DOI: 10.18372/2310-5461.69.20735
УДК 004.72.056.52

M. Odarchenko,
State University «Kyiv Aviation Institute»
orcid.org/0000-0002-7714-3558
e-mail: odarchenko.m.s@gmail.com;

M. Zaliskyi, Dr. Sc., Prof.
State University «Kyiv Aviation Institute»
orcid.org/0000-0002-1535-4384
e-mail: maximus2812@ukr.net

DETECTION OF ARTIFICIALLY INFLATED TRAFFIC IN A2P SMS USING PREFIX-BASED TIME-WINDOW ANALYSIS

Introduction

Application-to-Person (A2P) SMS remains one of the most widely used and trusted communication channels for authentication, financial transactions, and public safety services. As of 2024, mobile messaging is used by more than 5.5 billion subscribers worldwide [1]. Despite the rapid growth of internet-based messaging platforms, SMS continues to offer unmatched reach and interoperability, operating across approximately 1,000 mobile network operators in over 230 countries [2], [3]. These characteristics make SMS the default channel for One-Time Passwords (OTP), transactional notifications, and critical service alerts. Reflecting this role, the global A2P messaging market is projected to reach USD 78 billion by 2027, driven primarily by security-sensitive enterprise use cases [4].

The economic importance of A2P messaging has, however, made it an increasingly attractive target for fraud. Historically, SMS-related fraud was largely associated with signaling-layer vulnerabilities, including SS7 and Diameter exploitation, spoofing, and large-scale spam distribution [5], [6]. These threats motivated the widespread deployment of network-level SMS firewalls and signaling protection mechanisms.

In recent years, a new and significantly more costly fraud vector has emerged–Artificially Inflated Traffic (AIT), also commonly referred to as SMS pumping fraud. Unlike traditional attacks, AIT does not exploit weaknesses in signaling protocols. Instead, attackers deliberately trigger large volumes of OTP or transactional messages toward MSISDNs under their control or toward temporarily provisioned number ranges. From a technical and routing perspective, such traffic appears entirely legitimate: messages originate from authorized enterprise systems, are addressed to valid mobile numbers, and generate normal delivery receipts and timestamps. Moreover, the traffic often aligns with legitimate business workflows such as authentication, account verification, or promotional outreach (Fig. 1).

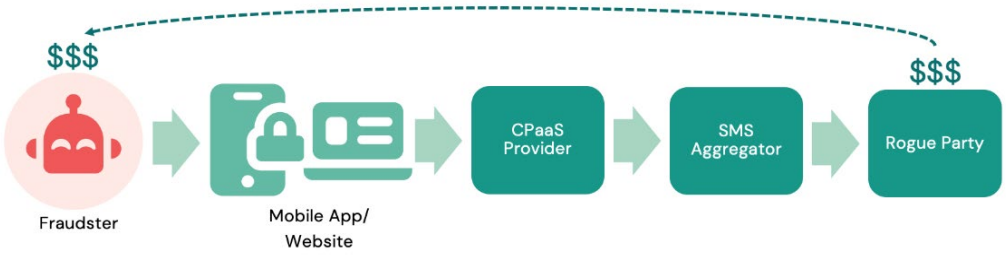

Fig. 1. A2P SMS Artificial inflation of traffi

Industry reports estimate that up to 19 % of global A2P OTP traffic may exhibit artificially inflated characteristics, resulting in annual enterprise losses exceeding USD 1.16 billion [7], [8]. According to the Mobile Ecosystem Forum (MEF), AIT was the fastest-growing category of A2P fraud in 2023 [9]. These figures highlight the scale of the problem and its growing impact on enterprise messaging economics.

A fundamental challenge in addressing AIT is its strong resemblance to legitimate high-volume messaging campaigns. Conventional SMS firewalls typically rely on signature-based detection, origin valida-

tion, or signaling anomalies, all of which provide limited protection when fraudulent traffic is generated through valid CPaaS interfaces. At the same time, enterprise-level fraud prevention systems often lack visibility into telecom-specific indicators such as MSISDN prefix concentration, delivery velocity, or short-term burst behavior.

Because real-time mitigation must avoid disrupting critical authentication and user-facing flows, any effective AIT detection mechanism must satisfy several constraints. It should operate without inspecting message content, adapt its thresholds to time-of-day usage patterns, distinguish between OTP and marketing traffic, and minimize false positives during legitimate promotional surges or seasonal events This paper proposes a lightweight, operator-agnostic method for detecting AIT based on statistical traffic behavior rather than content inspection. The approach combines fixed and rolling five-minute time windows, adaptive thresholds for daytime and nighttime activity, prefix-level MSISDN clustering (e.g., last-three-digit variability), and a temporary quarantine mechanism that blocks suspicious ranges for 24 hours to prevent economic escalation. The method is designed to be deployable at both the CPaaS aggregation layer and the enterprise perimeter, without introducing additional privacy risks.

While the proposed model demonstrates strong effectiveness against high-intensity AIT attacks, it also reveals limitations when applied to legitimate marketing traffic. As a result, future research should focus on integrating content classification and campaign fingerprinting techniques to further reduce false positives and improve differentiation between benign and adversarial traffic patterns.

Overall, this study contributes a practical and operationally feasible approach to securing enterprise A2P messaging flows, in alignment with guidance from the GSMA Fraud and Security Group (FASG) and prevailing CPaaS industry practices.

Analysis of recent research and publications

The continued expansion of A2P messaging underscores both its strategic importance and the growing financial risks associated with fraud. Recent industry analyses indicate that the global A2P SMS market exceeded USD 71.5 billion in 2024 and is projected to reach USD 96.73 billion by 2030, with an estimated compound annual growth rate of approximately 5.4 % [10]. Regional assessments show similar trends; for example, the North American A2P market alone was valued at USD 18.07 billion in 2024 and is expected to grow steadily throughout the decade [11]. Comparable estimates from Markets and Markets place the global market at USD 70.7 billion in 2023, with growth to USD 84.8 billion by 2029 [12]. Collectively, these figures reflect not only increasing enterprise adoption of A2P messaging but also a proportional rise in exposure to fraud-driven traffic.

Historically, both academic research and industry countermeasures focused on vulnerabilities at the signaling and transport layers of SMS networks. Well-documented threats included SS7 and Diameter exploitation, SMS spoofing, spam campaigns, and grey-route bypassing. In response, network-centric security frameworks and standards–such as GSMA FS.11 and FS.19 – significantly strengthened defenses against routing manipulation and inter-operator signaling abuse [13]. While these mechanisms have proven effective in protecting core network infrastructure, they offer limited protection against fraud that exploits legitimate enterprise messaging workflows.

More recently, Artificially Inflated Traffic (AIT)–often referred to as SMS pumping fraud–has emerged as one of the most damaging fraud categories within the A2P ecosystem. Unlike traditional attacks, AIT does not rely on protocol weaknesses. Instead, attackers exploit automated OTP, verification, or transactional messaging processes to generate large volumes of legitimate-looking SMS messages sent to MSISDNs under their control. Industry studies estimate that AIT accounts for a substantial share of global A2P fraud losses, with annual damages reaching hundreds of millions of USD [14]. The Mobile Ecosystem Forum (MEF) and leading CPaaS providers now identify AIT as the fastest-growing form of A2P fraud [15].

Prior research and operational analyses have also highlighted important behavioral differences between OTP traffic and marketing or promotional messaging. OTP traffic is typically transaction-driven, low in volume, and highly sensitive to delivery latency, whereas marketing traffic is batch-oriented, high-volume, and often scheduled within predefined time windows [16]. Although these distinctions are well understood conceptually, they are rarely translated into concrete detection rules or threshold-based models in the open literature.

In particular, existing academic and technical publications seldom provide explicit guidance on how detection thresholds should differ between OTP and marketing flows, how traffic patterns vary between daytime and nighttime usage, or how prefix-localized bursts can be systematically identified using delivery metadata alone. Similarly, while vendor whitepapers frequently reference techniques such as real-time anomaly detection, prefix intelligence, or API rate limiting, these descriptions are typically qualitative and lack transparent parameterization or reproducible evaluation methodologies [16], [17].

As a result, there remains a notable gap in publicly available, benchmarkable approaches for detecting AIT using only non-content metadata–specifically timestamps and MSISDN distribution patterns—

while preserving data privacy and ensuring deployability across heterogeneous CPaaS environments. Addressing this gap motivates the methodological contribution presented in this paper.

Problem statement

Artificially Inflated Traffic (AIT), also known as SMS pumping fraud, is increasingly recognized as one of the most financially damaging threats in the A2P messaging ecosystem [14], [15]. Unlike traditional SMS fraud, AIT does not rely on exploiting signaling-layer vulnerabilities. Instead, attackers abuse fully legitimate CPaaS access mechanisms – such as REST APIs or Short Message Peer-to-Peer Protocol (SMPP) connections – to generate large volumes of messages, most commonly OTP or transactional SMS, toward MSISDNs under their control. Because these messages are successfully delivered and billed immediately, financial losses accumulate rapidly and can reach tens or even hundreds of thousands of USD within a single incident [14].

From a technical perspective, AIT traffic is difficult to distinguish from legitimate enterprise messaging. Fraudulent messages typically share the same sender identities, routing paths, and delivery characteristics as genuine traffic. They are addressed to valid MSISDNs, generate normal delivery receipts, and exhibit high delivery success rates. As a result, AIT easily bypasses network-edge SMS firewalls, which are primarily designed to detect signaling manipulation or protocol abuse [13].

At the application level, fraud detection systems often lack visibility into telecom-specific indicators such as MSISDN prefix concentration, delivery velocity, or short-term burst behavior. This creates a critical visibility gap at the CPaaS aggregation layer, where most A2P traffic is consolidated and where both enterprise and network-level signals must be interpreted together [10–12].

A second major challenge lies in the risk of false positives. Legitimate marketing and promotional campaigns frequently exhibit traffic patterns that resemble AIT: they generate high message volumes over short periods, target subscribers within shared number ranges, and are often launched simultaneously according to commercial schedules. These characteristics closely mirror those observed in AIT attacks [16], [17].

If detection relies solely on simplistic volume or rate-based thresholds, legitimate campaigns may be incorrectly blocked. Such false positives directly affect enterprise revenue, reduce customer engagement, and erode trust in CPaaS reliability. Consequently, any practical AIT detection mechanism must balance sensitivity to fraud against the operational cost of misclassification.

In previous research, the authors introduced the Intelligent Guaranteed Delivery Performance (IGDP) and Price–Delivery Gap (PDG) as metrics for evaluating the efficiency and economic performance of messaging delivery systems. IGDP reflects the proportion of messages that reach and benefit real end users, while PDG quantifies the economic loss associated with unsuccessful or wasteful message delivery.

AIT negatively affects both metrics. Fraudulent messages generate cost without providing user value, thereby reducing IGDP, while simultaneously increasing enterprise expenditure and widening the PDG. From this perspective, AIT mitigation is not merely a fraud-prevention task but also a mechanism for improving delivery quality and optimizing messaging economics across CPaaS platforms.

For an AIT detection solution to be deployable in real-world environments, it must satisfy several operational constraints. Detection should be content-agnostic to preserve privacy and enable broad applicability across industries. Latency must be low to limit financial exposure during rapid OTP bursts. Thresholds should adapt to day and night usage patterns to account for natural variations in traffic intensity. Blocking actions must be applied at a prefix or range level to avoid unnecessary disruption to unrelated users. Finally, the solution must scale to CPaaS environments that process billions of messages per day across thousands of enterprise customers.

These constraints significantly limit the applicability of many machine-learning approaches that require message content access, extensive training data, or high-latency inference.

The objective of this study is to design and evaluate a transparent, threshold-based statistical model that addresses these challenges. Specifically, the proposed approach aims to detect AIT through high-density, prefix-localized traffic concentrations within short time windows; apply adaptive thresholds that account for time-of-day effects; differentiate OTP-like behavior from marketing-like traffic; and enforce an automated 24-hour quarantine on suspicious ranges. The evaluation further quantifies exposure before detection, blocked traffic volume after mitigation, and residual false-positive risk for legitimate campaigns.

The resulting model is intended as a standardized, reproducible baseline for CPaaS providers and enterprises, and as a foundation for future extensions incorporating content-aware analysis and statistical learning techniques.

Materials and methods

This section describes the datasets, analytical assumptions, and detection logic used to evaluate Artificially Inflated Traffic (AIT) in A2P messaging. The methodology is intentionally designed to be transparent, lightweight, and deployable in real-time environments, such as CPaaS platforms or enterprise messaging gateways (Fig. 2).

The study is based on multiple real-world datasets collected from operational A2P messaging environments. These datasets include two confirmed AIT incidents and one large-scale legitimate marketing campaign used to evaluate false-positive behavior. Each dataset contains only delivery metadata–specifically message timestamps and destination MSISDNs. No message content, sender identifiers, or personally identifiable user data were used.

This design choice ensures compliance with data-protection principles and reflects realistic deployment conditions, where content inspection may be restricted or undesirable. All datasets were analyzed in UTC time to maintain consistency across regions and traffic profiles.

AIT attacks typically concentrate traffic within narrow MSISDN ranges, often targeting sequential or closely related numbers. To capture this behavior, destination MSISDNs are grouped into prefix-level ranges by removing the last three digits of the number. Each resulting prefix therefore represents a range of up to 1,000 MSISDNs. This level of aggregation balances sensitivity and stability: it is granular enough to reveal localized traffic bursts while avoiding excessive fragmentation that could obscure statistically meaningful patterns. Prefix-based aggregation also aligns with operational blocking practices commonly used by CPaaS providers and mobile operators.

Traffic behavior is analyzed using two complementary windowing strategies (Fig. 3):

- Fixed windows, where traffic is aggregated into consecutive, non-overlapping five-minute intervals.
- Rolling windows, where a five-minute window advances in one-minute steps.

Fixed windows provide stable aggregation and are effective at capturing sustained bursts of activity. Rolling windows, by contrast, allow earlier detection of rapidly emerging anomalies. Both approaches are evaluated to understand the trade-off between detection latency and the amount of traffic that can be blocked after detection.

All detection decisions are triggered at the end of the window in which thresholds are exceeded. A short enforcement delay is applied to reflect realistic operational conditions.

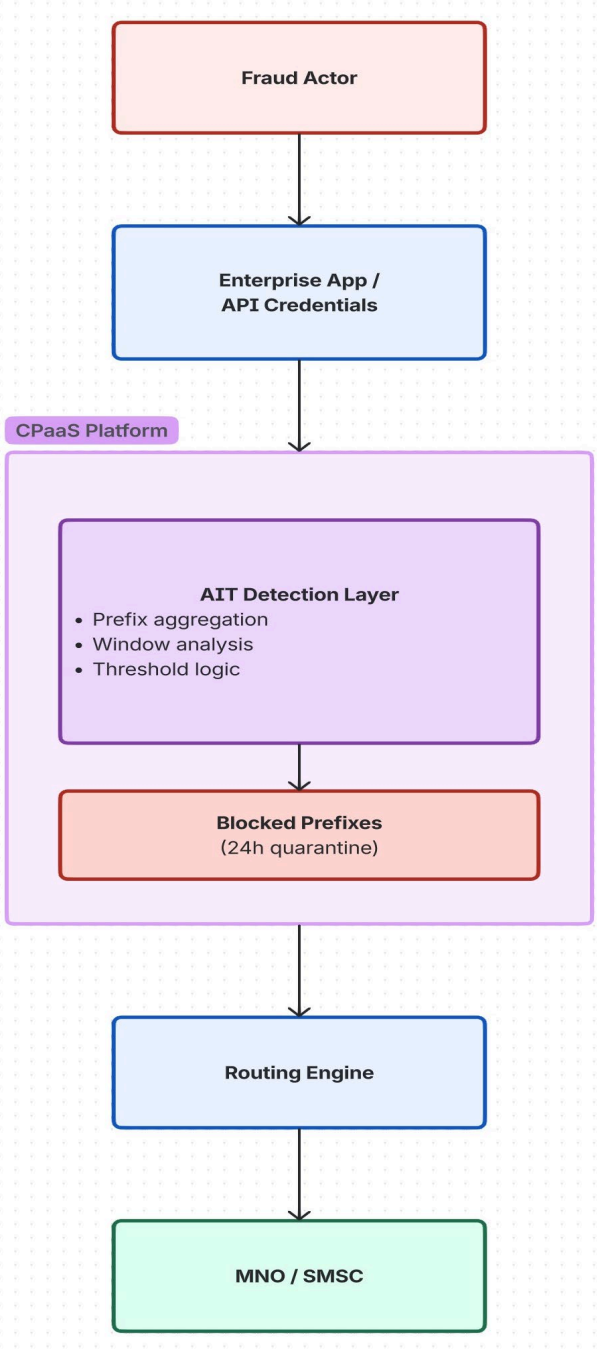

Fig. 2. Placement of the proposed AIT detection logic within a CPaaS platform. Detection occurs inline after API ingestion and before operator routing, enabling mitigation of artificially inflated traffic prior to billing and termination costs

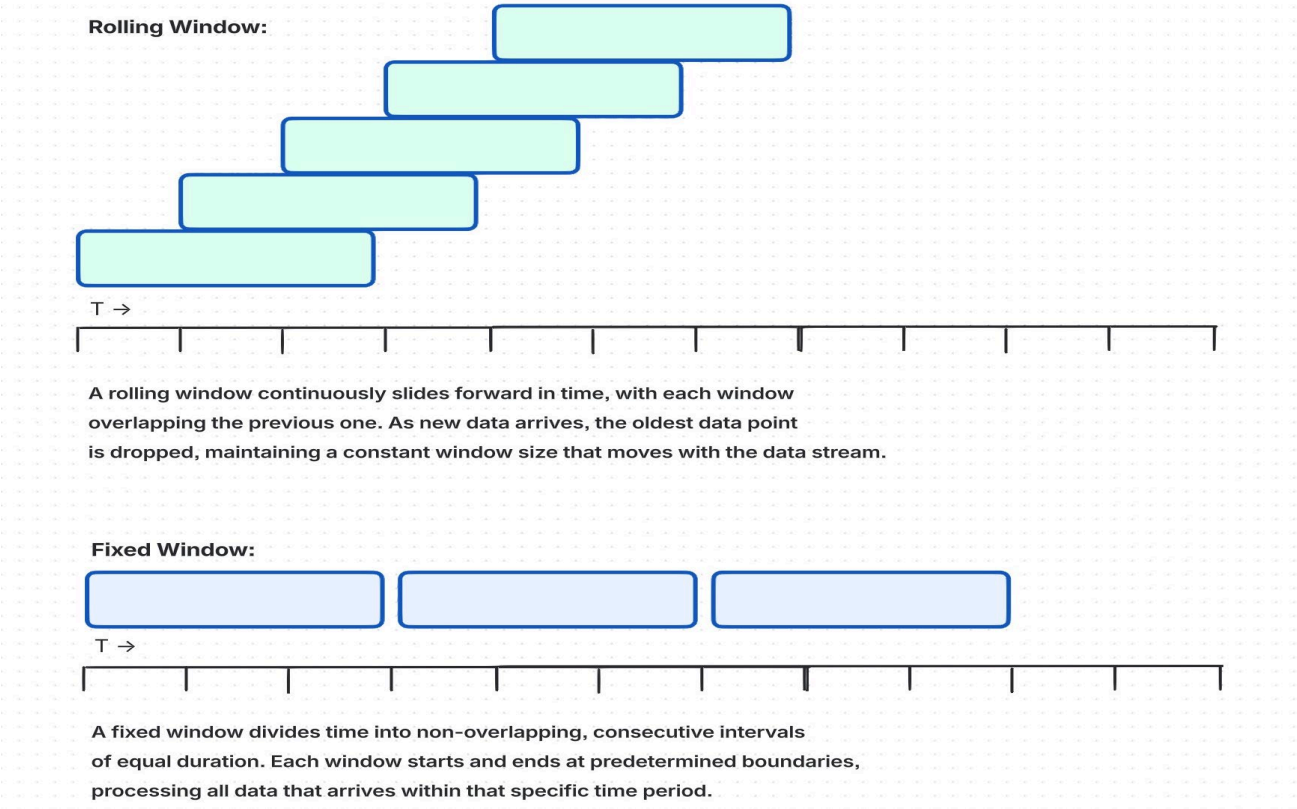

Fig. 3. Comparison of rolling and fixed time-window aggregation

Two primary detection metrics are considered:

1. Message count per prefix, which captures high-density bursts typical of AIT attacks.

2. Unique MSISDN count per prefix, which reflects recipient diversity and helps differentiate between repeated targeting of the same numbers and broad legitimate campaigns.

Message-count thresholds serve as the primary detection signal, while unique-recipient metrics are used to assess false-positive behavior and provide auxiliary validation. The study intentionally avoids more complex features in order to isolate the effectiveness of simple, interpretable rules.

Traffic intensity in A2P messaging varies significantly between daytime and nighttime hours. To acount for this, detection thresholds are adapted based on time of day. Lower thresholds are applied during low-activity periods, when anomalous bursts are more visible, while higher thresholds are used during peak hours to reduce false positives from legitimate campaigns.

This time-aware thresholding reflects common operational practices and improves robustness without introducing additional model complexity.

When a prefix exceeds the detection threshold, it is placed into a temporary quarantine state. All subsequent messages targeting that prefix are blocked for a fixed period of 24 hours. This duration is chosen to fully suppress typical AIT campaigns, which often operate in short, intense bursts but may reappear intermittently.

Blocking is applied strictly at the prefix level to minimize collateral impact on unrelated users or services. The quarantine mechanism is stateless across prefixes, allowing independent evaluation and enforcement.

For each dataset and detection configuration, the following metrics are computed:

- time to first detection;
- number of messages delivered before detection (exposure);
- number of messages blocked after detection;
- proportion of traffic blocked;
- number of prefixes triggering detection;
- false-positive impact on legitimate marketing traffic.

Both fixed and rolling window approaches are evaluated under identical conditions to ensure comparability. The results are then analyzed in relation to delivery efficiency metrics such as IGDP and PDG to assess the broader operational impact of AIT mitigation.

Based on the components described above, the detection process operates as follows: incoming messages are aggregated by MSISDN prefix and evaluated in fixed or rolling five-minute windows. For each window, message volume and optional unique-recip-

ient metrics are compared against time-of-day–adaptive thresholds. A prefix is flagged when thresholds are exceeded, with detection time defined at the end of the triggering window. The affected prefix is then quarantined for 24 hours, during which all subsequent messages are blocked. Exposure before detection and blocked volume after mitigation are recorded for evaluation.

From a signal-processing perspective, the proposed method employs a sliding-window aggregation model with a uniform (rectangular) weighting function, where all messages within the observation window contribute equally to the detection metric. Fixed windows correspond to non-overlapping rectangular segments, while rolling windows implement a sliding rectangular kernel with a one-minute step.

The detection system is designed as a soft real-time system, where strict deterministic latency guarantees are not required, but timely detection is essential to minimize economic exposure. The main temporal characteristics of the system are defined as follows: a window size of 5 minutes, rolling step of 1 minute, detection latency ranging from 0 to 5 minutes depending on window type, and near-immediate enforcement after threshold violation. This configuration provides a balance between responsiveness and statistical stability of detection.

Results and discussion

This section presents the results of applying the proposed AIT detection logic to multiple real-world datasets and discusses the observed trade-offs between detection efficiency, blocking effectiveness, and false-positive risk. The evaluation focuses on three traffic profiles: two confirmed AIT incidents with different behavioral characteristics and one large legitimate marketing campaign.

Detection performance on high-intensity AIT traffic

The first dataset represents a classic high-intensity AIT attack, characterized by rapid bursts of OTP messages concentrated within a small number of MSISDN prefixes. Under these conditions, the proposed detection logic performs very effectively.

Table 1

Summarizes the detection and blocking performance of the proposed method across different traffic profiles and windowing strategies

AIT detection and blocking performance across datasets Dataset	Window type	Total messages	Detected prefixes	Blocked messages	Blocked (%)
AIT Sample 1	Fixed	140,636	609	134,006	95,29
AIT Sample 1	Rolling	140,636	609	124,433	88,48
AIT Sample 2	Fixed	38,802	1,430	24,502	63,15
AIT Sample 2	Rolling	38,802	1,430	2,166	5,58
Marketing	Fixed	100,946	–	58,203	57,66
Marketing	Rolling	100,946	–	47,802	47,35

Using message-count thresholds, fixed five-minute windows block between 88 % and 96 % of the total traffic volume, depending on the configuration. Rolling windows detect anomalous behavior earlier but block a slightly smaller fraction of messages. This difference is explained by the mechanics of windowing: rolling windows trigger earlier in the attack lifecycle, leaving fewer messages remaining to be blocked once quarantine is applied, whereas fixed windows detect later but capture a larger post-detection tail of the burst.

Despite this trade-off, both windowing approaches demonstrate strong mitigation capabilities for dense, short-lived AIT attacks. These results confirm that simple prefix-level aggregation combined with short time windows is sufficient to detect and suppress high-impact AIT incidents.

Detection performance on distributed AIT patterns

The second AIT dataset exhibits a more distributed attack strategy. Traffic is spread across a larger number of prefixes, with lower per-prefix message rates, likely reflecting an attempt to evade simple volume-based detection.

In this scenario, detection effectiveness is notably lower. Fixed windows block approximately 63 % of the traffic, while rolling windows block only around 5–6 %. The reduced performance of rolling windows highlights an important limitation: when AIT traffic is intentionally diluted across time and prefixes, early detection results in minimal remaining traffic to block.

These findings illustrate that no single windowing strategy is universally optimal. Fixed windows are better suited to identifying sustained or distributed

attacks, while rolling windows excel at detecting rapid bursts. This observation supports the use of hybrid or multi-resolution detection strategies in operational deployments.

False-positive behavior on legitimate marketing traffic

The legitimate marketing dataset provides critical insight into the risks of false positives. Marketing campaigns often generate high message volumes within short timeframes and may target subscribers within shared number ranges, especially in geographically or demographically focused campaigns.

When applying message-count thresholds alone, the detection logic incorrectly blocks 47–58 % of marketing traffic, depending on the windowing strategy. This confirms that volume-based signals, while effective against AIT, are insufficient on their own to safely distinguish fraud from legitimate commercial activity.

In contrast, when detection relies on unique MSISDN counts rather than raw message volume, the false-positive rate drops sharply. This behavior reflects the inherent diversity of marketing campaigns, which typically distribute messages across many distinct recipients. However, unique-recipient thresholds alone are not effective for detecting AIT, as attackers often repeatedly target a limited set of numbers.

These results reinforce the need for combining multiple behavioral signals rather than relying on a single metric.

Table 2

Impact of detection metric selection on blocking behavior

Dataset	Metric	Window type	Blocked (%)	Observed behavior
AIT Sample 1	Message count	Fixed	95,29	High sensitivity to bursts
AIT Sample 1	Unique MSISDNs	Fixed	0,00	Low diversity in AIT
Marketing	Message count	Fixed	57,66	High false positives
Marketing	Unique MSISDNs	Fixed	2,36	Legitimate diversity preserved

As shown in Table 2, message-count thresholds are highly effective for detecting AIT but lead to significant false positives on marketing traffic, while unique-recipient metrics exhibit the opposite behavior.

Fixed versus rolling window trade-offs

The comparison between fixed and rolling windows highlights a fundamental design trade-off. Fixed windows tend to block a larger fraction of AIT traffic because detection occurs later, when a significant portion of the burst remains. Rolling windows, by contrast, detect anomalies earlier but reduce the amount of traffic that can be blocked after detection.

From an operational perspective, earlier detection may still be preferable in environments where even short exposure carries high risk, such as OTP abuse. In other cases, maximizing blocked volume may be more important. These findings suggest that windowing strategy should be selected based on business priorities rather than detection accuracy alone.

Impact on delivery efficiency metrics

The suppression of AIT has direct implications for delivery efficiency. Fraudulent messages consume network resources and generate cost without providing value to end users. By blocking AIT traffic, the proposed detection logic improves the proportion of meaningful deliveries, thereby increasing the Intelligent Guaranteed Delivery Performance (IGDP).

At the same time, preventing wasteful message delivery reduces enterprise expenditure and narrows the Price–Delivery Gap (PDG). In high-intensity attack scenarios, the observed blocking rates correspond to substantial reductions in unnecessary spend, demonstrating that AIT mitigation is not only a security measure but also an economic optimization.

Discussion and limitations

The results confirm that transparent, low-latency, rule-based detection can effectively mitigate AIT under a wide range of conditions. However, they also expose inherent limitations. Message-count thresholds alone are prone to false positives during legitimate marketing campaigns, while unique-recipient metrics lack sufficient sensitivity to detect fraud in isolation.

These limitations suggest that the proposed method should be viewed as a strong baseline rather than a complete solution. In practice, combining prefix-level statistics with additional signals–such as entropy measures, sender behavior baselines, or lightweight content classification–will be necessary to achieve robust performance across diverse traffic profiles.

To illustrate the operation of the proposed detection logic, consider a simplified example of message traffic directed to a single MSISDN prefix.

Assume that within a five-minute observation window, the following ten messages are generated by an enterprise application and delivered to destination numbers sharing the same prefix (MSISDN without the last three digits). The messages arrive within a time span of 94 seconds.

Under daytime conditions, the detection threshold for this prefix is set to 10 messages per five-minute window. As the total message count reaches this threshold before the end of the window, the prefix is flagged as suspicious. The detection timestamp is defined as the end of the triggering window.

From this moment, a quarantine period of 24 hours is applied to the prefix. All subsequent messages addressed to the same prefix during this interval are blocked. Messages delivered prior to the detection timestamp contribute to exposure, while blocked messages contribute to mitigation efficiency.

If the same traffic pattern were observed during nighttime hours, a lower threshold of five messages would apply, resulting in earlier detection and reduced exposure. Conversely, if the traffic belonged to a legitimate marketing campaign spanning multiple prefixes with higher MSISDN diversity, the threshold would not be exceeded and no blocking would occur.

This example demonstrates how message density, time-of-day thresholds, and prefix-based aggregation jointly determine detection outcomes in a transparent and interpretable manner.

Table 3

Example of AIT detection logic application

Time (UTC)	Destination MSISDN	Prefix	Window count
10:01:05	38631234567	38631234	1
10:01:18	38631234569	38631234	2
…	…	…	…
10:02:39	38631234580	38631234	10 → Trigger

Conclusions

Artificially Inflated Traffic represents a fundamental shift in how fraud manifests within the A2P messaging ecosystem. Rather than exploiting weaknesses in signaling protocols, AIT abuses legitimate enterprise messaging workflows, making it both economically damaging and difficult to detect using traditional network-based defenses. This study set out to evaluate whether simple, transparent, and deployable statistical techniques could meaningfully reduce the impact of such attacks without introducing excessive operational complexity or privacy concerns.

The results demonstrate that prefix-level aggregation combined with short time-window analysis is highly effective for detecting classic high-intensity AIT bursts. In these scenarios, the proposed approach blocks up to 96 % of fraudulent traffic, substantially reducing financial exposure. Fixed windowing achieves the highest overall mitigation by capturing the tail of sustained bursts, while rolling windows enable earlier detection of rapidly emerging attacks. These complementary behaviors suggest that windowing strategy should be selected based on operational priorities, such as minimizing exposure time versus maximizing blocked volume.

At the same time, the evaluation highlights an important limitation: message-count–based detection alone is insufficient for safely distinguishing AIT from legitimate marketing campaigns. High false-positive rates observed on marketing traffic underscore the need for additional behavioral signals. Unique-recipient metrics significantly reduce false positives but lack the sensitivity required to detect AIT in isolation. This reinforces the conclusion that effective AIT mitigation must rely on a combination of simple, interpretable indicators rather than a single threshold.

Beyond fraud prevention, the findings also have broader implications for delivery efficiency. By suppressing traffic that generates cost without delivering user value, AIT mitigation directly improves delivery performance metrics such as IGDP and reduces the Price–Delivery Gap (PDG). From this perspective, fraud detection becomes an integral component of routing optimization and economic efficiency in CPaaS platforms.

Overall, this study confirms that low-latency, content-agnostic detection rules can form a strong and practical foundation for AIT defense. While such rules are not sufficient on their own to address all attack variants, they provide a necessary baseline that can be implemented immediately and extended incrementally as more advanced techniques are introduced.

Future work

While the proposed approach demonstrates strong effectiveness against several classes of AIT attacks, the results also point to clear directions for future research and system evolution.

A key area for further development is the integration of content-aware and campaign-level classification. Although this study intentionally avoided payload inspection, lightweight mechanisms such as template identification, campaign fingerprinting, or message-type tagging could significantly reduce false positives by distinguishing legitimate marketing activity from adversarial traffic pumping. Such techniques could be applied selectively and in a privacy-preserving manner to complement metadata-based detection.

Another promising direction is sender and enterprise behavioral modeling. Rather than relying solely on global thresholds, future systems could establish dynamic baselines for individual senders or API credentials, capturing typical volume, prefix dispersion, and temporal patterns. Deviations from these baselines

would allow more precise anomaly detection, particularly in environments with diverse traffic profiles.

The evaluation also suggests benefits from multi-resolution and adaptive windowing strategies. Combining rolling and fixed windows of different durations could improve sensitivity to both short-lived bursts and low-rate distributed attacks. Thresholds could further adapt based on historical traffic patterns, geography, or service type, improving robustness without sacrificing transparency.

Statistical enrichment at the prefix level represents another avenue for improvement. More advanced entropy measures, run-length analysis, or probabilistic scoring models could be used to assign risk scores rather than binary decisions, enabling graduated responses instead of immediate blocking. Such approaches would allow operators to tune enforcement actions according to business risk tolerance.

Finally, future work should explore deeper integration between fraud detection and routing optimization frameworks. Embedding AIT risk signals directly into routing and fallback decisions – alongside metrics such as IGDP and PDG – would enable messaging platforms to jointly optimize security, reliability, and cost efficiency. This convergence represents a natural evolution toward adaptive, economically aware messaging infrastructures.

REFERENCES

[1] ITU. ICT Facts and Figures 2024. Geneva : International Telecommunication Union, 2024. URL: https://www.itu.int/itu-d/reports/statistics/facts-figures-2024/ (дата звернення: 13.12.2025).

[2] GSMA Intelligence. The Mobile Economy 2024. London : GSM Association, 2024. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/the-mobile-economy-2024/ (дата звернення: 13.12.2025).

[3] 3GPP TR 23.840. SMS Evolution and Migration Specifications. 3rd Generation Partnership Project, 2023. URL: https://www.3gpp.org/dynareport/23840.htm (дата звернення: 13.12.2025).

[4] Juniper Research. A2P Messaging Market Forecast 2023–2027. 2023. URL: https://www.juniperresearch.com/ (дата звернення: 13.12.2025).

[5] GSMA Fraud and Security Group. SS7 Interconnect Security Monitoring and Firewall Guidelines (FS.11). Version 3.0. 2021. URL: https://www.gsma.com/ (дата звернення: 13.12.2025).

[6] AdaptiveMobile Security. SS7 and Diameter Vulnerability Landscape: Threat Intelligence Report 2022. Dublin, 2022. URL: https://www.enea.com/solutions/signaling-security/ (дата звернення: 13.12.2025).

[7] Communications Fraud Control Association (CFCA). Global Fraud Loss Survey 2023. CFCA, 2023. URL: https://cfca.org/telecommunications-fraud-increased-12-in-2023-equating-to-an-estimated-38-95-billion-lost-to-fraud/ (дата звернення: 18.12.2025).

[8] Proofpoint. SMS Pumping Fraud: Threat Landscape Report 2024. Sunnyvale, 2024. URL: https://www.proofpoint.com/us/resources/threat-reports/state-of-phish (дата звернення: 18.12.2025).

[9] Mobile Ecosystem Forum (MEF). Business Messaging Fraud Report 2023. MEF, 2023. URL: https://mobileecosystemforum.com/mef-anti-fraud-yearbook-2024/ (дата звернення: 18.12.2025).

[10] Grand View Research. A2P Messaging Market Size, Share & Trends Analysis Report, 2025–2030. San Francisco, 2024. Report ID: GVR-1-68038-842-8. URL: https://www.grandviewresearch.com/ (дата звернення: 19.12.2025).

[11] Grand View Research. North America A2P Messaging Market Outlook 2030. San Francisco, 2024. URL: https://www.grandviewresearch.com/ (дата звернення: 19.12.2025).

[12] MarketsandMarkets. A2P Messaging Market – Global Forecast to 2029. Pune, 2024. Report Code: TC 2585. URL: https://www.marketsandmarkets.com/ (дата звернення: 20.12.2025).

[13] GSMA Fraud and Security Group. SMS Firewalls and Anti-Fraud Best Practices (FS.19). Version 2.0. London: GSMA, 2023. URL: https://www.gsma.com/ (дата звернення: 20.12.2025).

[14] Communications Fraud Control Association (CFCA). Artificially Inflated Traffic and SMS Pumping Fraud Overview. CFCA, 2023. URL: https://cfca.org/ (дата звернення: 20.12.2025).

[15] Mobile Ecosystem Forum (MEF). AIT and SMS Pumping: Industry Overview. MEF, 2023. URL: https://mobileecosystemforum.com/enterprise-mobile-messaging-fraud-framework/ (дата звернення: 20.12.2025).

[16] Juniper Research. Mobile Messaging Market Trends and Enterprise Use Cases 2024. Basingstoke, 2024. URL: https://www.juniperresearch.com/ (дата звернення: 20.12.2025).

[17] AdaptiveMobile Security. Messaging Abuse and A2P Fraud Trends. Threat Report Series 2022–2024. Dublin, 2024. URL: https://www.enea.com/ (дата звернення: 20.12.2025).

[18] Одарченко М. В. Сучасні виклики безпеці послуг SMS та сигнальних мереж. Київ: Національний авіаційний університет, 2024. 120 с.

Одарченко М. С., Заліський М. Ю.

ВИЯВЛЕННЯ ШТУЧНО ЗБІЛЬШЕНОГО ТРАФІКУ (AIT) В A2P-МЕСЕДЖИНГУ: МОДЕЛІ ЧАСОВИХ ВІКОН ТА ПРЕФІКСНА АНАЛІТИКА ДЛЯ ПРОТИДІЇ ШАХРАЙСТВУ

У статті досліджується зростаюча проблема штучно збільшеного трафіку (Artificially Inflated Traffic, AIT) у сервісах Application-to-Person (A2P) SMS та пропонується інтерпретована, метаданими керована методика для його виявлення й блокування в режимі реального часу. Показано, що AIT використовує легальні канали доставки, формуючи швидкі або розподілені «сплески» вихідних повідомлень до послідовних або близьких діапазонів MSISDN, що призводить до значних фінансових втрат для підприємств і CPaaS-платформ та знижує ефективність доставки.

Для вирішення цієї проблеми запропоновано префіксно-орієнтований підхід, що базується на аналізі фіксованих і ковзних часових вікон, порогах за кількістю повідомлень та унікальних MSISDN, а також на оцінці ентропії розподілу номерів. Методологія включає аналіз великомасштабних емпіричних датасетів, порівняння затримки виявлення й ефективності блокування, а також інтеграцію результатів із раніше розробленими метриками ефективності доставки — індексом гарантованої доставки (IGDP) та показником ціна–доставка (PDG).

Експериментальні результати демонструють, що аналіз фіксованих вікон забезпечує до 96% блокування високінтенсивних AIT-атак, тоді як ковзні вікна дають змогу раніше виявляти короткотривалі аномалії. Дослідження також показує ризик хибних спрацювань для легітимних маркетингових кампаній, що підкреслює необхідність використання додаткових ознак — профілю поведінки відправника, класифікації контенту та багаторівневих моделей аномалій. У висновках визначено, що прозорі статистичні моделі формують надійну основу для протидії AIT і наведено практичні рекомендації щодо підвищення стійкості CPaaS-інфраструктур до шахрайства.

Ключові слова: AIT, виявлення шахрайства, SMS, CPaaS, префіксна аналітика, ковзні вікна, ентропія, IGDP, PDG.

Odarchenko M., Zaliskyi M.

ARTIFICIALLY INFLATED TRAFFIC IN A2P MESSAGING: WINDOW-BASED DETECTION MODELS AND PREFIX-LEVEL ANALYSIS FOR FRAUD MITIGATION

The article examines the growing problem of Artificially Inflated Traffic (AIT) in Application-to-Person (A2P) SMS delivery and introduces an interpretable, metadata-driven framework for detecting and mitigating such fraud in real time. The study highlights how AIT exploits legitimate messaging pathways, generating rapid or distributed bursts of outbound traffic toward sequential or closely grouped MSISDN ranges. These patterns impose substantial financial losses on enterprises and CPaaS providers while degrading delivery efficiency metrics.

To address these vulnerabilities, the paper proposes a prefix-level analytical approach based on fixed and rolling time-window aggregation, message-count and unique-MSISDN thresholds, and entropy-based evaluation of MSISDN distributions. The methodological toolkit includes large-scale empirical dataset analysis, comparative evaluation of detection latency and blocking efficiency, and integration of outcomes with previously introduced delivery performance metrics–the Intelligent Guaranteed Delivery Index (IGDP) and Price–Delivery Gap (PDG).

Experimental results show that fixed-window analysis achieves up to 96% mitigation for high-intensity AIT bursts, while rolling windows enable earlier detection of short-lived anomalies. The study also reveals that threshold-based detection models may misclassify legitimate marketing campaigns, emphasizing the need for auxiliary features such as sender behavior profiling, content-aware classification, and multi-resolution anomaly scoring. The paper concludes that transparent statistical models provide a robust baseline for AIT defense and outlines practical recommendations for enhancing fraud resilience in large-scale CPaaS infrastructures.

Keywords: AIT, fraud detection, SMS, CPaaS, prefix clustering, rolling windows, entropy, IGDP, PDG.

Received: 20.12.2025 p.
Accepted: 10.03.2026 p.
Published: 27.04.2026 p.