УДК 621.396.1 DOI: 10.31673/2786-8362.2025.014820

Ветошко І.П.

FLASH CALLS IN MODERN TELECOMMUNICATION NETWORKS: THREATS, CHALLENGES, AND EFFECTIVE COUNTERMEASURES

Vetoshko I.P. Flash calls in modern telecommunication networks: threats, challenges, and effective countermeasures. This paper addresses the phenomenon of Flash Calls — ultra-short incoming calls used for user authentication in modern telecommunication networks — and their impact on network security and operator revenues. Unlike traditional SMS-based methods, Flash Calls allow cost-effective, high-speed verification, but simultaneously bypass billing systems and facilitate fraudulent practices such as CLI spoofing and international revenue sharing fraud (IRSF). The paper explores the technical characteristics of Flash Calls, highlights their detection complexity, and analyses vulnerabilities in current anti-fraud systems. A multi-layered countermeasure architecture is proposed, integrating real-time analytics, machine learning algorithms, and STIR/SHAKEN protocols to identify and block fraudulent activity. The effectiveness of the approach is supported by traffic analysis and experimental blocking results on global platforms such as Meta. The proposed system ensures reliable identification of malicious patterns while minimizing false positives and preserving service quality.

Keywords: Flash Calls, CLI spoofing, fraud detection, Voice firewall, Wangiri, machine learning, STIR/SHAKEN, telecom security, IMS network

Ветошко І.П. Флеш-дзвінки в сучасних телекомунікаційних мережах: загрози, виклики та ефективні заходи протидії. У статті розглядається феномен Flash-дзвінків – надкоротких вхідних дзвінків, що використовуються для аутентифікації користувачів у сучасних телекомунікаційних мережах – та їх вплив на безпеку мереж і доходи операторів. На відміну від традиційних SMS-методів, Flash-дзвінки дозволяють здійснювати економічно ефективну, високошвидкісну верифікацію, але водночас обходять білінгові системи і сприяють шахрайським діям, таким як підміна CLI і міжнародне шахрайство з розподілом доходів (IRSF). У статті досліджуються технічні характеристики флеш-дзвінків, підкреслюється складність їх виявлення та аналізуються вразливості в існуючих системах протидії шахрайству. Запропоновано багаторівневу архітектуру протидії, що поєднує аналітику в режимі реального часу, алгоритми машинного навчання та протоколи STIR/SHAKEN для виявлення та блокування шахрайських дій. Ефективність підходу підтверджується аналізом трафіку та результатами експериментальних блокувань на глобальних платформах, таких як Meta. Запропонована система забезпечує надійну ідентифікацію шкідливих патернів, мінімізуючи помилкові спрацьовування та зберігаючи якість обслуговування.

Ключові слова: Flash-дзвінки, підміна CLI, виявлення шахрайства, голосовий файрвол, Вангірі, машинне навчання, STIR/SHAKEN, телекомунікаційна безпека, мережа IMS

Introduction

Statement of the problem. Modern telecommunication networks are constantly facing new challenges related to the spread of non-standard and innovative approaches to information transmission. One such phenomenon is Flash Calls, short incoming calls used to authenticate users instead of traditional SMS codes. Despite the obvious advantages for end users and services (low cost, high speed), the massive introduction of Flash Calls creates significant risks for mobile operators, in particular in terms of financial losses, bypassing billing mechanisms, complicating the detection of fraudulent activity and CLI spoofing [1,2]. The subject area of this study is the information security of telecommunication networks, with a special emphasis on detecting and counteracting the fraudulent use of Flash Calls. The study analyzes the current threats associated with the use of Flash Calls on a global scale, examines the features of their technical profile, and proposes multi-level approaches to detecting and blocking them [3,4].

The problem is caused by the rapid growth of this type of traffic – according to the latest data, up to 10% of all incoming calls can be classified as Flash Calls, which creates an additional burden on the infrastructure and threatens to cause losses of operators' revenues. Also, in the context of the active development of digital platforms that use alternative authentication channels, the lack of effective mechanisms for detecting and controlling Flash Calls may contribute to the spread of cyber threats [1,4]. In view of this, the study is practically oriented, as it focuses on real mechanisms to counteract abuse in telecommunications networks, including the integration of artificial intelligence, real-time traffic profiling, and the use of STIR/SHAKEN protocols to confirm the authenticity of Caller ID [8,9,11]. The results obtained can be directly implemented in the practice of mobile operators and contribute to strengthening the security of the information environment [4,10,15].

Analysis of recent studies. The studies [1], [2] consider the phenomenon of Flash Calls as a new tool for user authentication, which is actively used by both legitimate services and attackers. In [3], [4], [6], the threats associated with bypassing traditional billing mechanisms of mobile operators, in particular through the use of CLI spoofing, SIM farms, and VoIP gateways, are analyzed. Special attention is paid to such fraud schemes as Wangiri and IRSF. Publications [8], [9], [10] describe the architectures of systems for actively detecting and blocking unwanted calls based on voice firewalls, as well as the implementation of STIR/SHAKEN protocols for verifying the authenticity of Caller ID. In [4], [5], signaling architectures with SBC, STP, and DRA elements are presented, which are critical for monitoring and routing traffic.

At the same time, a number of aspects have been identified that remain insufficiently disclosed in current research:

- [1], [3] insufficiently consider methods for identifying Flash Calls at the level of real-time signal traffic;
- [2], [4] lack practical recommendations for integrating artificial intelligence mechanisms into the analysis of such calls;
- [5], [15] only partially describe the consequences of blocking Flash Calls for OTT platforms (Meta, WhatsApp, Instagram), without a deep analysis of changes in the traffic structure;
- [6], [7] do not emphasize the scenarios of abuse through locally spoofed numbers and their impact on national operators.

The outcome of the review is the need to develop a comprehensive approach to detecting and countering Flash Calls, including machine learning, CLI verification, adaptive filters, and integration into the core of the operator's network. The research topic of this paper focuses on improving technical approaches to analyzing ultra-short calls and creating an adaptive protection system with minimal impact on legitimate services [4], [8], [10].

The purpose of this paper is to develop and substantiate an effective technical approach to detecting and blocking Flash Calls in telecommunication networks, which would protect operators from revenue losses, abuse by intruders and ensure compatibility with the existing network infrastructure. The proposed approach should take into account the characteristic features of Flash Calls, their short duration, high traffic volume, and the use of fake CLIs.

To achieve this goal, the study solves the following tasks: to investigate the architecture of Flash Calls in modern networks and determine their technical characteristics; to analyze the weaknesses of typical fraudulent call detection systems used by mobile operators; to assess the impact of Flash Calls on the billing system, signal traffic and caller ID security; to develop an architecture for a multi-level Flash Calls detection system using machine learning mechanisms, real-time traffic analysis

The result of the research should be an adaptive system for monitoring and counteracting Flash Calls, capable of responding to changes in traffic patterns, reducing the number of false positives and minimizing the impact on legitimate services.

Presentation of the main research material

Understanding flash calls. Flash calling is a relatively new phenomenon in the telecommunications sector, used primarily as an alternative to SMS-based one-time passwords (OTPs) for user authentication. The method involves a short, often zero-duration incoming call to a user's mobile device. The last few digits of the call number act as an authentication code that is automatically recognised by certain applications to complete the authentication process. This approach is becoming increasingly popular due to its cost advantages over traditional SMS-based authentication methods [1].

However, the use of Flash calls is associated with certain challenges, especially for mobile network operators (MNOs). Unlike SMS messages, which generate direct revenue for operators, Flash calls bypass traditional revenue streams, leading to potential financial losses. In addition, fraudsters and cybercriminals have begun to use Flash calls for malicious purposes, such as identity theft, account takeovers and unauthorised network access.

From a technical standpoint, Flash calls are often difficult to detect using conventional fraud prevention methods because they do not always follow known fraudulent traffic patterns. Conventional telecom fraud detection systems are primarily designed to identify irregularities in SMS and voice communications; however, they often struggle to effectively differentiate between authentic and malicious flash call activities. This problem calls for advanced detection mechanisms that analyse call patterns, call durations and the networks from which they originate to detect and reduce the number of fraudulent Flash calls. In addition, the legal and regulatory framework governing telecommunications services is not yet fully adapted to the growing number of ‘emergency calls’. Many existing regulations focus on the prevention of SMS and voice call fraud, leaving gaps in addressing the new risks associated with flash calls. As a result, telecommunications operators should take proactive steps to develop their own solutions to detect and mitigate the effects of flash calling, while remaining compliant with new and evolving industry standards [2].

Key features of flash calls. Flash calls have several unique characteristics that distinguish them from other forms of voice and SMS communication. These characteristics play a crucial role both in their legitimate use for authentication purposes and in their potential for fraudulent use:

1. **Extremely short call duration.** Flash calls are designed to end in a split second, often before the recipient has time to answer the call. In many cases, these calls do not even ring on the user's device. This behaviour makes them difficult to detect using traditional call monitoring methods.

2. **High volume and automated traffic.** Because Flash calls are primarily used for authentication, they are typically generated in high volumes, especially by businesses that need to verify users quickly and seamlessly. This leads to significant spikes in call traffic that may not always be distinguishable from fraudulent activity.

3. **Specific call patterns and number ranges.** Flash calls often come from predefined number ranges associated with authentication services. These ranges can be domestic or international, depending on the service provider's infrastructure. However, fraudsters can exploit this feature by using spoofed numbers or manipulating caller ID information.

4. **Use of SIM farms and VoIP infrastructure.** Fraudsters have been known to use SIM farms (large-scale operations with multiple SIM cards used to automatically generate traffic) and VoIP calls to initiate ‘flash calls’. These methods allow attackers to bypass authentication controls and generate large volumes of calls at minimal cost.

5. **Risks of circumventing revenue and monetisation controls**. One of the main concerns for telecommunications operators is that flash calls eliminate the need for SMS OTP, which leads to revenue losses. In addition, attackers can use flash calls as a means of circumventing traditional telecoms billing mechanisms, further increasing the financial losses for operators.

6. **Caller ID spoofing and geo-masking.** Fraudsters often spoof caller ID information to hide the true source of flash calls, misleading recipients as to the origin of the call. This allows them to mask the true source of the call and bypass detection mechanisms. Geographic masking further complicates the problem by making it appear that calls are coming from verified extensions.

7. **Challenges in tracking and preventing fraud.** Due to their short duration and the nature of their use, flash calls are often difficult to track and analyse using conventional telecommunications monitoring systems. To effectively detect fraudulent calls, operators must implement advanced analytics based on artificial intelligence, machine learning models and real-time call profiling [3].

Approaches to detecting and counteracting flash calls in telecommunications networks.

To effectively counteract flash calls, modern telecommunications networks must deploy a multi-layered security strategy that includes real-time analytics, adaptive security policies, and advanced fraud detection mechanisms. Table 1 summarizes the fundamental elements that contribute to a robust anti-flash call defence system.

A comprehensive fraud management system should provide complete network transparency to detect and analyse abnormal call behaviour. Implementing real-time monitoring and profiling techniques helps operators distinguish legitimate authentication calls from fraudulent attempts [2]. Real-time call profiling: Machine learning algorithms and analytics-based approaches are used to evaluate call patterns in real time, identifying the behaviour of flash calls based on the frequency, duration, and numbers they come from. Improved customer experience: Advanced filtering mechanisms reduce false positives by ensuring that legitimate calls are not accidentally blocked and preventing unauthorized authentication attempts.

Table 1

Key elements for creating a reliable anti-flash call defence system

Transparency and control in fraud management	Comprehensive Voice Fraud Handling	Safe deployment in MNO's core network	Flexible and Adaptive Call Filtering Techniques	Diversity & Ease of Use (Support of multiple actions)
Real-time call profiling	Tackling diverse voice scams	High availability	Supports multiple integration protocols SIP, CAMEL etc.	Call block
Improved customer experience	Wangiri, Simbox	Geo redundancy	Customised rules configuration	Temporary call hold
	Robo/Scam Calls, Flash Calls	Fallback routing		CLI suppression / modification
	CLI spoofing			Call redirection
	IRSF			

A comprehensive fraud management solution includes a 360-degree approach and is crucial. Modern networks face a variety of voice fraud schemes beyond flash calls. A robust fraud detection framework should include: Wangiri and Simbox fraud detection option to identify one-ring call fraud and illegitimate call rerouting, Robo and scam calls filtering to differentiate between genuine business calls and automated fraudulent campaigns, caller ID spoofing prevention to ensure the integrity of the caller ID by detecting anomalies in call routing, International Revenue Share Fraud (IRSF) detection option to identify of fraudulent traffic redirection schemes used to manipulate international call termination rates and generate illegal profits. To effectively combat flash calls, security mechanisms must be built directly into the operator's core network, ensuring that requirements are met in real time. High availability and geo-redundancy involve deploying distributed fraud detection solutions to ensure resilience against large-scale attacks. Fallback routing mechanisms enable the redirection of suspected fraudulent calls to analysis engines without disrupting legitimate traffic [2,3].

Fraudulent call patterns are constantly changing, requiring operators to implement flexible fraud prevention systems: Customizable rule configuration enables operators to define criteria for detecting flash calls and dynamically adjust thresholds based on real-time traffic patterns; Multi-protocol support ensures compatibility with SIP, CAMEL, and other telecommunications standards, allowing fraud detection across different signaling environments. Operators should implement a variety of call filtering and blocking mechanisms that can be easily adapted to different fraud scenarios. Multiple action support enables dynamic measures such as blocking, temporary call holds, and call redirection to mitigate fraud. CLI suppression and modification help adjust caller ID visibility to counter spoofing attempts and fraudulent call rerouting [1,3].

CLI (Calling Line Identification) spoofing has become a serious problem in modern telecommunications, which leads to considerable financial impact for mobile network operators (MNOs) and other stakeholders. According to the CFCA's Fraud Loss Survey Report 2021, CLI spoofing alone has caused losses of around USD 2.63 billion for telecoms operators worldwide [2].

CLI spoofing occurs when fraudsters intentionally manipulate the caller ID to disguise the origin of a call. This fraudulent practice allows attackers to impersonate legitimate institutions, government agencies, or trusted companies, misleading unsuspecting users into disclosing sensitive information. In addition, interconnect operators can also alter CLI data, resulting in a loss of interconnect revenue when the operator does not receive the proper termination fees due to distorted call routing. The two main causes of CLI spoofing include the following:

- Fraudulent imitation: Attackers spoof trusted numbers to trick people into enabling financial fraud, phishing attacks, or unauthorised access to user accounts. This tactic is commonly used in social engineering attacks, where victims are manipulated into believing they are communicating with legitimate entities.
- Manipulation of interconnection revenues: Some telecommunications intermediaries modify CLI data for financial gain, either to misclassify call routing or to circumvent higher interconnect fees, resulting in direct revenue losses for operators [2].

To combat CLI spoofing, modern telecoms networks require real-time detection mechanisms based on machine learning and call analysis using artificial intelligence. Operators should implement CLI authentication methods, such as STIR/SHAKEN protocols, that verify the integrity of caller information, ensuring that incoming calls come from legitimate sources.

Wangiri calls, another common fraudulent scheme, are estimated to have caused telecoms operators $2.23 billion in losses. The term ‘Wangiri’ comes from Japanese and translates to ‘one call and that's it’, describing a scam in which fraudsters make thousands of automated calls that are disconnected after a single ring [4]. The objective is to prompt the recipient to return the missed call, inadvertently connecting them to premium international numbers, which incurs substantial charges. Wangiri fraud is often linked to international revenue sharing fraud (IRSF), where fraudsters partner with premium number providers to generate revenue from prepaid calls. This type of fraud is particularly challenging for telecommunications operators as it exploits the curiosity of users and cannot be easily blocked without advanced monitoring of call patterns. Key characteristics of the Wangiri scam include: short duration calls (Calls end before the caller has time to answer, increasing the likelihood of a callback); large-scale automated campaigns (scammers generate huge numbers of Wangiri-style calls using bot-driven dialling systems); the use of international numbers (calls usually come from expensive international destinations, which can lead to significant financial losses for unsuspecting individuals). To prevent Wangiri, flash calls, CLI spoofing, and other types of fraud, telecom providers should implement adaptive filtering mechanisms, using real-time call analytics to identify and block suspicious call patterns before they reach end users [4]. Implementing blacklists of known fraudulent number ranges and using artificial intelligence to detect anomalies can help operators prevent financial losses.

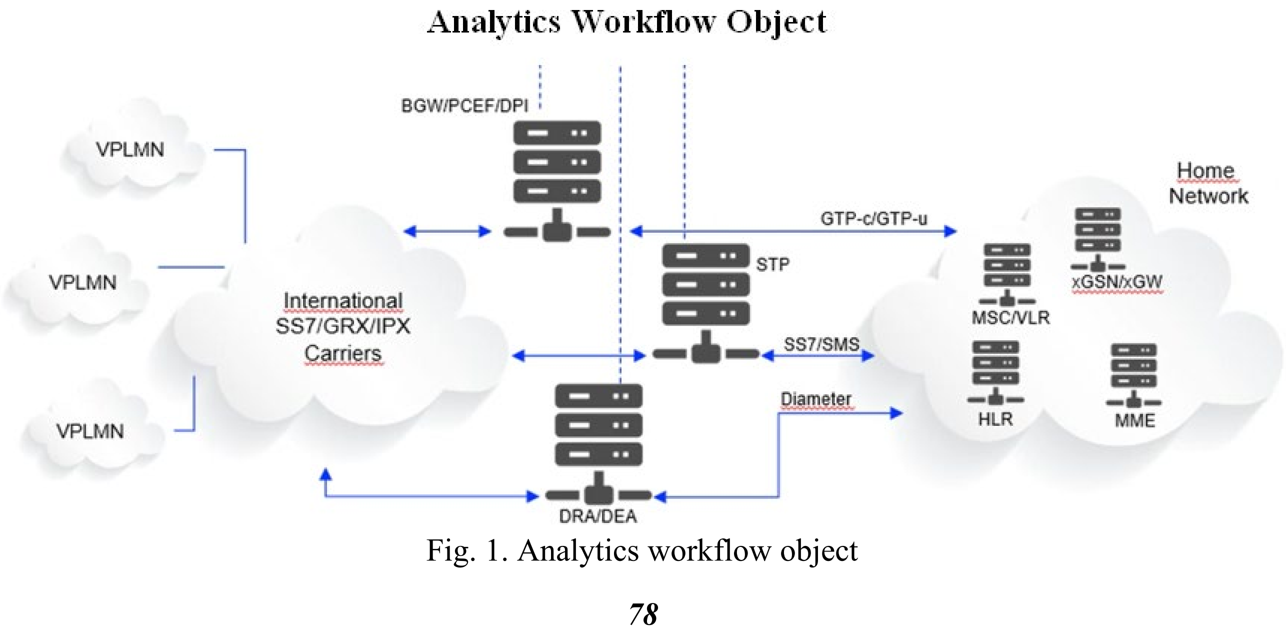

Fig. 1. Analytics workflow object

The analytics workflow object (Fig. 1) illustrates the interaction between the main elements of the telecommunications network and international signalling operators in the fraud detection process. It shows the flow of traffic through the VPLMN, international SS7/GRX/IPX operators, STP, DRA/DEA and home network components, showing the critical points for fraud prevention. STP routes SS7/SMS and GTP traffic, making it a focal point for detecting flash calls, CLI spoofing, and Wangiri fraud. DRA/DEA-driven diameter signalling is analysed for authentication and billing anomalies to prevent unauthorised access. Fraudsters manipulate CLI data, introduce unauthorised signalling and redirect calls through operated interconnect operators, resulting in revenue losses. Security mechanisms include GTP-c/GTP-u analytics, SS7/SMS firewalls, and diameter-based anomaly detection, allowing operators to profile call behaviour and block fraudulent patterns in real time. Artificial intelligence-based traffic analysis enhances fraud detection at different levels of the telecoms infrastructure [1,4].

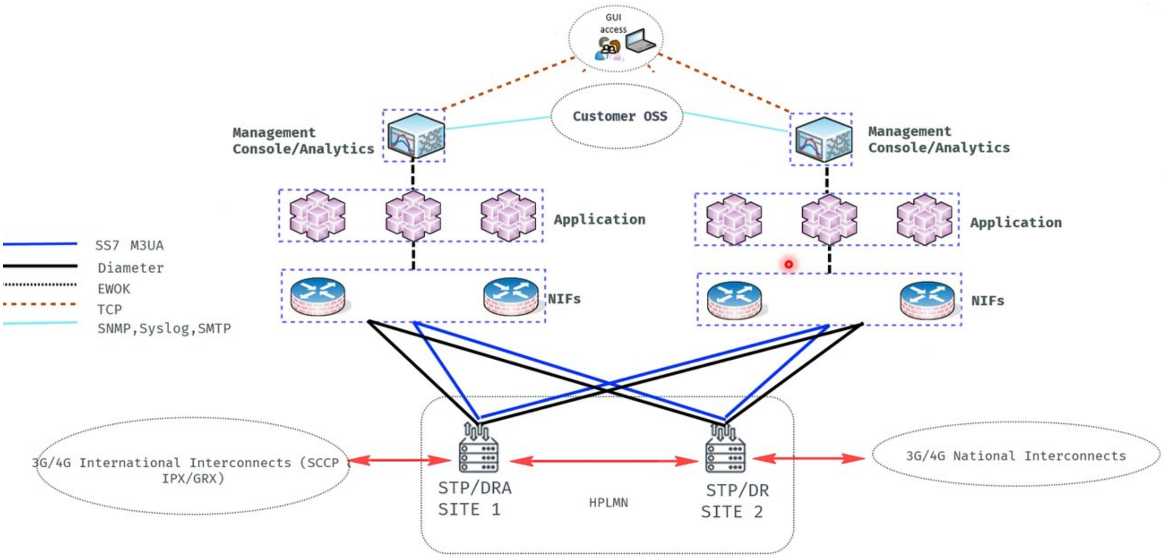
GUI access
Customer OSS
Management Console/Analytics
Management Console/Analytics
Application
Application
NIFs
NIFs
SS7 M3UA
Diameter
EWOK
TCP
SNMP,Syslog,SMTP
3G/4G International Interconnects (SCCP IPX/GRX)
STP/DRA SITE 1
HPLMN
STP/DR SITE 2
3G/4G National Interconnects

Fig. 2. Signaling traffic routing and fraud detection architecture

The diagram (Fig. 2) illustrates a telecommunications signaling architecture designed for efficient traffic routing, real-time analytics and fraud detection, including GUI-based management for operational oversight. It depicts the interactions between STP/DRA sites, international and domestic connections, network interface functions (NIFs), application processing layers, and management consoles. This infrastructure supports important signaling protocols such as SS7 (M3UA), Diameter, and IPX/GRX, which are fundamental to the operation of mobile networks, including fraud prevention. STP/DRA is based on the exchange of signaling messages, which provides optimized routing between multiple connection points [5]. These components support redundancy, load balancing, and security mechanisms to prevent traffic congestion and unauthorized manipulation. International connections facilitate roaming and cross-border signaling, while national connections enable traffic exchange at the domestic level. Given the vulnerability of SS7 and Diameter to fraud, such as CLI and IRSF spoofing, real-time analysis of signal flows is essential [2,4].

The application layer processes and filters incoming signal traffic, integrating with fraud detection algorithms and policy enforcement mechanisms. This layer connects to management consoles and analytical platforms that centralize monitoring, apply anomaly detection methods, and generate detailed network security reports. The Customer OSS interacts with these analytical tools through a graphical user interface (GUI), allowing operators to manage security policies, analyze fraud trends, and implement countermeasures in real time. The architecture utilizes multiple transport layers, including TCP-based EWOK messaging for transport signaling and SNMP/Syslog/SMTP for logging and alerting. These mechanisms enhance network security and real-time monitoring by enabling automated fraud detection, CLI inspection, and Flash attack prevention. By integrating GUI-driven analytics, operators gain a visual and interactive framework for monitoring traffic behavior, enforcing security policies, and responding quickly to new threats [6].

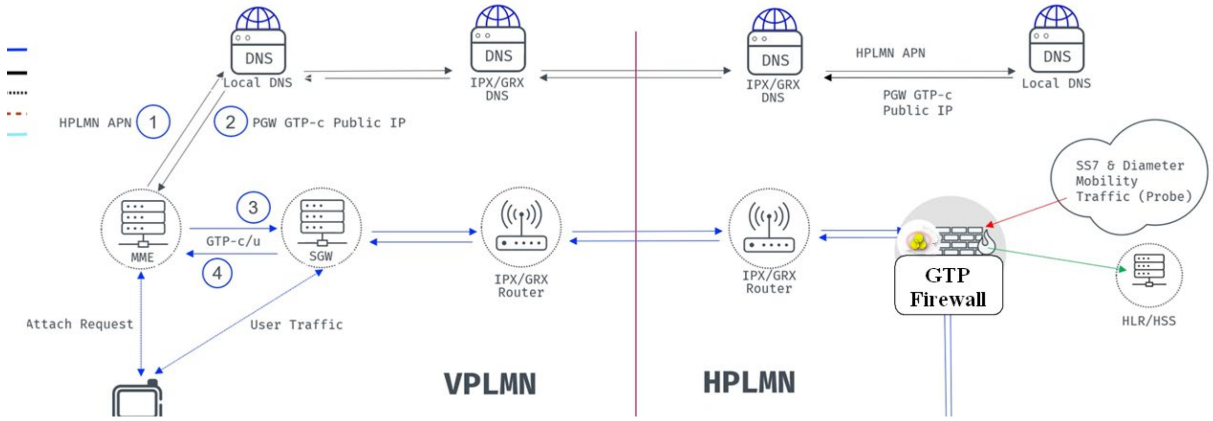

Fig. 3. GTP Traffic routing and firewall protection in VPLMN-HPLMN communication

The diagram (Fig. 3) illustrates the flow of GTP-C/U signaling and security between a visited public land mobile network (VPLMN) and a home public land mobile network (HPLMN), emphasizing the role of the GTP firewall (GTP-FW) in securing the signaling and mobility management. The process begins with the Mobility Management Entity (MME) processing connection requests from user devices, which runs DNS resolutions to obtain the HPLMN APN and PGW GTP-c Public IP address. The MME then establishes a GTP control plane (GTP-C) session with the service gateway (SGW), followed by a GTP user plane (GTP-U) traffic stream to transfer data [7]. The session is routed over IPX/GRX connections that facilitate inter-operator communication. The GTP firewall acts as a critical security layer by inspecting GTP-C, SS7 signaling and Diameter mobile traffic. It monitors session establishment to detect IMSI/TMSI leakage, session hijacking, unauthorized tunneling, and abnormal signaling behavior. The firewall does not modify GTP-U data traffic, ensuring minimal impact on the user experience, while applying strict policy enforcement to GTP-C messages. When fraud, spoofed requests, or policy violations are detected, the firewall either drops or rejects compromised signaling packets. By integrating GTP's real-time security mechanisms, the firewall strengthens fraud prevention strategies by mitigating threats such as flash calling, CLI spoofing, and international revenue sharing fraud (IRSF). This architecture strengthens subscriber privacy, signal integrity, and inter-operator security, ensuring that mobile networks remain resilient to new fraud techniques [4, 6].

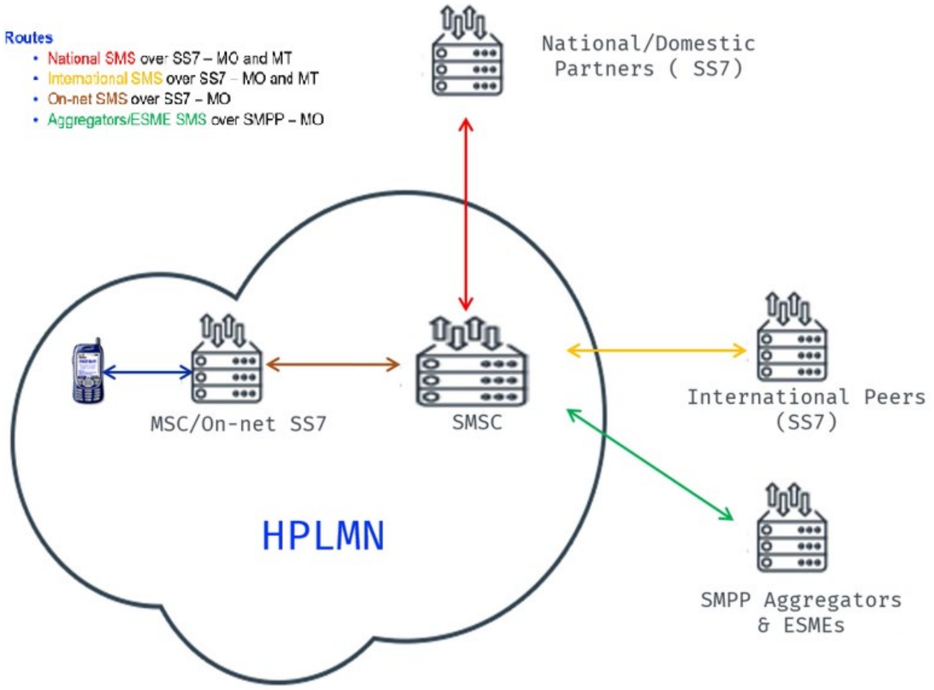
Routes
• National SMS over SS7 – MO and MT
• International SMS over SS7 – MO and MT
• On-net SMS over SS7 – MO
• Aggregators/ESME SMS over SMPP – MO
National/Domestic Partners (SS7)
MSC/On-net SS7
SMSC
International Peers (SS7)
HPLMN
SMPP Aggregators & ESMEs

Fig. 4. SMS interconnect architecture

The architecture of SMS interconnection (Fig. 4) is based on the use of SS7 signalling for messaging between mobile networks, as well as the SMPP protocol for integrating external service platforms [14]. Key nodes, such as the SMSC (Short Message Service Centre), perform the functions of routing, storing and subsequent delivery of SMS messages to end users. In the case of inter-operator interaction, messages pass through international SS7 ports, where attempts at route spoofing and Caller ID manipulation are possible. A critical threat for operators is the use of SIM boxes, devices that contain a large number of SIM cards and allow SMS to be redirected through illegal gateways, bypassing standard tariffs. This leads to significant financial losses for operators, as such SMS clusters operate in A2P mode and are disguised as regular P2P communications. An additional risk is the misuse of SMPP channels, which allows third-party providers (ESMEs, AGGs) to carry out illegal terminal traffic using opaque routing schemes [8].

In the process of detecting fraud, operators need to implement solutions based on the analysis of traffic behavioural patterns, including checking abnormal number range activity, frequency and structure of SMPP requests, as well as detecting non-standard SMS routes. To prevent attacks, it is necessary to use signalling firewalls that can track CLI modifications, identify SMS flows with unusual characteristics, and block attempts to circumvent tariff policies. An effective monitoring system should support real-time anomaly logging using machine learning to automatically update detection rules. Integration with DPI (Deep Packet Inspection) allows for in-depth analysis of signalling traffic content and rapid response to new threats, such as SMPP and SS7 manipulations used by attackers to route traffic through opaque channels [7].

One possible approach to detecting and blocking fraudulent calls is to implement a comprehensive voice traffic policy management and analytics system such as the one offered by Mobileum. Such solutions are based on the integration of several key components, including a Voice Policy Engine (VPE) and a Fraud Management System (FMS), which work in real-time to detect anomalous activity. The architecture has two main components. The first component, the VPE, is responsible for enforcing security policies for voice calls, monitoring calls, processing rules, and analysing suspicious calls [8]. Through the use of STIR/SHAKEN mechanisms, this system is able to effectively detect and counteract caller ID spoofing, which is one of the key methods of fraud in modern telecommunications networks [9].

The second component, the FMS, is a powerful analytics engine that analyses signalling traffic and customer data record (CDR). Using machine learning algorithms and big data analytics, the system can detect anomalies in call routing, analyse fraudulent schemes such as Wangiri and IRSF, and generate recommendations for action. The FMS also performs logging, auditing and data modelling functions, enabling operators to gain a deeper understanding of attacks and adaptively update their security mechanisms. The combination of enforcement and analytics mechanisms ensures effective network protection against a wide range of threats. Integration with signalling nodes such as HLR, HSS, GMSC SBC, and IGW provides the ability to interact with major telecommunications protocols, including SIP, CAMEL/INAP, and MAP, which is critical for detecting fraudulent calls regardless of the technology environment [10].

The architecture (Fig. 5) provides for deep interaction between the VPE and key signalling nodes, including the SBC, CSCF, HLR, HSS, MNP and FNP. This provides the ability to cross-analyse number authenticity, detect spoofed callers, and block anomalous traffic in real time. Caller ID tracing technologies allow you to quickly compare information about the origin of a call, checking the compliance of routing parameters and avoiding geographic masking attacks. The FMS performs global anomaly monitoring, receiving data from CDR records and signalling sources such as TAP, NRTRDE, SS7 and Diameter. Analytical algorithms detect irregularities in traffic behaviour that are typical of Wangiri attacks, IRSF schemes, and automated attempts to bypass billing mechanisms. Dynamic training of models based on big data allows predicting and responding to new types of fraud without the need for manual pre-configuration. An important advantage of the architecture is its integration with SMS gateways, network management systems (NMS) and authentication services. This allows you to extend protection not only to voice traffic, but also to related messaging services, providing end-to-end control over information flows in the mobile network. SMTP alerts and automated reports enable operators to respond to incidents and adjust security settings instantly [10].

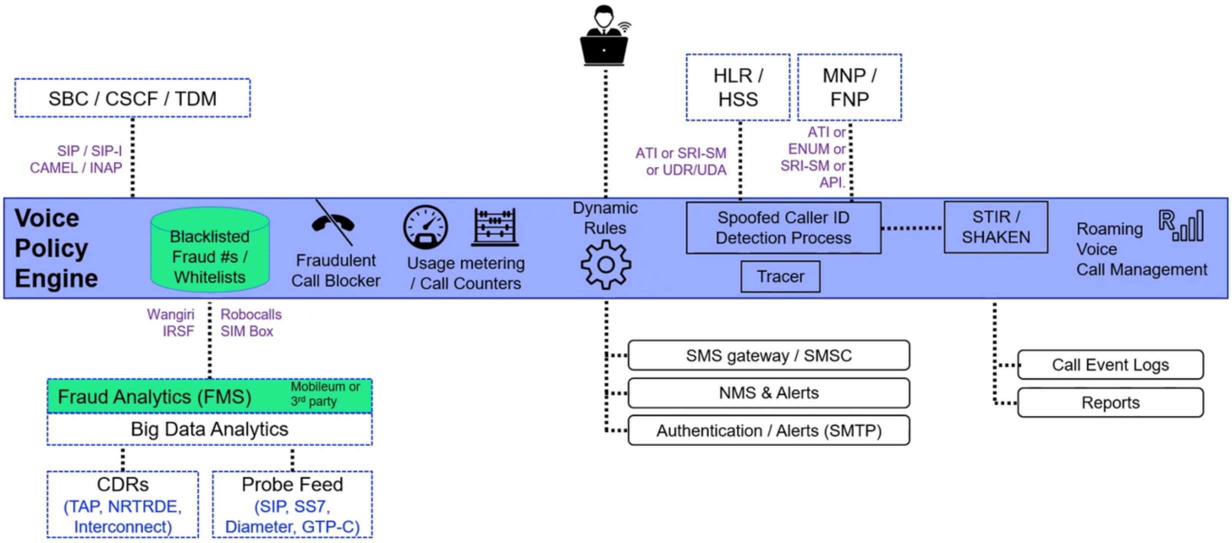

Fig. 5. Integration of Voice Policy Engine (VPE) and Fraud Analytics System (FMS) to combat fraudulent calls

It is worth considering an integrated approach based on Voice Firewall, which combines immediate response mechanisms (Enforcement Engine) and an analytical component (Analytics Engine). The main function of Voice Firewall is real-time monitoring and control of voice traffic at the international and national inter-operator levels, which allows you to block threats and warn subscribers about potentially fraudulent calls. Such solutions not only prevent financial losses for operators, but also ensure compliance with regulatory requirements for consumer protection [8,10].

Enforcement Engine provides a rapid response to detected threats by applying routing and traffic control rules. The system allows you to apply flexible actions to suspicious calls: from blocking with a fixed or random reason code to modifying the CLI, redirecting to IVR or Voicemail, and dynamically configuring scenarios based on analytics. Its functionality includes:

- Blocking fraudulent calls through flexible lists (Deny/Allow/DND), integration with FMS, and the use of various call processing mechanisms (IVR, Voice Captcha).
- Detection of fake Caller ID through number validation, roaming status, correlation with other calls, and certified verification via STIR/SHAKEN.
- Load control through call counting, minutes used, and individual subscriber restrictions.
- Dynamic routing rules that support SRISM/ATI queries to central roaming databases and real-time call routing adjustments.

The analytics engine performs in-depth traffic analysis using:

- Rule-Based Analytics to quickly detect fraudulent calls.
- Artificial intelligence and machine learning (AI/ML Analytics) to predict and adapt protection based on behavioural traffic patterns.
- Case Management mechanisms for maintaining a history and correlating events containing potential risks.
- Data analysis and threat modelling (Data Modeling, Fraud Engines) to identify complex schemes to bypass traditional detection mechanisms [8,10].

One of the most critical aspects of security is the detection of CLI Spoofing, which is used to deceive subscribers and bypass operators' billing mechanisms. The following checks are used for this purpose:

- Empty Caller ID analysis – checking calls with an empty number, which is a characteristic sign of spoofing.
- Number length check – checks whether the MSISDN length matches the minimum and maximum values according to the operator's settings.

- Mirror Call detection – a situation where A-Party and B-Party have the same number, which is a sign of fraud.
- Number Subset Analysis – when a part of the called party's number matches the calling party's number, indicating possible CLI manipulation.
- Checking roaming status and subscriber activity via ATI, SRISM, UDR/UDA, which allows you to track the user's true location and prevent call route tampering.

Intelligent routing and traffic analysis is based on the use of a comprehensive verification mechanism that allows for the accurate location and status of subscribers in mobile networks. One of the key elements of this process is the use of SRI-SM (SIGTRAN) to HLR, which allows to obtain MSC/VLR details and determine the current status of the subscriber. Additionally, the ATI (SIGTRAN) to HLR check is performed, which analyses the Age of Location to help assess whether the subscriber is active in the network. For a deeper analysis, the UDR/UDA (Diameter Sh) mechanism to the HSS is used, which allows making requests to the MME to assess the current status of the subscriber and its movement between networks. In addition, the company continuously monitors subscriber location updates in international links using network probes (SS7/Diameter Probe), which allows timely detection of anomalies in user behaviour and prevention of potential threats [10].

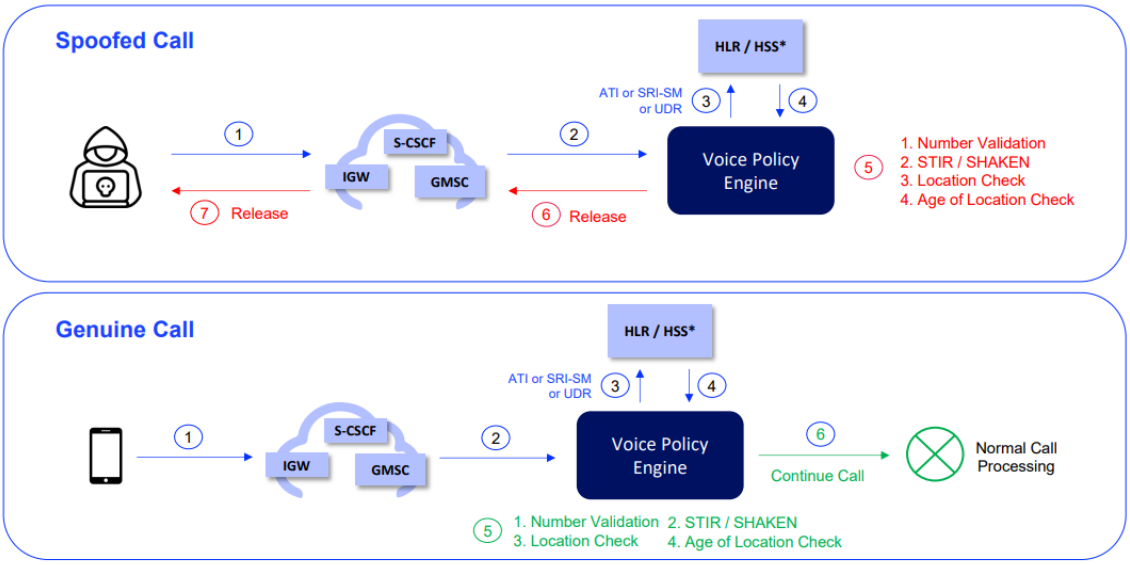

Fig. 6. CLI Spoofing / Refiling Detection Call Flow (* Home Network HLR / HSS or Other Network HLR / HSS in Home Country)

A fraudulent call is initiated by an attacker and transmitted through the Interconnection Gateways (IGWs) before being routed to the S-CSCF and GMSC. At this stage, standard call routing is performed, but before final processing, the request is sent to the Voice Policy Engine (VPE), which performs an in-depth analysis of the call parameters (Fig. 6). The VPE interacts with the HLR/HSS via ATI, SRI-SM or UDR mechanisms to perform several critical checks:

1. Number Validation – comparing the number structure with international and local standards to detect anomalies.

2. STIR/SHAKEN authentication – verification of the digital signature of the call to determine its authenticity and detect attempts to spoof Caller ID.

3. Location Check – compares the physical location of the calling party with the routing information.

4. Age of Location Check – assesses the time of the last location update to determine if the caller is within the expected geographic area [8,10].

If any of the checks fail, the call is classified as potentially fraudulent, and the VPE generates a Release Call command. This allows operators to block attempts by attackers to use fake CLIs to bypass billing mechanisms or conduct social engineering.

When a call is initiated from a trusted source, it passes through similar routing nodes (IGW, S-CSCF, GMSC) and is passed to the VPE for verification. The same set of checks is performed in the HLR/HSS, including number verification, STIR/SHAKEN authentication, location and age of last registration. If all the checks pass, the VPE confirms the authenticity of the call and allows it to be routed to (Continue Call), which leads to (Normal Call Processing). This ensures that legitimate callers do not experience delays or false blocking.

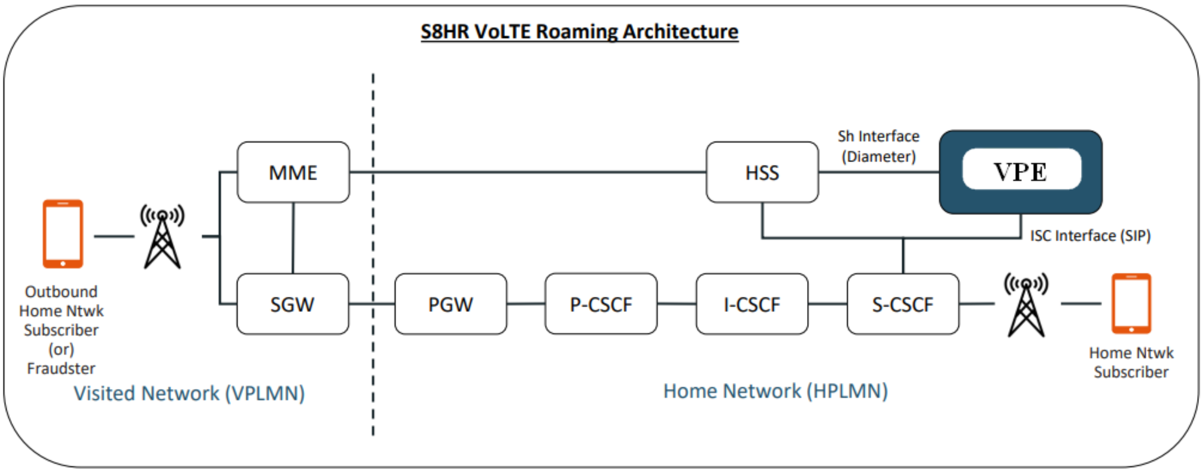

Fig. 7. Roaming status check in S8HR VoLTE roaming scenario

S8HR VoLTE roaming provides end-to-end routing of voice calls through the operator's home network, which allows for centralised control of signalling traffic and fraud protection mechanisms (Fig. 7). In this architecture, all calls, regardless of their destination, are routed through the IMS infrastructure of the home network, which minimises the impact of the visited network on call processing. This approach allows operators to implement effective strategies to combat CLI manipulation, international refiling and other types of fraud. An important security element is the verification of roaming status and subscriber identification at the Home Subscriber Server (HSS) level. This process includes data exchange between the HSS and the Voice Policy Engine (VPE) via the Diameter Sh Interface, which allows the user to estimate the location of the user, his or her movement history and recent activity on the network [6]. The VPE analyses call routing parameters and checks them for abnormal patterns typical of fraudulent schemes such as IRSF or Wangiri. Integration with the Session Initiation Protocol (SIP) allows the use of additional authentication mechanisms, including STIR/SHAKEN, to verify the authenticity of Caller ID. One of the key challenges in S8HR VoLTE is controlling rate bypass, when fraudsters use roaming routes to disguise outgoing international calls. In this case, the VPE analyses signal traffic for atypical calls with fake identifiers and checks whether the subscriber's registration information matches their actual location [9]. If inconsistencies are detected, the system can automatically block suspicious calls or reroute them for additional analysis. Using S8HR in conjunction with VPE provides mobile operators with an effective tool to detect and prevent fraudulent calls, as well as to protect against revenue losses due to the misuse of international interconnect connections. Through centralized alarm monitoring, operators can not only reduce the risks associated with caller ID spoofing and cross-network attacks, but also improve the quality of service for legitimate subscribers.

Call Flow Analysis for Flash Calls and CLI Spoofing. The following call flow diagram (Fig. 8) illustrates the handling of Flash Calls and detection of CLI spoofing attempts within a modern telecommunications network architecture. It demonstrates how various scenarios, including legitimate call handling, call blocking, and timeout handling, are addressed using a Voice Firewall (Voice FW) integrated with the Session Border Controller (SBC) and Home Location Register/Home Subscriber Server (HLR/HSS) [12].

The diagram outlines the processes involved in managing Flash Calls and CLI spoofing attempts through three distinct scenarios: Normal Scenario, Blocking Scenario, Default Timeout Scenario.

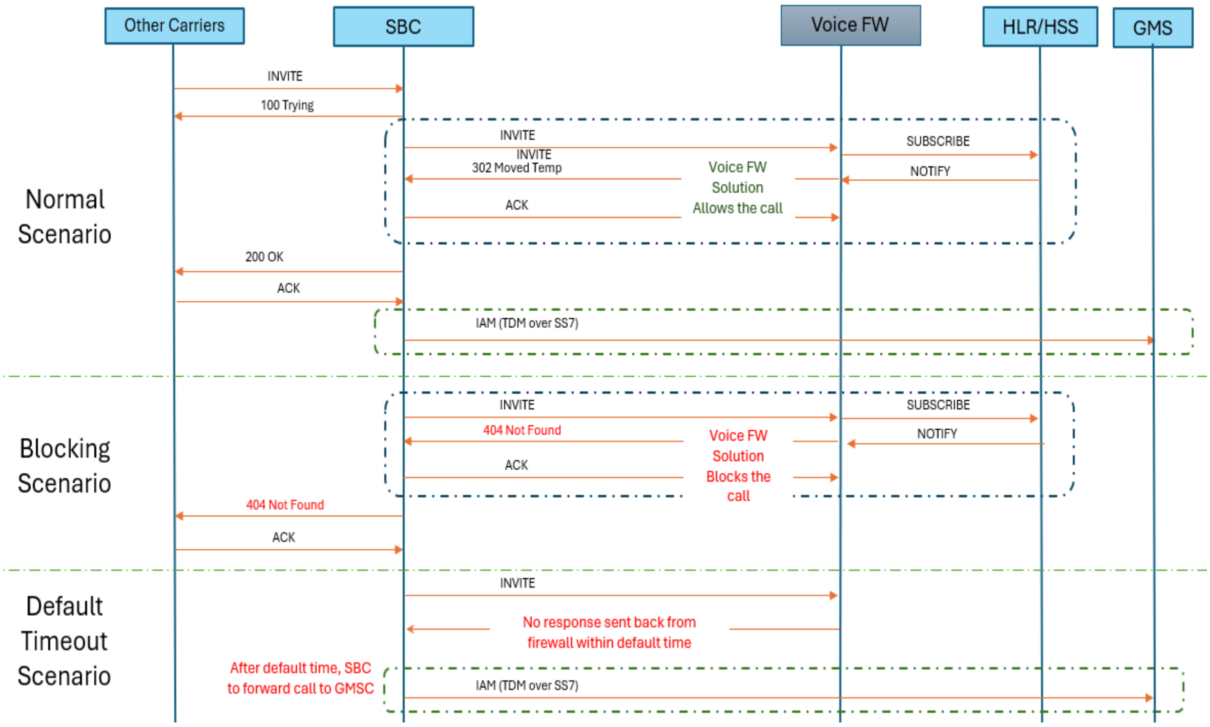

Fig. 8. Call Flow for Flash Calls and CLI Spoofing

Normal Scenario describes the typical handling of legitimate calls that pass through the network's security infrastructure without triggering any alert mechanisms. The process begins with an INVITE request sent from other carriers to the Session Border Controller (SBC). The SBC immediately relays this request to the Voice Firewall for further analysis [13]. Upon receiving the request, the Voice Firewall initiates a thorough validation process, which includes:

- Authentication and verification of the Calling Line Identification (CLI) using protocols such as STIR/SHAKEN to ensure the integrity of the caller ID.
- Cross-referencing the incoming call's details with known trusted sources and whitelisted ranges to detect any discrepancies.
- Applying Artificial Intelligence (AI) and Machine Learning (ML) algorithms to analyze patterns and determine whether the call's attributes match legitimate profiles [11].

If the call passes all validation checks, the Voice Firewall allows the call to proceed, returning a 302 Moved Temp response, followed by an ACK from the SBC. The SBC then sends an IAM (Initial Address Message) using TDM over SS7 towards the GMSC for final processing and call establishment. The successful processing of the call signifies that the network infrastructure is correctly identifying and authorizing genuine traffic, ensuring high-quality service without compromising security.

Blocking Scenario illustrates how the network architecture reacts when a potential security breach or fraudulent activity is detected. The process begins similarly, with the INVITE request being forwarded from the SBC to the Voice Firewall. The Voice Firewall initiates comprehensive analysis by performing the following checks:

- Inspecting the CLI for signs of manipulation or spoofing, including empty numbers, incorrect length, or geographical masking attempts.
- Utilizing advanced pattern recognition techniques driven by AI/ML models to detect abnormal calling behavior or volume anomalies.
- Comparing the call's origin against blacklists containing known fraudulent numbers or previously identified malicious sources.
- Validating the caller's identity using STIR/SHAKEN protocols to confirm that the digital signatures match expected standards [9].

Upon detecting suspicious activity, the Voice Firewall returns a 404 Not Found response, indicating that the call has been blocked. This blocking mechanism provides immediate protection against unauthorized calls, preventing them from progressing further within the network architecture. The SBC acknowledges the block with an ACK message, ensuring that the attempt is terminated effectively and recorded for future analysis.

Default Timeout Scenario highlights potential vulnerabilities associated with timeouts and the fallback mechanisms that networks must employ to ensure reliability. After the SBC forwards the INVITE request to the Voice Firewall, the network expects a timely response to determine the appropriate course of action. If the Voice Firewall fails to respond within the predefined timeout period, the SBC initiates a fallback procedure:

- The call is automatically forwarded to the GMSC, following the standard procedure for unverified calls.
- This can occur due to network congestion, processing delays, or inefficiencies within the Voice Firewall itself.
- Allowing calls to proceed without thorough validation poses significant risks, particularly when handling Flash Calls or spoofed CLI attempts.

Such scenarios necessitate the implementation of adaptive timeout management systems capable of dynamically adjusting thresholds based on real-time traffic conditions and network load.

Analysis of Flash Calls Traffic and Blocking Results. The analysis of flash call traffic statistics gathered throughout 2024 provides crucial insights into the scale of flash call activity and its impact on various network operators globally. These statistics, shared by GSMA and collected from several operators worldwide, highlight the growing prominence of flash call usage and the challenges associated with effectively managing this phenomenon [2].

Flash call traffic comprises a substantial portion of incoming telecommunications traffic, with 10.16% of all inbound traffic identified as flash calls. This metric underscores the prevalence of flash calls as a method of bypassing traditional authentication channels, particularly in applications that employ flash calls as an alternative to SMS-based OTPs. Additionally, the statistics reveal that approximately 78% of all flash calls originate from US prefixes, indicating a significant concentration of traffic from a specific geographical region [2]. This trend suggests that many authentication services are leveraging US-based numbers to initiate flash calls, either for legitimate purposes or as part of fraudulent schemes. Furthermore, 34% of countries report that the majority of flash calls are generated using locally spoofed numbers. This tactic allows attackers to disguise the true origin of a flash call, complicating efforts to detect and prevent fraudulent activity. The use of local spoofing also presents unique challenges for network operators attempting to differentiate between legitimate authentication calls and malicious attempts to exploit the network [2].

The bar chart titled “Top flash call number ranges excluding US” (Fig. 9) provides a breakdown of flash call activity by country, with Italy, the UK, and Mexico leading the list with 3.00%, 2.75%, and 2.50% of the overall flash call traffic, respectively. Other countries such as Brazil, Saudi Arabia, UAE, and Nigeria also contribute significantly to the global flash call volume. This data emphasizes the widespread nature of flash call usage across diverse geographical regions and the need for a unified approach to combating associated threats [2].

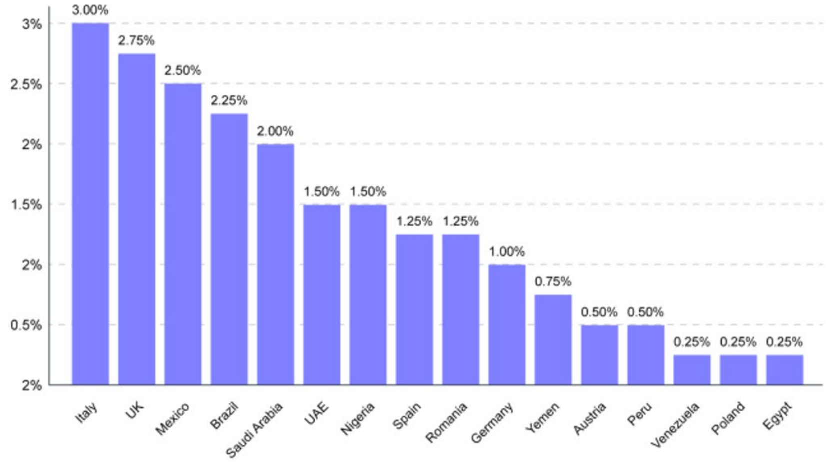

Fig. 9. Top flash call number ranges excluding US

The second set of graphs (Fig. 10) illustrates the dramatic impact of emergency call blocking on traffic on various Meta platforms, including Facebook, WhatsApp, Instagram, as well as the entire Meta ecosystem. The research was conducted on the equipment of telecommunications vendor GMS, using a modern solution based on Voice Firewall. The start of blocking “fast” calls, introduced at the end of December 2023, led to a significant increase in traffic on all monitored platforms.

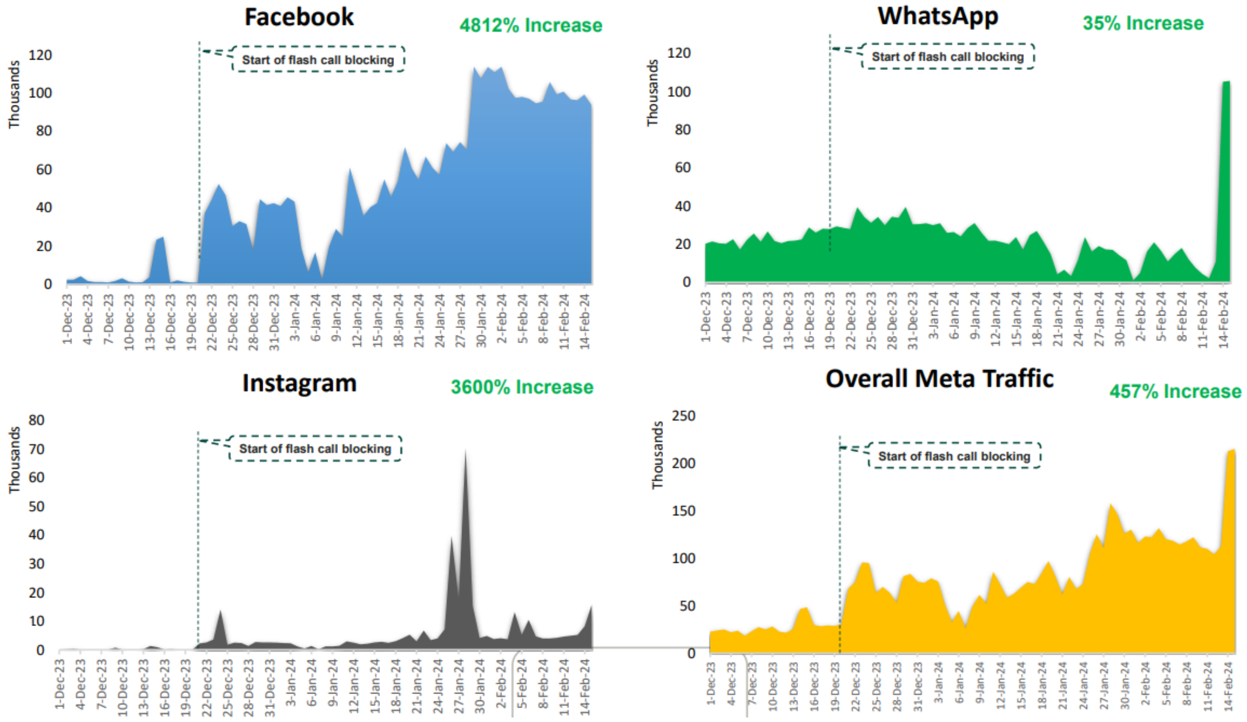

Fig. 10. Results of blocking flash calls in all Meta services

The most significant increase was observed on Facebook, with traffic volumes spiking by 4812% following the initiation of flash call blocking. This exponential growth reflects the platform’s reliance on alternative authentication mechanisms after flash call blocking measures were enforced. Instagram experienced a dramatic increase of 3600% in traffic, indicating a similar dependency on flash calls for authentication processes. The noticeable peaks in traffic suggest periods of particularly high demand for user verification services. Although the increase in WhatsApp traffic was comparatively modest at 35%, the consistent upward trend highlights the adaptation of the platform to alternative methods of traffic routing and authentication. The cumulative effect across all Meta platforms resulted in a 457% increase in traffic. This substantial growth demonstrates the broader impact of flash call blocking on the Meta ecosystem, suggesting that flash call authentication played a significant role in user verification across these services [15].

These findings indicate that flash call blocking mechanisms implemented by operators have led to considerable changes in traffic patterns across popular platforms. The sudden shift in traffic underscores the importance of implementing robust detection and prevention mechanisms that can adapt to rapidly evolving traffic conditions. Moreover, these statistics highlight the need for collaborative efforts among operators, application developers, and industry regulators to develop comprehensive solutions that can effectively manage the challenges associated with flash call traffic.

Conclusions

The article analyses current challenges and identifies approaches to countering Flash Calls in telecommunications networks. Flash Calls, as a relatively new authentication method, are used to bypass SMS messages and other traditional means of verification. However, this method poses significant threats to telecommunications operators due to loss of revenue, fraudulent attacks and difficulty in detection due to the short duration of calls and manipulation of Caller ID. Research shows that existing security systems are often unable to effectively detect Flash Calls due to their specific nature, including extremely short call durations, automated high-frequency traffic flows, and the use of spoofed or masked numbers. Typical methods used by attackers include Caller ID spoofing, the use of SIM farms, VoIP infrastructure and geographic masking, which makes it difficult to track and counter. Developing and implementing effective mechanisms to counter Flash Calls requires a multi-layered approach, including:

- Integration of artificial intelligence and machine learning systems to detect anomalous calls that deviate from normal user behaviour patterns.
- Adaptive filtering policies that allow you to configure thresholds and system responses to suspicious calls without interrupting legitimate traffic.
- Implementation of Caller ID authentication and integrity checking protocols to ensure the authenticity of call origin.
- Implementation of distributed mechanisms for high availability and geo-redundancy to ensure uninterrupted operation of systems in the face of large-scale attacks.

The proposed approach emphasizes the need for seamless integration with existing network management systems to provide end-to-end protection. A robust security framework should be capable of detecting and mitigating Flash Calls in real time while preserving the quality of legitimate services. Moreover, international cooperation among telecom operators, application developers, and industry regulators is crucial for developing unified standards and protocols to combat Flash Calls effectively. Future research should focus on evaluating the proposed methods in real-world deployments, enhancing the adaptability of AI-driven models, and exploring new techniques for proactive fraud detection. By continually updating and refining detection systems, telecommunications networks can remain resilient against evolving threats and provide secure and reliable services to their users.

References:

1. Hitchins D. What are flash calls and how do they work? Infobip. URL: https://www.infobip.com/blog/what-is-a-flash-call.
2. Taylor L. CFCA 2021 Global Fraud Loss Survey. New York, NY, USA : CSFA, 2021. 67 p, URL: https://cfca.org/document/2021-fraud-loss-survey/.
3. GSMA. Flash Call Traffic Analysis Report 2024. GSMA Intelligence.
4. Sinapsio. Wangiri Fraud and Flash Calls | Betatel LTD. Blog. URL: https://api.betatel.com/blog/wangiri-fraud-and-flash-calls.
5. Vetoshko I.P., Kravchuk S.O. Opportunities to Improve the Quality of Voice Services in 5G Networks // 2023 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), ISBN: 979-8-3503-4848-4, 13-18 November 2023, Kyiv, Ukraine. https://doi.org/10.1109/UkrMiCo61577.2023.10380376.
6. 3GPP TS 23.501 version 16.7.0 Release 16. 5G; System architecture for the 5G System (5GS). Effective from 2021-01-21. Official edition. FRANCE : 650 Route des Lucioles F-06921 Sophia Antipolis Cedex, 2021. 451 p.
7. Vetoshko I.P., Kravchuk S.O. Possibilities of improving the voice services quality in 5G networks // Information and Telecommunication Sciences. – 2023. – Vol.14, No 2. – P. 9-16, https://doi.org/10.20535/2411-2976.22023.9-16
8. Flash calls. Mobileum. URL: http://www.mobileum.com/products/risk-management/business-assurance/flash-calls.
9. ATIS-1000080.v004. Signature-based Handling of Asserted information using toKENs (SHAKEN): Governance Model and Certificate Management. Effective from 2021-10-05. Official edition. New York, NY: ATIS Packet Technologies and Systems Committee (PTSC), 2021. 29 p.
10. Voice Firewall. Mobileum. Comprehensive Voice Traffic Policy Management and Analytics System. URL: https://www.mobileum.com/products/roaming-and-core-network/network-services/voice-firewall/.
11. 3GPP TS 31.102 version 17.14.1 Release 17. Universal Mobile Telecommunications System (UMTS); LTE; 5G; Characteristics of the Universal Subscriber Identity Module (USIM)

application. Effective from 2024-10-08. Official edition. FRANCE : 650 Route des Lucioles F-06921 Sophia Antipolis Cedex, 2024. 371 p.

12. Ветошко І.П. Кравчук С.О. Розгортання голосових сервісів у мережах 5G // Grail of Science. – 2023. - № 24. – с. 278–281, https://doi.org/10.36074/grail-of-science.17.02.2023.051.

13. 3GPP TR 33.835 V16.1.0. Study on authentication and key management for applications based on 3GPP credential in 5G. Effective from 2020-07-09. Official edition. FRANCE : 650 Route des Lucioles F-06921 Sophia Antipolis Cedex, 2020. 83 p.

14. Dahlman E., Parkvall S., Sköld J. 5G Standardization. 5G NR: the Next Generation Wireless Access Technology. 2018. 442p. URL: https://doi.org/10.1016/b978-0-12-814323-0.00002-8.

15. Team G. Why Do You Need a Voice Firewall? GMS | AI-driven communications solutions | GMS. URL: https://gms.net/blog/why-do-you-need-a-voice-firewall.

Автор статті

Ветошко Іван – аспірант, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

ORCID: 0000-0002-0009-7610

Author of the article

Vetoshko Ivan – postgraduate, National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine.

ORCID: 0000-0002-0009-7610