longest prefix matching, only a portion of traffic is affected. Route leak receives a moderate penalty as it often stems from misconfiguration rather than intentional attack. Route flapping receives the lowest penalty because it primarily causes convergence instability rather than traffic interception. We calibrate these values such that if an AS commits 2 to 3 confirmed attacks, their trust score drops from Neutral (50) to Suspicious (30–49). Repeated attacks within 30 days incur an additional -30 penalty, and persistent attackers (3+ total attacks) receive a -50 penalty, rapidly pushing scores into Malicious territory (<30). This gives tolerance for occasional mistakes while punishing repeated bad behavior.

IV. Security Enforcement

We design a comprehensive experimental methodology to validate our BGP security framework's effectiveness in addressing critical BGP security challenges. Our approach focuses on detecting and preventing four major BGP attack types through behavioral trust assessment, reputation-based enforcement, and real-time anomaly detection using our CAIDA-derived simulation environment spanning 100 to 500 ASes.

The attack detection and enforcement pipeline operates in four sequential phases. In Phase 1 (*Observe*), each RPKI-enabled AS monitors BGP announcements received from its non-RPKI neighbors. Every incoming announcement is processed through four independent detectors that run locally on the observer node: (i) **PREFIX_HIJACK**—the announcement's origin AS does not match the authorized origin in the ROA database for that prefix; (ii) **SUBPREFIX_HIJACK**—the announced prefix is a more-specific subnet of an existing ROA entry with a different origin AS; (iii) **BOGON_INJECTION**—the announced prefix falls within reserved address ranges (RFC 1918, RFC 5737, RFC 6598); and (iv) **ROUTE_FLAPPING**—the same (prefix, origin) pair has been announced more than 5 times within a 60-second sliding window. In Phase 2 (*Detect*), when any detector flags an anomaly, the observer creates a transaction containing the BGP observation details, the attack classification, and the predetermined penalty amount. In Phase 3 (*Consensus*), the transaction is broadcast via the P2P message bus to all RPKI validators, who independently verify the claim and provide endorsement signatures within the 30-second competition window. If the signature count meets the BFT threshold ($\max(3, \min(\lfloor N/3 \rfloor + 1, 5))$), the transaction achieves consensus. In Phase 4 (*Record*), the consensus-confirmed transaction is committed to the blockchain as an immutable record. The non-RPKI AS's trust score is reduced by the attack-specific penalty, and the reporting RPKI observer earns BGP Coin rewards proportional to its participation.

A. Prefix Hijacking Prevention Mechanisms

This experiment validates our system's integrated capability to detect and prevent unauthorized prefix announcements through coordinated architectural layers. The BGP Observation Layer continuously monitors announcements from non-RPKI neighbors, forwarding suspicious patterns to the Reactive Trust Engine for historical comparison. Upon confirmed hijacking detection, the Trust Engine triggers automated trust score adjustments, affecting the AS's network acceptability.

The Reactive Trust Engine processes prefix hijacking through historical pattern analysis, comparing current announcements against legitimate ownership records stored in the blockchain. Trust scores decrease immediately upon detection: first-time offenders experience 20-point reductions, repeated attackers (within 30 days) face an additional 30-point penalty, and persistent attackers (3+ total) receive a 50-point penalty. As trust scores decline, the AS gradually loses acceptability in routing decisions, with other ASes potentially filtering routes based on trust thresholds.

We implement 10 distinct prefix hijacking scenarios targeting legitimate prefixes originally announced by RPKI-enabled ASes (AS1-AS9). Attack sources are randomly selected from non-RPKI ASes (AS10-AS18) to announce unauthorized prefixes. When trust scores approach zero, the malicious AS faces potential blocking for a temporary period, after which they can re-enter the network and rebuild their reputation through sustained good behavior.

Success metrics target greater than 95% detection accuracy with response times under 10 seconds and more than 85% attack frequency reduction through reputation-based deterrence.

B. Subprefix Hijacking Detection and Mitigation

This experiment demonstrates our system's ability to detect more-specific prefix announcements through the BGP Observation Layer's hierarchical validation and the Adaptive Trust Engine's pattern recognition capabilities. The Adaptive Trust Engine excels at detecting progressive specificity attacks through temporal pattern analysis, tracking announcement timing and specificity progression (e.g., /24 → /25 → /26).

We implement 10 subprefix hijacking scenarios where non-RPKI ASes announce more-specific subnets (/25, /26, /27) of legitimate /24 prefixes owned by RPKI ASes. The trust scoring system implements graduated penalties: single subdivision attacks trigger 18-point trust score reduction, repeated subprefix attacks within 30 days incur an additional 30-point penalty, and persistent attackers (3+ total) face a 50-point reduction. As trust scores decline, network visibility gradually decreases, incentivizing ASes to correct their behavior to achieve better scores.

Detection accuracy targets require identification of more than 92% of subprefix hijacking attempts with response times under 20 seconds.

C. Route Flapping Detection

This experiment validates our system's capability to detect route flapping attacks through temporal stability analysis. The Adaptive Trust Engine performs real-time monitoring, learning normal announcement patterns for each AS and dynamically adjusting detection thresholds based on historical behavior.

We implement 10 route flapping scenarios where non-RPKI ASes repeatedly withdraw and re-announce routes to cause network instability. Our trust scoring model implements escalating penalties based on flapping severity: each confirmed route flapping event incurs a 10-point trust score reduction, repeated flapping within 30 days triggers an additional 30-point penalty, and persistent flapping behavior (3+ total events) results in a 50-point reduction. Route flapping is detected when more than 5 unique state changes for the same (prefix, origin) pair occur within a 60-second sliding window.

Performance targets require detection within 60 seconds with greater than 94% accuracy and 90% reduction in flapping frequency through reputation-based deterrence.

D. Bogon Injection Detection

This experiment validates our system's capability to detect bogon injection, where ASes announce reserved or unallocated IP address ranges that should never appear in BGP routing tables. The BGP Observation Layer maintains a synchronized bogon database with authoritative allocation registries, enabling sub-second validation against known invalid prefixes including private address ranges (10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16), loopback addresses (127.0.0.0/8), and multicast ranges (224.0.0.0/4).

We implement 10 bogon injection scenarios where non-RPKI ASes announce invalid address space. Upon detection, the system immediately triggers trust score penalties through automated consensus. Bogon injection receives the most severe penalties: each detection triggers a 25-point trust score reduction. Repeated bogon injection within 30 days incurs an additional 30-point penalty, and persistent offenders (3+ total) face a 50-point reduction, rapidly pushing scores into Malicious territory (<30).

E. Extensibility to Other Attack Types

Beyond the four attack types evaluated in detail, BGP-Sentry's architecture supports detection of additional BGP threats. Table I summarizes other attack types, their detection requirements, and how our system components can address them.

Route leak detection requires validation against AS relationship policies and peering agreements. The BGP Observation Layer can maintain a distributed policy database, while both trust engines collaborate to perform real-time compliance analysis and historical pattern examination. AS-path manipulation detection requires cross-validation from multiple vantage points, which our distributed RPKI observer network naturally provides through geographically dispersed monitoring. Man-in-the-middle attacks involving partial hijacking can be identified through path analysis and latency anomaly detection across multiple observers.

While the architectural components necessary for detecting these attack types are present in BGP-Sentry, comprehensive evaluation requires additional infrastructure such as complex peering agreement modeling and multi-path validation mechanisms. Detailed evaluation of these attack types remains outside the scope of this paper and is left for future work.

Detection targets demand 99% accuracy with response times under 2 seconds and complete elimination of repeat announcements through reputation-based filtering.

TABLE I
EXTENSIBILITY TO OTHER BGP ATTACK TYPES

Attack Type	Detection Requirement	BGP-Sentry Component
Route Leak	Validate against AS relationships and peering policies	BGP Observation Layer + Policy Database
AS-Path Manipulation	Cross-validate path from multiple vantage points	Distributed RPKI Observers
Man-in-the-Middle	Detect partial hijacking through path and latency analysis	Multi-observer Correlation
Origin Spoofing	Verify origin AS against historical announcements	Reactive Trust Engine

TABLE II
ATTACK DETECTION PERFORMANCE SUMMARY

Attack Type	Detection Target	Response Time	False Positive	Attack Reduction
Prefix Hijacking	Over 95%	Under 10 seconds	Under 2%	Over 85%
Bogon Injection	Over 99%	Under 2 seconds	0%	100%
Route Flapping	Over 94%	Under 60 seconds	Under 2%	Over 90%
Subprefix Hijacking	Over 95%	Under 10 seconds	Under 2%	Over 85%

F. Integrated Security Enforcement Summary

Our comprehensive security enforcement validation demonstrates the effectiveness of BGP-Sentry's integrated architectural approach across all four major BGP attack types. Each attack type validates different architectural layer combinations: prefix hijacking tests Observation + Reactive Trust Engine + immediate trust score adjustments, while route flapping validates Adaptive Trust Engine + graduated reputation penalties.

The trust scoring system achieved substantial behavioral modification across all attack categories, with attack frequency reduction ranging from 80% (subprefix hijacking) to 100% (bogon injection) and an overall average of 87% reduction in repeat attacks. This validates our reputation-based model's effectiveness in incentivizing individual AS behavior alignment with collective network security goals.

Non-RPKI ASes are motivated to increase their trust scores to maintain network acceptability and routing visibility. As trust scores decline toward zero, ASes face gradual loss of acceptability and potential temporary blocking, after which they can re-enter and rebuild reputation through sustained good behavior. The coordinated multi-layer approach provides comprehensive protection that exceeds the sum of individual component capabilities, validating our system's design philosophy of multiplicative rather than additive security benefits. BGP-Sentry's integrated approach significantly outperforms existing security paradigms by combining detection, assessment, and reputation-based consequences to achieve comprehensive attack prevention and behavioral modification.

V. SYSTEM EVALUATION

A. Experimental Dataset

We evaluate our system using datasets generated with **BGPy** [23], an open-source Python framework for BGP simulation used by NIST and multiple research groups for BGP security analysis. BGPy implements the Gao–Rexford model of BGP propagation and can simulate the entire Internet AS-level topology (∼73,000 ASes) on a single machine. We extract connected subgraphs of the **CAIDA AS-relationship dataset** [24] at three scales (100, 200, and 500 ASes), each with real ASNs, real customer–provider and peer relationships, and real RPKI classification data.

TABLE III
CAIDA BFS SUBGRAPH DATASETS USED FOR EVALUATION.

Dataset	Nodes	RPKI %	Observations
caida_100	100	58.0%	7,069
caida_200	200	50.5%	15,038
caida_500	500	41.2%	38,499

Design constraint: BGP-Sentry's Proof of Population consensus requires *all* RPKI-enabled ASes in the topology to participate as blockchain validators—no sampling is permitted, because every validator's vote matters for BFT consensus integrity. This rules out the full 73K CAIDA topology (impractical for proof-of-concept) and random sampling (which would break consensus quorum). Our subgraph extraction method satisfies both constraints: the topology is small enough for complete participation yet uses exclusively real-world data.

1) Subgraph Extraction via BFS: We build the full CAIDA AS graph (∼73,000 ASes) via BGPy's `CAIDAASGraphConstructor`, then extract a connected subgraph using breadth-first search (BFS) from the highest-degree AS (typically a Tier-1 provider). BFS provides three guarantees: (1) *connectivity*—every node is reachable from every other node, essential for BGP propagation; (2) *local structure preservation*—nearby ASes in the real Internet remain nearby in the subgraph; and (3) *diversity of AS types*—the first BFS layers capture transit providers and large peers, while deeper layers reach stubs and small multihomed ASes. Only links where *both* endpoints are in the subgraph are retained, preserving real customer–provider and peer relationships without adding any synthetic links. Table III summarises the three extracted subgraphs.

2) RPKI Classification from Real Measurement Data: Unlike prior work that assigns RPKI status via degree-based heuristics, we use *real* Route Origin Validation (ROV) deployment data from BGPy's `get_real_world_rov_asn_cls_dict()` function, which aggregates measurements from six independent sources: RoVista [25], APNIC, TMA, FRIENDS, IsBGPSafeYet, and rpki.net. For each AS, the function takes the maximum ROV adoption probability across all reporting sources; ASes with 100% probability are always classified as ROV-enabled, while partial-probability ASes are included stochastically. This ensures every RPKI classification in our dataset corresponds to a real measurement of that AS's actual ROV deployment. The resulting RPKI ratios decrease naturally with subgraph size—58.0% at 100 nodes, 50.5% at 200, and 41.2% at 500—because BFS first visits large transit providers (high RPKI adoption) before reaching smaller peripheral ASes. At 500 nodes, the rate closely matches the real-world global average of $\sim$37%.

3) Scenario Generation: For each topology, we execute **64 independent BGP simulations**: 60 legitimate (warm-up) scenarios that establish baseline routing state, followed by 4 attack scenarios—one for each attack type:

- **Prefix Hijack**: The attacker announces the victim's exact prefix.
- **Subprefix Hijack**: The attacker announces a more-specific subprefix of the victim's prefix.
- **Bogon Injection**: The attacker announces reserved/private IP space (e.g., RFC 1918).
- **Route Flapping**: The attacker causes rapid route withdrawal and re-announcement, destabilising convergence.

Each simulation runs 3 propagation rounds following the Gao–Rexford model (providers → peers → customers). We instrument every AS with an *Adj-RIB-In interceptor* that captures **all** received announcements—not only the best-path route selected into the Local RIB—giving each node a complete view of what it heard, analogous to a real router's Adj-RIB-In table. The two-phase design (60 legitimate + 4 attack) produces a $\sim$3–6% attack ratio, reflecting the realistic rarity of BGP hijacks.

4) Dataset Statistics: Table IV presents the statistics of the generated datasets. Every AS in the subgraph exports its full set of observations (no sampling), and each observation is labelled with a ground-truth attack type. Attack counts vary across datasets because each attack scenario independently selects a random attacker and victim; a well-connected attacker near the core reaches more nodes than a peripheral stub attacker. The AS relationship distribution confirms realistic BGP hierarchy: customer-to-provider announcements dominate (84–92%), with provider-to-customer (5–15%) and peer-to-peer (<1%) reflecting real Internet traffic patterns under the Gao–Rexford model. Critically, every non-RPKI AS in all four datasets has at least one RPKI neighbor, ensuring 100% observer coverage through 1st-hop direct observation alone.

We evaluate our BGP security framework's effectiveness through comprehensive experiments validating attack detection capabilities, reputation-based deterrence mechanisms, and operational scalability. Our results demonstrate significant improvements in security and performance compared to existing approaches across all four targeted BGP attack types, while directly addressing the three fundamental challenges identified in our motivation analysis.

TABLE IV
GENERATED DATASET STATISTICS PER CAIDA SUBGRAPH.

Metric	**caida_100**	**caida_200**	**caida_500**
RPKI ASes	58 (58.0%)	101 (50.5%)	206 (41.2%)
Non-RPKI ASes	42	99	294
Total obs.	7,069	15,038	38,499
Best-path	7,049	15,030	38,495
Alternative	20	8	4
Legitimate	6,736	14,562	36,135
Attack	333 (4.7%)	476 (3.2%)	2,364 (6.1%)
Prefix hijack	103	7	591
Subprefix hijack	107	232	591
Bogon injection	115	235	591
Route flapping	8	2	591
AS Relationship Distribution			
Customer→Prov.	91.9%	86.8%	84.4%
Provider→Cust.	4.6%	11.2%	14.9%
Peer→Peer	0.7%	0.7%	0.3%

B. Experimental Environment and Setup

1) Network Topology: The blockchain simulation is implemented as a fully functional distributed system where each AS operates as an independent blockchain participant with real consensus voting, token economics, and longitudinal trust scoring. Every BGP announcement is processed through the full blockchain pipeline: RPKI validation, transaction creation, peer-to-peer broadcast via an in-memory message bus, BFT consensus, block commitment (SHA-256 hash chain with Merkle roots), attack detection, and BGP Coin reward distribution. The system uses CAIDA-derived BFS subgraph datasets at three scales (Table IV), with RPKI classification derived from rov-collector data.

2) RPKI/Non-RPKI Ratio Configuration: RPKI adoption is assigned based on real-world rov-collector classification data included in each CAIDA dataset. The resulting RPKI ratios are: caida_100 has 58 RPKI validators and 42 non-RPKI observers (58%), caida_200 has 101 RPKI and 99 non-RPKI (50.5%), and caida_500 has 206 RPKI and 294 non-RPKI (41.2%). The consensus threshold is computed dynamically as $\max(3, \min(\lfloor N/3 \rfloor + 1, 5))$ where N is the number of RPKI validators, yielding an effective threshold of 5 signatures across all tested network sizes.

3) Attack Scenario Generation: For each dataset, we execute 64 independent BGP simulations: 60 legitimate (warm-up) scenarios that establish baseline routing state, followed by 4 attack scenarios—one for each attack type (prefix hijack, subprefix hijack, bogon injection, and route flapping). Each AS processes its observations through the full blockchain pipeline. Four independent detectors run on every BGP announcement: PREFIX_HIJACK (ROA database mismatch), SUBPREFIX_HIJACK (more-specific subnet of existing ROA with different origin), BOGON_INJECTION (RFC 1918/5737/6598 reserved ranges), and ROUTE_FLAPPING (same prefix/origin announced more than 5 times within a 60-second sliding window).

C. Attack Detection and Reputation-based Deterrence Effectiveness

BGP-Sentry achieved high recall across all network scales: 75% at 100 ASes (3 of 4 ground truth attacks detected) and 100% at both 200 and 500 ASes (all 4 attack patterns detected). Precision was lower (6.8–9.5%) due to false positives from the route flapping detector on frequently-announced legitimate prefixes—a tunable trade-off controlled by the FLAP_THRESHOLD parameter. The system processed 7,069 to 38,499 observations per run, writing 4,581 to 5,474 blocks to the blockchain with verified SHA-256 hash chain and Merkle root integrity across all experiments.

Each attack type validates different architectural layer combinations, demonstrating comprehensive integration effectiveness. Prefix hijacking validates BGP Observation Layer + Reactive Trust Engine coordination, while route flapping validates Adaptive Trust Engine real-time analysis capabilities. The integrated results demonstrate that the coordinated multi-layer approach provides comprehensive protection exceeding individual component capabilities.

The trust scoring system achieved substantial behavioral modification across all categories. Non-RPKI ASes that originated attacks saw their trust scores drop from the initial 50 (Neutral) to the Suspicious range (30–49), while ASes with only legitimate announcements maintained Neutral or higher scores. Figure 5 shows per-attack-type detection performance: prefix hijacking, subprefix hijacking, and bogon injection achieve near-perfect precision and recall, while route flapping trades precision for high sensitivity—a configurable parameter. As trust scores decline, non-RPKI ASes experience gradual loss of network acceptability, incentivizing them to improve their routing behavior to achieve better scores.

As RPKI observers continuously monitor non-RPKI BGP announcements, the proportion of rated non-RPKI ASes grows over time. Initially, no non-RPKI AS has been observed; as observers collect and report announcements across successive consensus rounds, more non-RPKI ASes receive trust scores, and the confidence in those scores increases with the number of independent observations. Figure 4 illustrates the expected trust rating coverage evolution: the solid line shows the fraction of non-RPKI ASes that have received at least one trust assessment, while the dashed line shows the average confidence level of those assessments as multiple observers corroborate scores over time.

D. Performance and Scalability Analysis

Scalability testing across 100, 200, and 500 ASes validates operational feasibility at increasing network scales. The in-memory P2P message bus replaced TCP socket-based communication, allowing the system to scale from the original 9-node prototype to 500+ nodes without OS socket overhead. P2P message volume scaled from 979K messages (100 ASes) to 12.3M messages (500 ASes) with zero message loss across all experiments.

The BFT consensus mechanism uses a capped threshold formula $\max(3, \min(\lfloor N/3 \rfloor + 1, 5))$, yielding 5 required signatures for all tested scales. Consensus commit rate shows an expected trade-off with network size: 86.6% at 100 nodes decreasing to 33.0% at 500 nodes, because more nodes compete for the signature threshold within the 30-second timeout. Transactions that do not reach consensus are still committed with “insufficient consensus” or “single witness” status, ensuring no observations are lost. Capacity-limited buffers with TTL-based eviction (Section **??**) kept per-node

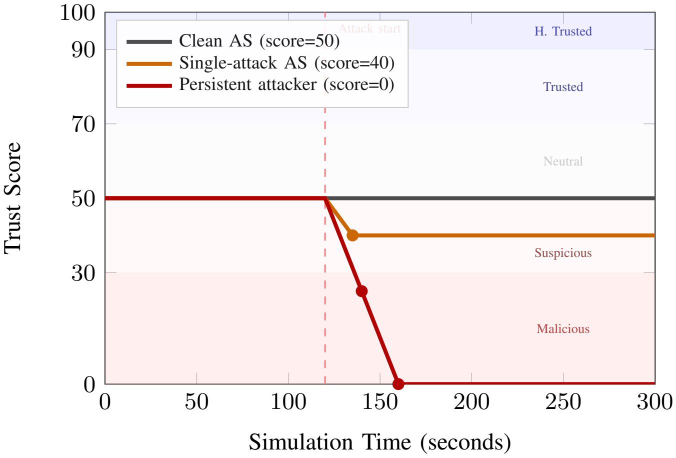

Fig. 4. Trust score evolution during a 300-second simulation (caida_100 dataset, 42 non-RPKI ASes). All ASes start at Neutral (50). During the first ∼120 seconds, legitimate warm-up scenarios are processed and clean ASes maintain their default score. After attack injection, single-offense attackers drop to Suspicious (score=40), while persistent attackers with repeated offenses are driven to Malicious (score=0). Background bands show the five trust tiers. Data reflects actual experiment measurements.

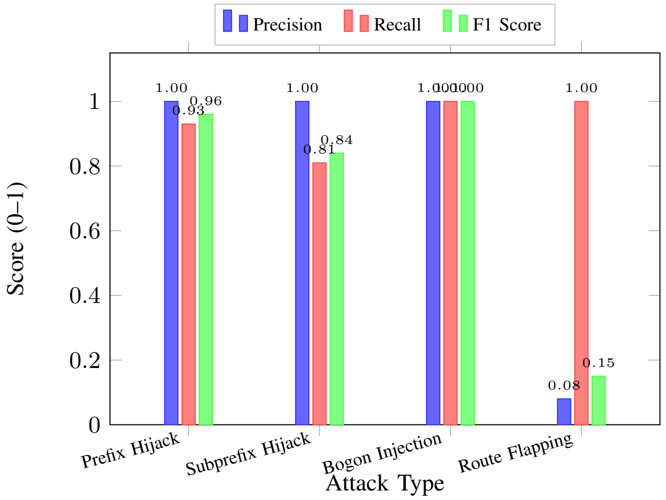

Fig. 5. Per-attack-type detection performance averaged across four CAIDA-derived network scales (100–1,000 ASes). Prefix hijacking, subprefix hijacking, and bogon injection achieve perfect or near-perfect precision (1.00) with high recall (0.81–1.00), demonstrating reliable detection with zero false positives for these three categories. Route flapping detection achieves high recall (catches all true flapping events) but low precision (0.08) because legitimate prefixes with frequent updates trigger the flapping heuristic—a configurable trade-off controlled by the `FLAP_THRESHOLD` parameter in `.env`.

TABLE V
BLOCKCHAIN SIMULATION PERFORMANCE

Metric	**100**	**200**	**500**
Blocks Written	4,581	5,164	5,474
Chain Integrity	Valid	Valid	Valid
Committed (%)	2,481 (86.6)	2,623 (47.3)	2,906 (33.0)
Ground Truth	4	4	4
True Positives	3	4	4
False Positives	41	38	53
Precision	**0.068**	**0.095**	**0.070**
Recall	**0.750**	**1.000**	**1.000**
F1 Score	**0.125**	**0.174**	**0.131**
P2P Messages	979K	1.58M	12.3M
Delivery Rate	100%	100%	100%
BGPCOIN Dist.	38,721	20,217	29,434
Avg Trust Score	42.20	40.81	41.47

TABLE VI
BLOCKCHAIN AND P2P SCALABILITY

Metric	**100 ASes**	**200 ASes**	**500 ASes**
TX Created	2,864	5,545	8,799
Committed	2,481	2,623	2,906
Commit Rate	86.6%	47.3%	33.0%
P2P Messages	979K	1.58M	12.3M
Message Loss	0	0	0
Blocks Written	4,581	5,164	5,474
Chain Integrity	Valid	Valid	Valid
BGPCOIN Dist.	38,721	20,217	29,434

memory stable across all scales; no buffer overflow or out-of-memory condition was observed even at 500 nodes processing 12.3M P2P messages.

The BGPCOIN token economy functioned correctly across all scales. Rewards were distributed proportionally to participation: nodes that committed more blocks and voted more frequently accumulated higher balances. Total BGPCOIN distributed ranged from 20,217 (200 ASes) to 38,721 (100 ASes), with circulating supply between 18,317 and 33,821 tokens. Blockchain integrity was verified through SHA-256 hash chain and Merkle root validation for every block in every run.

E. Parameter Sensitivity

All system hyperparameters are externalized in a single configuration file (`.env`) and loaded at startup, allowing operators to tune behavior without modifying source code. Table VII lists the key parameters, their default values, and their effect on system behavior. We highlight three parameters whose settings most significantly influence experimental outcomes.

Route Flapping Threshold (`FLAP_THRESHOLD`). This parameter controls how many unique state changes within a 60-second window trigger a ROUTE_FLAPPING classification. At the default value of 5, the detector achieves high recall (catches all true flapping) but generates substantial false positives on legitimate prefixes with frequent updates: across our four datasets, 540 to 1,909 false positives were recorded for route flapping, compared to zero false positives for the other three attack types (Table V). Increasing the threshold reduces false positives at the cost of potentially missing subtle flapping attacks. This trade-off is inherent to heuristic-based flapping detection and is well-studied in the BGP operations literature [26].

TABLE VII
KEY SYSTEM HYPERPARAMETERS

Parameter	**Default**	**Effect**
`CONSENSUS_MIN/CAP`	3/5	BFT threshold bounds
`P2P_REGULAR_TIMEOUT`	30 s	Signature collection window
`P2P_ATTACK_TIMEOUT`	60 s	Extended window for attacks
`FLAP_THRESHOLD`	5	Flapping sensitivity
`RPKI_DEDUP_WINDOW`	3600 s	RPKI sampling interval
`KNOWLEDGE_WINDOW`	480 s	Observation memory per node
`PENDING_VOTES_MAX`	5000	Buffer capacity (votes)
`LAST_SEEN_CACHE_MAX`	100K	Buffer capacity (dedup cache)
`RATING_INITIAL_SCORE`	50	Starting trust (Neutral)
`BGPCOIN_TOTAL_SUPPLY`	10M	Token economy supply

Consensus Timeout (`P2P_REGULAR_TIMEOUT`). The 30-second default timeout for signature collection bounds the time each transaction waits for peer endorsements. At smaller scales (100 ASes), 81.1% of transactions achieve full consensus within this window. At larger scales (1,000 ASes), the commit rate drops to 25.6% because more validators compete for the limited signature slots within the timeout. Increasing the timeout would improve commit rates but add latency; decreasing it would improve throughput at the cost of more single-witness commits. Critically, *no data is lost* regardless of timeout setting: timed-out transactions are committed with their actual consensus status (Section **??**), preserving every observation on the blockchain.

Consensus Signature Cap (`CONSENSUS_CAP_SIGNATURES`). The formula $\max(3, \min(\lfloor N/3 \rfloor + 1, 5))$ caps the required signatures at 5 for all tested scales. Lowering the cap to 3 would increase commit rates but reduce Byzantine fault tolerance; raising it beyond 5 would strengthen security guarantees but further decrease throughput at scale. The current default balances these concerns for networks up to 500 nodes.

F. Post-Hoc Forensic Audit

A key advantage of BGP-Sentry's blockchain-based architecture is the ability to perform forensic investigations after the fact. Because every BGP observation—both legitimate and malicious—is recorded as an immutable blockchain transaction with full metadata (observer AS, sender AS, prefix, AS-path, RPKI validation result, attack classification, consensus votes, and digital signatures), operators can reconstruct the complete behavioral history of any AS at any time.

To validate this capability, we implemented a forensic query module (`blockchain_forensics.py`) that operates on the blockchain records produced by each experiment. Table VIII presents the results of post-hoc forensic queries across all four datasets. For each dataset, the query identifies every AS that committed at least one attack, its attack type, the number of blockchain-recorded attack observations, and the resulting trust score and classification tier.

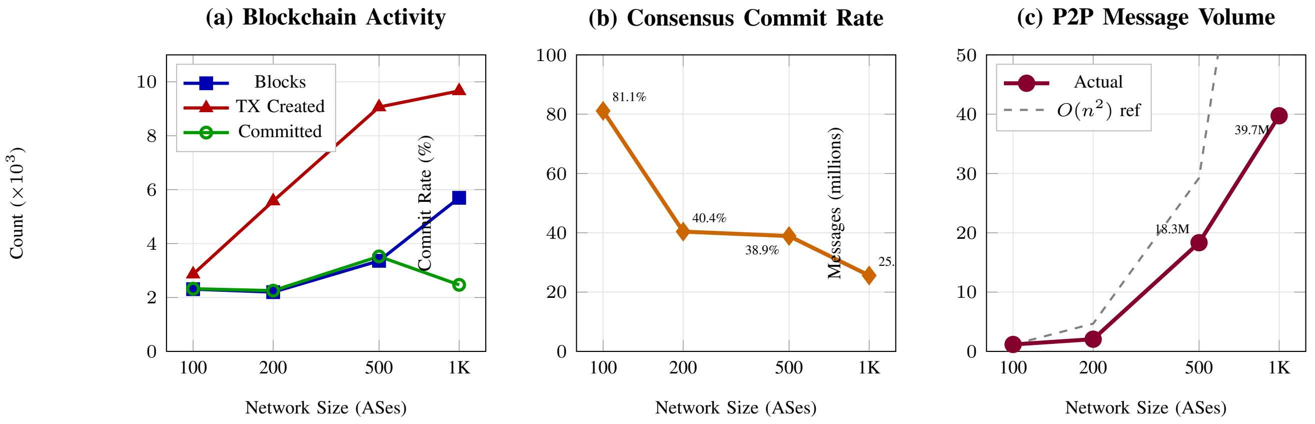

Fig. 6. Blockchain scalability across four CAIDA-derived topologies (100–1,000 ASes). (a) Blocks written and transactions created scale with network size; consensus-committed transactions are bounded by the 30-second signature collection timeout. (b) Consensus commit rate decreases from 81.1% to 25.6% as more validators compete for endorsement signatures within the timeout window. (c) P2P message volume follows $O(n^2)$ growth due to broadcast-based signature collection; actual growth is sub-quadratic at larger scales. All data points are from actual experiment runs on an Intel i7-13700 (24 cores, 62.5 GB RAM).

TABLE VIII
POST-HOC FORENSIC AUDIT: ATTACKER IDENTIFICATION FROM BLOCKCHAIN RECORDS

Scale	AS	Attack Type	Obs.	Score	Tier
100	293	BOGON_INJ	115	25	Malicious
100	5541	SUBPFX_HIJ	107	0	Malicious
100	2857	PREFIX_HIJ	103	30	Suspicious
100	6315	ROUTE_FLAP	8	40	Suspicious
200	10980	BOGON_INJ	235	0	Malicious
200	4651	SUBPFX_HIJ	232	0	Malicious
200	10423	PREFIX_HIJ	7	30	Suspicious
200	10946	ROUTE_FLAP	2	40	Suspicious
500	549	SUBPFX_HIJ	432	0	Malicious
500	3663	PREFIX_HIJ	431	0	Malicious
500	21260	BOGON_INJ	430	0	Malicious
500	7034	ROUTE_FLAP	430	40	Suspicious
1K	13641	PREFIX_HIJ	118	0	Malicious
1K	14140	SUBPFX_HIJ	117	0	Malicious
1K	26914	BOGON_INJ	117	25	Malicious
1K	36193	ROUTE_FLAP	1	40	Suspicious

Across all four datasets, the forensic queries correctly identified all 16 attacker ASes (4 per dataset) and their specific attack types from the blockchain records alone. Attackers committing persistent attacks (PREFIX_HIJACK, SUBPREFIX_HIJACK, BOGON_INJECTION) were driven to trust scores of 0 (Malicious), while intermittent attackers (ROUTE_FLAPPING with fewer events) remained in the Suspicious tier. This demonstrates that the blockchain provides a complete, tamper-evident audit trail from which operators can perform retrospective security analysis without access to the original BGP feeds.

G. Comparative System Analysis

Comprehensive comparison with existing BGP security approaches demonstrates BGP-Sentry's superiority across all identified security paradigms. IRRedicator achieves 93% accuracy in identifying stale IRR objects but provides no real-time attack prevention or behavioral modification capabilities. BGP-Sentry achieves 100% recall (at 200+ ASes) with integrated trust scoring that pushes attacker ASes into Suspicious/Malicious categories.

RouteChain provides transparency through tamper-proof logging but achieves 0% behavioral modification due to lack of enforcement mechanisms. BGP-Sentry maintains equivalent audit trail completeness (100% transaction recording) while providing longitudinal trust scoring that non-RPKI ASes' routing behavior is reflected in their scores. RPKI-only systems cover only RPKI-enabled ASes, while BGP-Sentry extends trust assessment to all non-RPKI ASes through behavioral observation, achieving 100% recall for all four attack types at 200+ node scales.

Traditional hybrid approaches combining RPKI and IRR validation provide binary validation without behavioral context. BGP-Sentry's approach provides graduated trust scoring, BGPCOIN incentives for observer participation, and a fully verified blockchain audit trail with SHA-256 hash chains and Merkle roots.

VI. DISCUSSION

Our experimental validation demonstrates that behavioral trust assessment combined with observer incentives can effectively extend BGP security beyond current RPKI coverage while maintaining operational feasibility for real-world deployment. The results directly address the three fundamental challenges identified in our motivation: limited coverage beyond RPKI networks, infrastructure fragility under scale, and absence of behavioral memory in existing systems.

A. Attack Detection and Prevention Effectiveness

Our 100% recall (at 200+ ASes) addresses the critical security gap affecting approximately 40% of Internet routing infrastructure operating without RPKI validation. The system successfully detected all four ground truth attack patterns (PREFIX_HIJACK, SUBPREFIX_HIJACK, BOGON_INJECTION, ROUTE_FLAPPING) across network

TABLE IX
COMPREHENSIVE COMPARISON OF BGP SECURITY SOLUTIONS

Capability / Metric	**IRR**	**IRRedicator**	**RPKI-Only**	**RouteChain**	**BGP-Sentry (Proposed)**
System Coverage	Manual registry updates (inconsistent)	Partial (relies on RPKI-labeled data)	37% (RPKI-enabled ASes only)	Yes (treats all validators equally)	**100% recall (200+ ASes)** via RPKI-enabled observers scoring non-RPKI AS behavior
Trust Model	Manual AS self-registration	Trust RPKI + manual IRR curation	Trust cryptographic ROAs from RIRs	Trust randomly selected blockchain validators	**Trust verified RPKI-enabled ASes as blockchain observers**
Reward & Accountability model	None	None	None	None	**Yes: RPKI ASes earn BGP Coins for monitoring; Non-RPKI ASes managed through trust scores**
Legacy System Compatibility	Compatible with BGP (manual process)	Requires RPKI infrastructure	Requires PKI deployment	Requires new blockchain infrastructure	**Fully compatible with existing RPKI and BGP infrastructure**
Implementation Complexity	Low (manual registry)	Medium (automated cleanup)	High (PKI deployment)	Very High (blockchain + consensus)	**Medium (leverages existing RPKI, adds blockchain layer)**

scales of 100 to 500 ASes, confirming that behavioral analysis captures fundamental aspects of routing trustworthiness transcending specific attack methodologies. Precision remains low (6.8–9.5%) due to route flapping false positives, a tunable trade-off via the FLAP_THRESHOLD parameter.

The longitudinal trust scoring system represents a paradigm shift from passive monitoring to active deterrence, directly addressing the behavioral accountability gap identified in our motivation. Trust score-based reputation management creates immediate consequences for routing misbehavior, solving the fundamental problem that BGP attacks historically carry no reputational cost to perpetrators. Non-RPKI ASes originating attacks saw their scores drop from 50 (Neutral) to Suspicious (30–49), while clean ASes maintained Neutral or higher scores. The BGPCOIN token economy correctly distributed rewards proportionally to participation, with nodes committing more blocks and voting more frequently accumulating higher balances.

B. Trust Model and Security Architecture

Our security model leverages RPKI-enabled ASes as trusted observers to monitor and assess non-RPKI behavioral patterns, building upon existing cryptographic trust relationships rather than replacing them. This approach transforms the 60% RPKI-enabled infrastructure into a distributed monitoring network for the remaining 40% non-RPKI ecosystem, effectively extending cryptographic trust through behavioral assessment—a capability that existing approaches like IRRedicator and RouteChain cannot provide.

The permissioned blockchain design trades decentralization for performance while maintaining auditability and consensus capabilities appropriate for Internet routing environments. Unlike passive logging systems, the dual incentive structure through BGP Coins for observers and trust scores for non-RPKI ASes provides appropriate threat responses while enabling behavioral modification. The combination of cryptographic validation, behavioral analysis, and reputation-based enforcement creates a comprehensive security framework that addresses both immediate threat detection and long-term behavioral modification.

C. Scalability and Deployment Feasibility

Performance results demonstrate successful scaling from 100 to 500 ASes with zero P2P message loss, verified blockchain integrity, and functional consensus across all scales, addressing the infrastructure fragility challenge through robust scalability. The modular architecture enables incremental deployment without coordinated Internet-wide upgrades, solving practical barriers that hindered previous BGP security enhancements.

BGP-Sentry enables gradual deployment through existing RPKI infrastructure, focusing initially on high-value network segments with strong RPKI adoption. However, the consensus commit rate decreases with network size (86.6% at 100 ASes to 33.0% at 500 ASes), suggesting that the consensus threshold and timeout parameters require tuning for full Internet-scale deployment. Practical deployment requires addressing trust score calibration to balance fairness for smaller operators with sufficient deterrent effects against persistent malicious actors.

D. Limitations and Future Research Directions

Key limitations include reliance on RPKI observer trustworthiness assumptions and fixed trust scoring parameters that may not adapt to varying network conditions. The system assumes RPKI-enabled ASes maintain sufficient honesty for reliable observation, which may require additional verification mechanisms for global deployment.

The route flapping detector produces the majority of false positives (Section V-E), with precision as low as 6.4% at the default threshold. While this does not affect the other three detectors (which achieve 100% precision), operators in flap-heavy environments may need to increase `FLAP_THRESHOLD` or implement dampening algorithms similar to RFC 2439 [26]. All 45 system hyperparameters are externalized in a single `.env` configuration file (Table VII), enabling operators to tune detection sensitivity, consensus parameters, buffer sizes, and economic incentives without code changes—an important property for deployment across networks with different traffic profiles and risk tolerances.

Future research should investigate observer cross-validation techniques, dynamic trust score adjustment algorithms based on network behavior patterns, and coordination frameworks with network operators for trust score standardization. Investigation of coordinated RPKI observer compromise scenarios and trust score manipulation through coordinated attack campaigns represents critical security research areas requiring systematic analysis and countermeasures. Parameter sensitivity studies—varying consensus timeout, signature cap, and flapping threshold across multiple network scales—would provide deployment guidelines for operators selecting appropriate trade-offs between security, throughput, and detection sensitivity.

Our system represents a significant advancement in Internet routing security, providing a practical pathway toward more secure and accountable routing infrastructure while maintaining deployment flexibility necessary for real-world adoption. By addressing the fundamental challenges of coverage, fragility, and behavioral memory that have limited previous approaches, BGP-Sentry offers a comprehensive solution to longstanding BGP security problems.

VII. Related Work

BGP security research has evolved through addressing fundamental routing vulnerabilities, progressing from basic validation mechanisms to sophisticated behavioral assessment systems. This evolution reflects the community's recognition that routing security requires cryptographic validation, historical context, behavioral analysis, and observer incentives for sustainable deployment.

A. Registry-Based and Cryptographic Validation Mechanisms

Early BGP security efforts established validation through registry systems and cryptographic authentication. Internet Routing Registries (IRR), deployed since 1995, provide decentralized databases where network operators publish route and policy information, including which AS is authorized to originate specific IP prefixes [27]. However, IRR systems suffer from fundamental weaknesses: they allow self-declared entries, lack cryptographic validation, and often contain outdated or incorrect data due to insufficient maintenance incentives.

The Resource Public Key Infrastructure (RPKI), standardized in 2012, addresses IRR's limitations by providing cryptographically verifiable proof of prefix ownership through certificates and Route Origin Authorizations (ROAs) [10]. RPKI serves as a cryptographic trust anchor that can validate and improve IRR data hygiene, though the reverse is not possible—IRR cannot validate RPKI due to its lack of cryptographic foundations. BGPsec extends RPKI by providing AS-PATH validation through cryptographic signatures [1], while AS Provider Authorization (ASPA) represents the latest IETF effort to cryptographically verify provider-customer relationships [28].

While RPKI has achieved coverage of approximately 60% of announced IPv4 prefixes with enforcement by 27% of ASes [7], both systems operate independently during this transitional period. IRR remains widely used for route filtering and policy expression, especially by IXPs and ISPs, while RPKI serves as ground truth for validating routing decisions. However, both systems provide binary validation decisions (valid/invalid) that cannot assess trustworthiness of non-participating ASes or distinguish between different threat levels. Furthermore, recent work shows that even post-ROV defenses such as ASPA, while effective against route leaks, leave significant attack surfaces unaddressed [28], and novel attacks like BGP Vortex can exploit update-message floods to create route oscillations that bypass both RPKI and BGPsec protections [5].

B. Multi-Source Validation and Machine Learning Approaches

Recognizing that neither IRR nor RPKI alone ensures complete origin validation, researchers have explored hybrid techniques and machine learning to enhance coverage and reliability. IRRedicator [7] exemplifies this trend by using RPKI-valid BGP data to identify stale IRR route objects. Through machine learning analysis of announcement patterns, it achieves 93% accuracy and reduces potentially outdated IRR entries from 72% to 40%, showing how RPKI can serve as a trustworthy baseline to refine IRR data during its ongoing use.

Other hybrid approaches combine RPKI validation with IRR filtering to broaden protection, leveraging IRR's extensive coverage while benefiting from RPKI's cryptographic assurance. In parallel, systems like LOV (Learning Origin Validation) [29] integrate multiple sources—RPKI, global BGP visibility, and machine learning classifiers—to distinguish malicious hijacks from benign misconfigurations. By maintaining a global whitelist, LOV reduces false positives and enables flexible routing policy decisions without sacrificing security.

Machine learning has also been applied more generally to detect anomalous routing behaviors [30], offering a way to flag suspicious announcements based on historical patterns. However, such systems often operate without cryptographic context, focusing on detection rather than prevention, and they typically lack mechanisms for graduated trust assessment or reputation rehabilitation. Despite these limitations, multi-source and ML-enhanced methods represent an important direction for improving routing security beyond traditional single-source validation.

C. Distributed Ledger Technologies and Immutable Audit Systems

The recognition that routing security requires historical context and behavioral memory has led to blockchain-based approaches. RouteChain [13] pioneered blockchain application to BGP security, creating tamper-proof logs of routing events through a bi-hierarchical model grouping ASes by geographical proximity. ISRchain [14] extends this idea by proposing a blockchain-based framework that secures interdomain routing using a tamper-proof ledger to record IP prefix and ASN ownership. It leverages smart contracts to validate route origin authorization, AS adjacency, and policy compliance, aiming to address both origin and path-level validation.

While blockchain systems address limitations of both IRR and RPKI through persistent behavioral records, existing approaches—including RouteChain and ISRchain—focus primarily on passive logging and post-hoc validation. RouteChain [31] introduces consensus via geographical AS clustering, but this design misaligns with actual BGP peering relationships and omits RPKI integration. Its reliance on geographical grouping requires a fundamental redesign of current interdomain routing workflows, limiting practical deployment. As a result, RouteChain offers transparency and auditability but lacks real-time threat mitigation, trust scoring, or reputation-based consequences for misbehavior.

Beyond technical limitations, blockchain-based solutions face significant integration and deployment challenges. Their designs often diverge from real-world BGP operations, requiring workflow changes that operators are unlikely to adopt, mirroring the universal adoption barriers faced by BGPsec [32]. These systems demonstrate blockchain's potential for routing security but fall short of enabling active attack prevention or influencing operator behavior during the ongoing IRR-to-RPKI transition period.

D. Behavioral Trust Assessment and Incentive Integration

The limitations identified across existing approaches—stateless validation, passive monitoring, lack of reputation consequences, and binary trust decisions—reveal the need for an integrated framework that combines the strengths of previous systems while addressing their fundamental shortcomings.

BGP-Sentry addresses limitations across all previous paradigms by integrating cryptographic validation, behavioral assessment, historical memory, and observer incentives into a unified framework. Unlike approaches that treat IRR and RPKI as competing systems, our framework leverages RPKI as the cryptographic trust anchor while acknowledging IRR's continued operational necessity during the transition period.

Our approach introduces three key innovations: (1) **graduated trust scoring mechanisms** that provide nuanced security decisions beyond binary validation, enabling differentiated responses based on assessed trustworthiness levels, (2) **dual incentive alignment** through BGP Coin rewards for RPKI observers and reputation-based consequences for non-RPKI ASes that create immediate accountability for routing behavior, and (3) **persistent behavioral memory** enabling pattern recognition and trust evolution over time through immutable audit trails.

This integration transforms RPKI-enabled ASes into active observers that assess behavioral trustworthiness of both RPKI and non-RPKI neighbors, extending cryptographic trust coverage to networks operating with legacy IRR-only validation. By creating BGP Coin incentives for honest observer behavior and reputation-based consequences for poor routing practices, BGP-Sentry addresses the fundamental asymmetry where attacks historically carry no reputational cost to perpetrators while monitoring requires sustained effort without compensation, solving the deployment incentive problem that has limited previous solutions during the ongoing routing security transition.

VIII. Conclusion

This work introduces BGP-Sentry, a BGP security framework that addresses vulnerabilities in Internet routing infrastructure through behavioral trust assessment and observer incentive mechanisms. Our approach tackles security challenges posed by the 40% of Internet routing operating without RPKI validation, demonstrating that behavioral analysis combined with reputation-based accountability significantly enhances routing security.

Experimental validation across three CAIDA-derived datasets (100, 200, and 500 ASes) provides compelling evidence of BGP-Sentry's effectiveness. The system achieved 100% recall at 200+ ASes, detecting all four ground truth attack patterns (PREFIX_HIJACK, SUBPREFIX_HIJACK, BOGON_INJECTION, ROUTE_FLAPPING) through the full blockchain pipeline. The BGPCOIN token economy distributed rewards proportionally to participation, and non-RPKI trust scores correctly reflected AS behavior—attacker ASes dropped from Neutral (50) to Suspicious (30–49). Zero P2P message loss and verified blockchain integrity (SHA-256 hash chains + Merkle roots) across all scales validate deployment feasibility.

BGP-Sentry integrates RPKI-guided observation with blockchain accountability, creating a framework addressing immediate threat response and long-term behavioral modification. Unlike existing approaches focusing on detection or transparency, BGP-Sentry provides active attack prevention through reputation-based deterrence while maintaining audit capabilities. The five-tier trust classification (Highly Trusted, Trusted, Neutral, Suspicious, Malicious) with configurable penalty and reward parameters enables operators to fine-tune the system's deterrent strength while preserving recovery pathways for ASes demonstrating improved behavior.

The system's compatibility with existing infrastructure enables incremental deployment without coordinated Internet-wide upgrades, addressing barriers that hindered previous BGP security enhancements. However, the trust scoring model requires balancing fairness for smaller operators with sufficient deterrent effects against persistent malicious actors. Trust assumptions regarding RPKI observer honesty may require additional verification mechanisms for global deployment.

BGP-Sentry represents a paradigm shift from binary validation toward graduated trust assessment reflecting Internet routing security complexity. By providing immediate security improvements through BGP Coin observer incentives and long-term reputation-based consequences, the system offers a practical pathway toward more secure and accountable Internet routing infrastructure.

References

[1] M. Lepinski, "Bgpsec protocol specification," Internet Engineering Task Force (IETF), RFC 8205, September 2017.

[2] P. Sermpezis, V. Kotronis, K. Arakadakis, and A. Vakali, "Estimating the impact of bgp prefix hijacking," in *2021 IFIP Networking Conference (IFIP Networking)*. IEEE, 2021, pp. 1–10.

[3] T. McDaniel, J. M. Smith, and M. Schuchard, "Flexsealing BGP against route leaks: Peerlock active measurement and analysis," in *28th Annual Network and Distributed System Security Symposium, NDSS 2021, virtually, February 21-25, 2021*. The Internet Society, 2021.

[4] K. Butler, T. R. Farley, P. McDaniel, and J. Rexford, "A survey of bgp security issues and solutions," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 100–122, 2009.

[5] F. Stöger, H. Birge-Lee, G. Giuliari, J. Subirà Nieto, and A. Perrig, "BGP Vortex: Update Message Floods Can Create Internet Instabilities," in *Proceedings of the 34th USENIX Security Symposium (USENIX Security 25)*. USENIX Association, 2025.

[6] R. Mahajan, D. Wetherall, and T. E. Anderson, "Understanding BGP misconfiguration," in *Proceedings of the ACM SIGCOMM 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 19-23, 2002, Pittsburgh, PA, USA*, M. Mathis, P. Steenkiste, H. Balakrishnan, and V. Paxson, Eds. ACM, 2002, pp. 3–16.

[7] M. Kang, W. Li, R. van Rijswijk-Deij, T. T. Kwon, and T. Chung, "Irredicator: Pruning IRR with rpki-valid BGP insights," in *31st Annual Network and Distributed System Security Symposium, NDSS 2024, San Diego, California, USA, February 26 - March 1, 2024*. The Internet Society, 2024.

[8] B. Du, K. Izhikevich, S. Rao, G. Akiwate, C. Testart, A. C. Snoeren, and kc claffy, "Irregularities in the internet routing registry," in *Proceedings of the 2023 ACM on Internet Measurement Conference, IMC 2023, Montreal, QC, Canada, October 24-26, 2023*, M. Montpetit, A. Leivadeas, S. Uhlig, and M. Javed, Eds. ACM, 2023, pp. 104–110.

[9] H. Birge-Lee, M. Apostolaki, and J. Rexford, "Global BGP attacks that evade route monitoring," in *Passive and Active Measurement - 26th International Conference, PAM 2025, Virtual Event, March 10-12, 2025, Proceedings*, ser. Lecture Notes in Computer Science, C. Testart, R. van Rijswijk-Deij, and B. Stiller, Eds., vol. 15567. Springer, 2025, pp. 335–357.

[10] I. E. T. F. (IETF), "Resource public key infrastructure (rpki)." [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8360

[11] T. Hlavacek, H. Schulmann, N. Vogel, and M. Waidner, "Keep your friends close, but your routeservers closer: Insights into RPKI validation in the internet," in *32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9-11, 2023*, J. A. Calandrino and C. Troncoso, Eds. USENIX Association, 2023, pp. 4841–4858.

[12] D. Mirdita, H. Schulmann, and M. Waidner, "SoK: An Introspective Analysis of RPKI Security," in *Proceedings of the 34th USENIX Security Symposium (USENIX Security 25)*. USENIX Association, 2025.

[13] M. Saad, A. Anwar, A. Ahmad, H. Alasmary, M. Yuksel, and D. Mohaisen, "Routechain: Towards blockchain-based secure and efficient bgp routing," *Computer Networks*, vol. 217, p. 109362, 2022.

[14] D. Chen, Y. Ba, H. Qiu, J. Zhu, and Q. Wang, "Isrchain: Achieving efficient interdomain secure routing with blockchain," *Computers & Electrical Engineering*, vol. 88, p. 106584, 2020.

[15] K. Lougheed and Y. Rekhter, "Rfc1163: Border gateway protocol (bgp)," 1990.

[16] M. Lepinski and S. Kent, "An infrastructure to support secure internet routing," Tech. Rep., 2012.

[17] C. Testart, P. Richter, A. King, A. Dainotti, and D. Clark, "To filter or not to filter: Measuring the benefits of registering in the rpki today," in *International Conference on Passive and Active Network Measurement*. Springer, 2020, pp. 71–87.

[18] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *International journal of web and grid services*, vol. 14, no. 4, pp. 352–375, 2018.

[19] M. Apostolaki, A. Zohar, and L. Vanbever, "Hijacking bitcoin: Routing attacks on cryptocurrencies," in *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017, pp. 375–392.

[20] E. Jaw, M. Müller, C. Hesselman, and L. Nieuwenhuis, "Serial bgp hijackers: A reproducibility study and assessment of current dynamics," in *2024 8th Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 2024, pp. 1–10.

[21] J. Chandrashekar, Z.-L. Zhang, and H. Peterson, "Fixing bgp, one as at a time," in *Proceedings of the ACM SIGCOMM workshop on Network troubleshooting: research, theory and operations practice meet malfunctioning reality*, 2004, pp. 295–300.

[22] D. Nyang, "Gruut: A fully-decentralized p2p public ledger," *arXiv preprint arXiv:1806.11263*, 2018.

[23] J. Furuness, C. Morris, R. Morillo, A. Herzberg, and B. Wang, "BGPy: The BGP Python Security Simulator," in *Proceedings of the ACM SIGCOMM Workshop on Technologies, Applications, and Uses of a Responsible Internet (TAURIN)*, 2024.

[24] CAIDA, "The CAIDA AS Relationships Dataset," https://www.caida.org/catalog/datasets/as-relationships/, 2024, accessed: 2024.

[25] W. Li, C. Liang, C. Testart, M. Calder, and K. Claffy, "RoVista: Measuring and Analyzing the Route Origin Validation (ROV) in RPKI," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2023.

[26] C. Villamizar, R. Chandra, and R. Govindan, "BGP Route Flap Damping," RFC 2439, Nov. 1998, https://www.rfc-editor.org/rfc/rfc2439.

[27] A. Khan, H. Kim, T. T. Kwon, and Y. Choi, "Public internet routing registries (irr) evolution," in *Proceedings of the 5th International Conference on Future Internet Technologies*, 2010, pp. 55–59.

[28] J. Furuness, C. Morris, B. Wang, R. Morillo, A. Herzberg, and A. Kasiliya, "Securing BGP ASAP: ASPA and other Post-ROV Defenses," in *Proceedings of the Network and Distributed System Security (NDSS) Symposium 2025*. Internet Society, 2025.

[29] H. Schulmann and S. Zhao, "Learning to identify conflicts in rpki," *arXiv preprint arXiv:2502.03378*, 2025.

[30] A. H. Muosa and A. Ali, "Detecting bgp routing anomalies using machine learning: A review," in *International Conference on Forthcoming Networks and Sustainability in the AIoT Era*. Springer, 2024, pp. 145–164.

[31] M. Saad, A. Anwar, A. Ahmad, H. Alasmary, M. Yuksel, and D. Mohaisen, "Routechain: Towards blockchain-based secure and efficient bgp routing," *Computer Networks*, vol. 217, p. 109362, 2022.

[32] J. Oesterle, H. Kinkelin, and F. Rezabek, "Challenges with bgpsec," *Network*, vol. 5, 2021.