Ericsson White Paper
GFMC-18:000011 Uen
July 2022

ERICSSON

Signaling security

Achieving adaptive security in the signaling network

Content

Introduction

According to the Ericsson Mobility Report, June 2022 [1], 5G will account for nearly half of all mobile subscriptions by 2027. This strong growth shows that a solid digital infrastructure is essential for today's society. 5G opens up a wide range of new applications, not only for consumers but for businesses and industries too. Billions of subscribers worldwide are counting on communications service providers (CSPs) to protect their privacy and offer secure services. The same applies to businesses and industries maintaining billions of devices that must be secured and protected from unauthorized access. To maintain a high level of trust with their subscriber base, CSPs must be able to ensure confidentiality, data integrity, accountability, and availability with their service offerings. Furthermore, investing in secure network solutions enables them to gain commercial advantage through the reduction of subscriber churn and the accelerated transition toward new and innovative services on account of a higher level of customer acceptance.

5G is an enabler for business and industry-specific applications. It opens up new business opportunities and acts as a catalyst for digital transformation and Industry 4.0. This is a huge growth market for mobile CSPs and is putting new demands on network security. Private networks put a new demand on the secure connectivity to a distributed infrastructure as required for the internet of things. Secure connectivity is a basic requirement that is described further in the Ericsson white paper on IoT security [2].

CSPs need a robust strategy to protect their networks from known security risks. A typical protection strategy first addresses central routing functions at the network edge. It then broadens to become a defense-in-depth strategy that extends to the target nodes inside the network to provide multilayer protection. While this is a good start, it is recommended that CSPs use advanced analytics as well to raise their level of security protection even further.

Modern security monitoring and analytics tools can reveal known and new security risks, allowing CSPs to take preemptive action and implement the necessary countermeasures before their networks become subjected to attacks. Regular assessments make it possible to continuously identify potential security risks and verify the measures that protect against them. The results from security analytics should be integrated into the security risk assessment to turn unknown security risks into known ones.

The introduction of cloud native network functions increases flexibility, making it easy to add or remove network elements based on CSPs' needs. Achieving effective threat management in a rapidly changing environment requires a high level of process automation to assess vulnerability and address security risks.

The challenge

The evolving threat landscape around today's telecom networks drives the need for innovative threat management solutions at the network level. Legacy networks offering 2G, 3G or 4G services are built on the principle that trusted network elements communicate with each other. Signaling protocols that are used in those networks like the international Signaling System 7 (SS7) standard, including Mobile Application Part (MAP) and IP-based protocols such as Session Initiation Protocol (SIP), Diameter and GPRS Tunneling Protocol (GTP) can be transported via secure tunnels, but it is not mandatory to support the secure transport. 5G networks use HTTP signaling, which is commonly used for internet services. In contrast to legacy networks, 5G has security built in from the start as described in A guide to 5G network security 2.0 [3]. 5G core network functions can authenticate each other and make use of encrypted signaling. Also, the roaming connections are end-to-end security protected. Nevertheless, 5G and legacy networks are still vulnerable to attacks if a node gets compromised, for example, through exploitation of a zero-day vulnerability. Insider threats are also of high concern when a network function is abused by personnel. Network CSPs do offer third parties connectivity to their signaling networks. This allows for the injection of malicious signals through user-to-network interfaces and network-to-network interfaces by trusted network elements. The procedures used to manipulate signaling sequences are widespread. Authentication and encryption are not sufficient to prevent these attacks but trustworthiness of signaling communication relies on the integrity of peering network functions too.

Establishing an adaptive security strategy

To protect their networks from signaling security threats, CSPs should follow a three-step strategy as depicted in Figure 1 that includes adopting a signaling security framework, employing analytics and process automation, and carrying out regular security assessments.

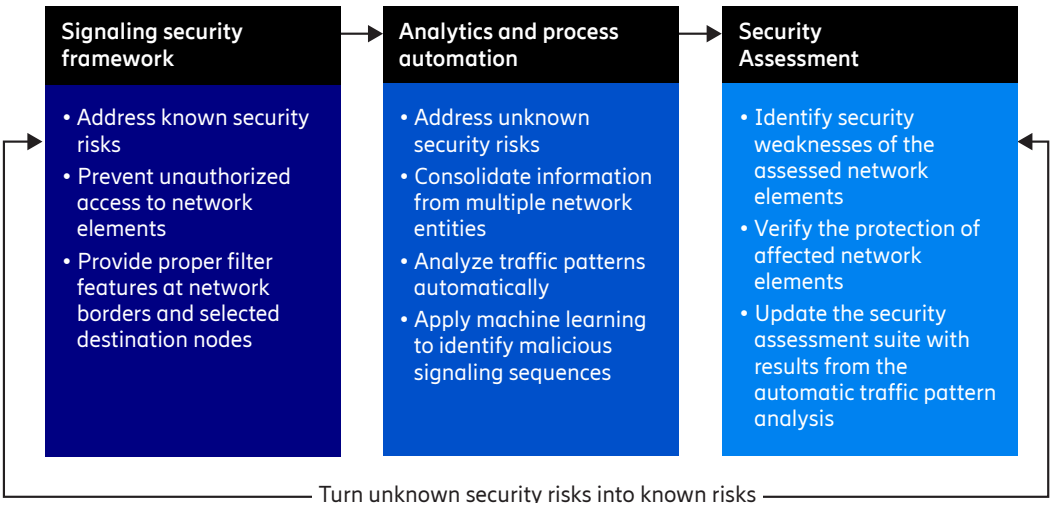

Figure 1. Signaling network - security protection strategy

A signaling security framework makes it easier to address known security risks by preventing unauthorized access to network elements and providing proper filter features at network borders as well as selected destination nodes. Analytics and process automation extends the protection to include multi-vector threats, insider threats and unknown security risks by consolidating data from multiple network entities and interpreting traffic patterns automatically. Analytics and machine learning should also be used to identify malicious signaling sequences.

Performing compliance monitoring on network functions ensures that security configurations are in-line with industrial recommendations and corporate policies . Regular security assessments on network elements enables the identification of known vulnerabilities as well as verification of the protection of these elements. An analytics and process automation suite carries out an automatic traffic pattern analysis of services, evaluating the risk status of these elements, providing feedback, and turning unknown vulnerabilities into known ones.

Step 1: Adopting a signaling security framework

To establish a basis for a secure signaling network, a CSP must protect network equipment from unauthorized access in the following ways:

- Apply proper node hardening methods to all network elements so that unused interfaces are closed, and only authorized network interfaces can be used to establish communication links with the network elements.
- Protect IP connectivity towards the network elements with an IP firewall, so that only authorized network elements can establish connections.
- Take advantage of the new 5G security concepts to apply for Transport Layer Security (TLS) protection on signaling interfaces and Open Authorization (OAuth) to authorize network functions to use network services.
- Perform authorization and authentication of operations and maintenance (O&M) accounts, so that only well-known users can modify the configuration of a node in line with the given permission. Any changes to the node configuration are logged so that they are traceable.
- Configure dedicated network elements to deal with external network signaling traffic, keeping them separated from the network elements that deal with internal network traffic. In this way, if a CSP's network faces a denial-of-service attack from the outside, internal network traffic will not be affected.
- Define the services that can be triggered by third parties with access to the signaling network to protect it from harmful Ingress signaling traffic.

Following the above recommendations typically prevents the CSP's own network from being used as a source of malicious signaling traffic. Nevertheless, the signaling network remains exposed to signaling security risks that can be injected into both user-to-network and network-to-network interfaces. Figure 2 shows how to set up a signaling security framework to protect the CSP's network from known signaling security risks.

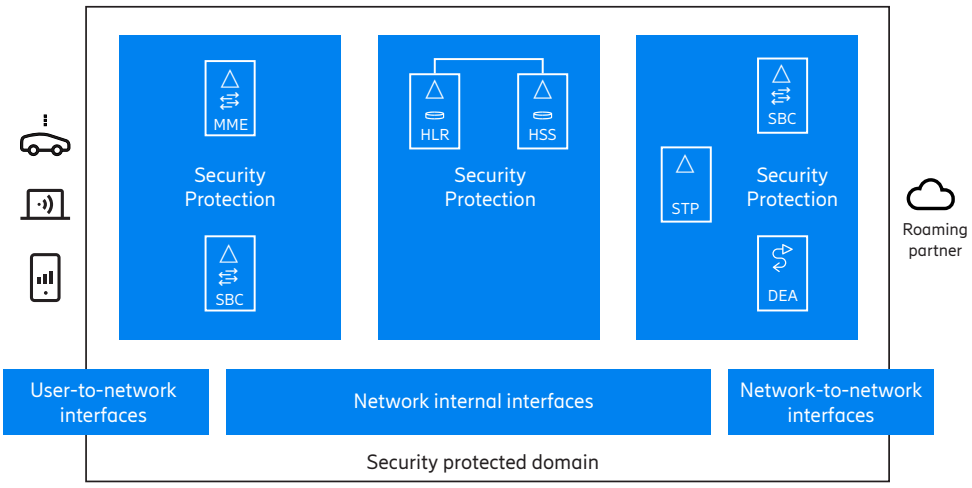

Figure 2. Signaling security framework

The user equipment (UE) needs to be authenticated and authorized before it can access the network through a user-to-network interface. The authentication and authorization can be based on the Subscriber Identity Module (SIM) or the Universal Subscriber Identity Module (USIM) in the user terminal. Other means of authentication are username and password combinations and certificate-based credentials. With 5G the user terminal authentication procedure has been enhanced so that the home network obtains confirmation about a successful authentication of a UE in the visited network.

Verifying the injected signaling procedures is also recommended when users are correctly authenticated and authorized. For instance, in the case of an IP Multimedia Subsystem (IMS), the session border controller (SBC) located at the edge of the network performs signaling and media rate control, and the SIP requests validation and encryption to protect the subscribers' privacy and integrity.

On network-to-network interfaces, CSPs need to verify the trustworthiness of incoming signaling procedures in their own administrative domains. This is typically done in nodes acting as the first point of contact at the edge of the signaling network. The signaling transfer point (STP) acts as the first point of contact for SS7 signaling.

The Diameter Edge Agent (DEA) takes on this role for diameter signaling. SIP signaling from interfacing networks is terminated first in an SBC before it is propagated into the own CSPs' networks.

In 5G, the Security Edge Protection Proxy (SEPP) acts as the first point of contact.

The defense-in-depth principle can be applied in the signaling network as well, introducing an additional layer of security checks in case the first layer is bypassed. Consequently, target nodes such as the Home Location Register (HLR) or Home Subscriber Server (HSS) perform sanity checks on the signaling messages as well to filter out any that are obviously wrong.

Recommended security checks on network-to-network interfaces can be separated into two types: stateless and stateful. Stateless security checks only take into consideration the

message content and internal configuration data. Stateful security checks involve more sophisticated handling processes. A stateful security check is designed to prevent location-based fraud, where voice calls or text messages are redirected, resulting in unlawful interception or impersonation of subscriber identities.

In large network deployments, there are multiple interconnection points to roaming networks. Stateful security checks have to be executed in at all these interconnection points, meaning that the signaling firewall will need an efficient mechanism to synchronize location information on subscribers networkwide. New location information can be received on any of the interconnect points for a dedicated subscriber. Information about the last trusted location must be the same in all the signaling firewalls. Another use case for the synchronization of location events among multiple firewall instances is the cross-check of location events among different mobile network generations. 2G and 3G events are received through SS7 signaling, 4G events are received by diameter signaling and 5G events are received by HTTP/2 signaling. 5G has enhanced the security of subscribers' location events due to the enhanced home control where the authentication information of the UE in the visited network is forwarded to the home network. Consequently, 5G location events are very trustworthy and can be compared against 2G/3G and 4G location events.

Using encrypted signaling transport is a complementary strategy providing additional security in signaling networks. Internet Protocol Security (IPsec), TLS, or Datagram Transport Layer Security (DTLS) provides confidentiality, integrity, authentication, and replay protection for a signaling connection between two peers. External parties cannot read or modify the signaling information. Neighboring peers can be authenticated in a more trustful way, and attackers cannot replay recorded signaling streams to harm the network. In 5G TLS protected signaling connections are specified by the standards right from the start so that interfacing network functions are prepared to use them.

Secure signaling connections can be established between two peers. This works easily on user-to-network interfaces where the communication from a user terminal to a trusted network node can be encrypted. Certain limitations will however become apparent when extending this concept to an end-to-end session involving multiple CSP networks. In 2G, 3G, and 4G networks, end-to-end encryption is not easily possible when intermediate network nodes must read and modify certain information elements of a signaling message to facilitate routing decisions. CSPs can agree on a secure signaling connection at their interconnection links, but none of these CSPs can influence how the signaling is treated behind the agreed security endpoints so that it is possible to continue with an unprotected signaling connection. 5G introduces the concept of an end-to-end protected roaming connection. Roaming interfaces can be either protected using TLS or Application Layer Security based on Protocol for N32 Interconnect Security (PRINS). The end-to-end roaming security concept has been specified as well for 4G networks using Diameter End-to-End Security, but this is not yet widely deployed.

Another issue that counteracts secure signaling transport is the fact that attacks on the signaling infrastructure are launched from trusted network elements. This is possible due to the fact that network nodes get compromised , for example, through exploitation of a zero-day vulnerability.

Considering this limitation, the added value of end-to-end protected roaming connections is to clearly authenticate the remote party so that location fraud cannot be easily committed. Together with a signaling firewall, attacks can be made visible, and countermeasures can be taken to block fraudulent traffic.

Step 2: Employing analytics and process automation

While boundary protection combined with improved 5G security controls increases overall security level, it is still not sufficient for the detection of all kinds of threats and to ensure trustworthiness of signaling network. Complex network functions are the potential targets of supply chain attacks with hidden integration of backdoors. Software stacks comprised of many layers and components are exposed to zero-day vulnerabilities, which may be exploited by attackers. Adversaries may use stolen identities, or disgruntled employees may abuse their privileges for non-legitimate purposes. How can CSPs ensure trustworthy network operations in the face of potentially compromised functions?

Perimeter protection defense approaches build an implicit trust in entities inside the perimeter assuming that attacks are coming from outside. In contrast, Zero Trust approach starts from the point that the adversary is already inside the network and no implicit trust is made in network location or previous verification of an identity. Instead, resource access requests are evaluated on a per-session basis using dynamic policies, and grant decisions are based on the confidence level in the requestor identity and integrity. Trust level computation can be powered by behavioral and environmental attributes of requestors and assets, which requires integrity monitoring, behavior analytics and threat detection capabilities in the network.

For highly secure operations, boundary protection should evolve toward Zero Trust, using a unified security and fraud governance solution, as illustrated in Figure 3. Such a solution provides end-to-end network knowledge for securing assets across different layers and facilitating remediation across all relevant assets.

Understanding adversary behavior is often key for successful threat detection. MITRE's Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) is an industry-wide preferred methodology and knowledge base for describing adversary behaviors. In ATT&CK, tactics represent technical objectives what the adversary wants to achieve, and techniques are the actions performed to achieve those goals. Behavior-based threat detection is oriented around identification of traces of malicious techniques using various sensors. Those sensors could capture and analyze authentication events, message flows, console commands, Operational System calls, system and application logs and so on. Allocation of sensors and analytics are optimized for covering techniques of the behavior knowledge base.

ATT&CK is organized around specific technology domains in which the adversary operates. Adversary techniques in telecom networks can be different than in enterprise networks. CSPs are advised to use a threat detection platform whose knowledge base covers telecom domain specific adversary behaviors, and which provides telecom specific sensors and analytics.

An adversary may reach its objective through multiple consecutive steps, where individual actions may be similar to legitimate network operations. Alerting on all suspicious operations may lead to high level of false positives unless those events are evaluated within a wider context. Efficient threat detection does not only identify but also correlate these weak signals across multiple layers of the protocol and software stacks, hence allows tracing adversary actions over subsequent steps providing superior detection accuracy.

Unified security analytics is evolving toward the aggregation of information elements from different points in the network, and data collection from multiple sources enables nodal information to be combined and thus increases situational awareness at the network level.

Detection of insider threats is of concern with traditional perimeter-based security approaches since the adversary is already located inside a trusted zone and has the necessary privileges. Zero trust approaches may rely on state-of-the-art user behavior analytics methods to tackle insider threats and to decline suspicious resource access requests. User behavior analytics traces activities and attributes associated with various identities and detects suspicious behaviors deviating from the norm. This methodology addresses not only insiders but also misuse of stolen credentials.

The emergence of new types of threats creates new challenges in keeping threat information databases up to date. Anomaly detection techniques, frequently powered by Machine Learning and AI algorithms, can identify abnormalities, drawing security analysts' attention to suspects at an early stage. Machine Learning has also important roles in finding similarities to known bad behaviors and in clustering suspicious ones. With the help of

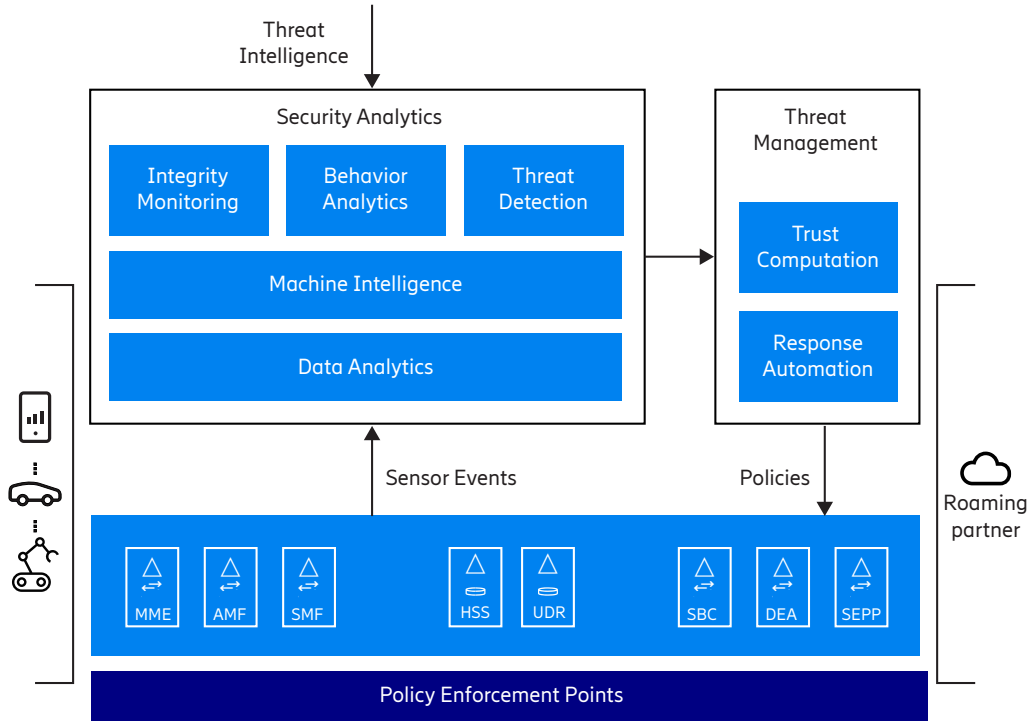

Figure 3. Security analytics and process automation

machine intelligence, input from security analysts as well as from central threat intelligence can be rapidly incorporated into predictive analytics. In this way, threat detection becomes significantly more adaptive compared with a traditional programmatic approach.

Ericsson recommends that network CSPs select a consolidated security analytics solution with the combined power of integrity monitoring, behavior analytics and adversary technique detection. This proactive security approach provides the benefits of end-to-end security risk awareness, sophisticated threat detection capabilities, and significantly shortened mitigation time.

For enhanced protection, CSPs should subscribe to threat intelligence information, which can alert them to globally affected threats and, in some cases, even targeted threats applicable to their realm. Threat intelligence facilitates an understanding of risks and allows threat information to be turned into deployable mitigation actions. CSPs can also decide to share threat information by submitting threat reports.

A high degree of automation is needed to ensure a speedy response to any threat identified. Security process automation and policy orchestration should deploy and adjust security controls dynamically. The process can act upon threats and anomalies that signaling security analytics have identified or the received in threat intelligence reports, and decisions can be made based on confidence level and impact.

Step 3: Carry out regular security assessments

A security assessment is an essential procedure carried out to gain an understanding of the risk level a signaling network is exposed to, and to what extent known security issues are mitigated by the network functions. Two different strategies can be applied: passive monitoring and active attack initiation.

As shown in Figure 4, passive monitoring is based on observance of actual network traffic and reports of known attack scenarios, which make a CSP aware of the actual security risks observed in the network and what countermeasures to take to prevent them.

The passive monitoring approach can be enhanced with an assessment of the node hardening and privacy protection of the network elements involved, covering the following points:

- security policy set definition (at network level), including policies about access control, data masking, hardening, audit logging, and so on
- continous policy compliance monitoring to ensure that security configuration is in-line with corporate security policies
- vulnerability assessments by matching node software level and configuration against vulnerability databases

Applying the strategy of active attack initiation goes a step further, as Figure 5 illustrates.

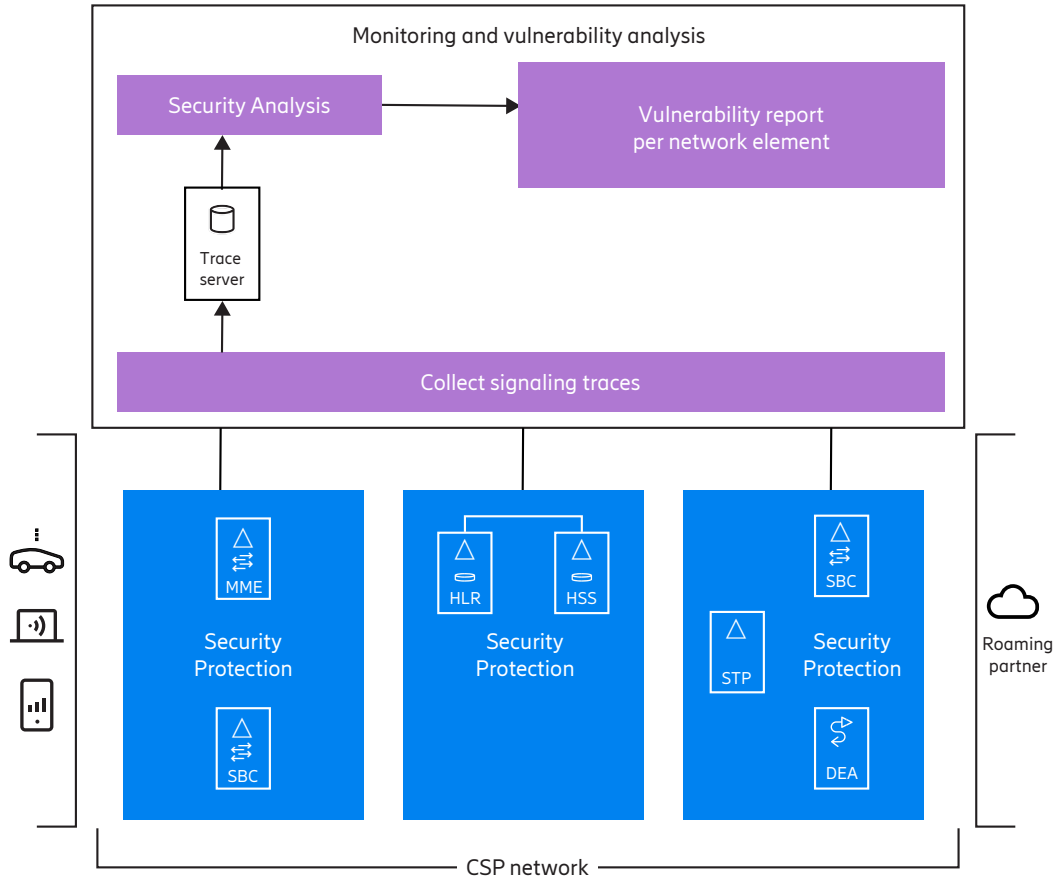

Figure 4. Passive monitoring

Known attack scenarios are targeted toward network nodes from special equipment—either network internally in a kind of lab environment, or network externally in a realistic end-to-end environment. The advantage of this approach is that it is possible to systematically target attack scenarios against the different network entities and verify protection mechanisms against them. Thus, a CSP gets a verified security configuration at the node and network level that can mitigate the injected attack scenarios.

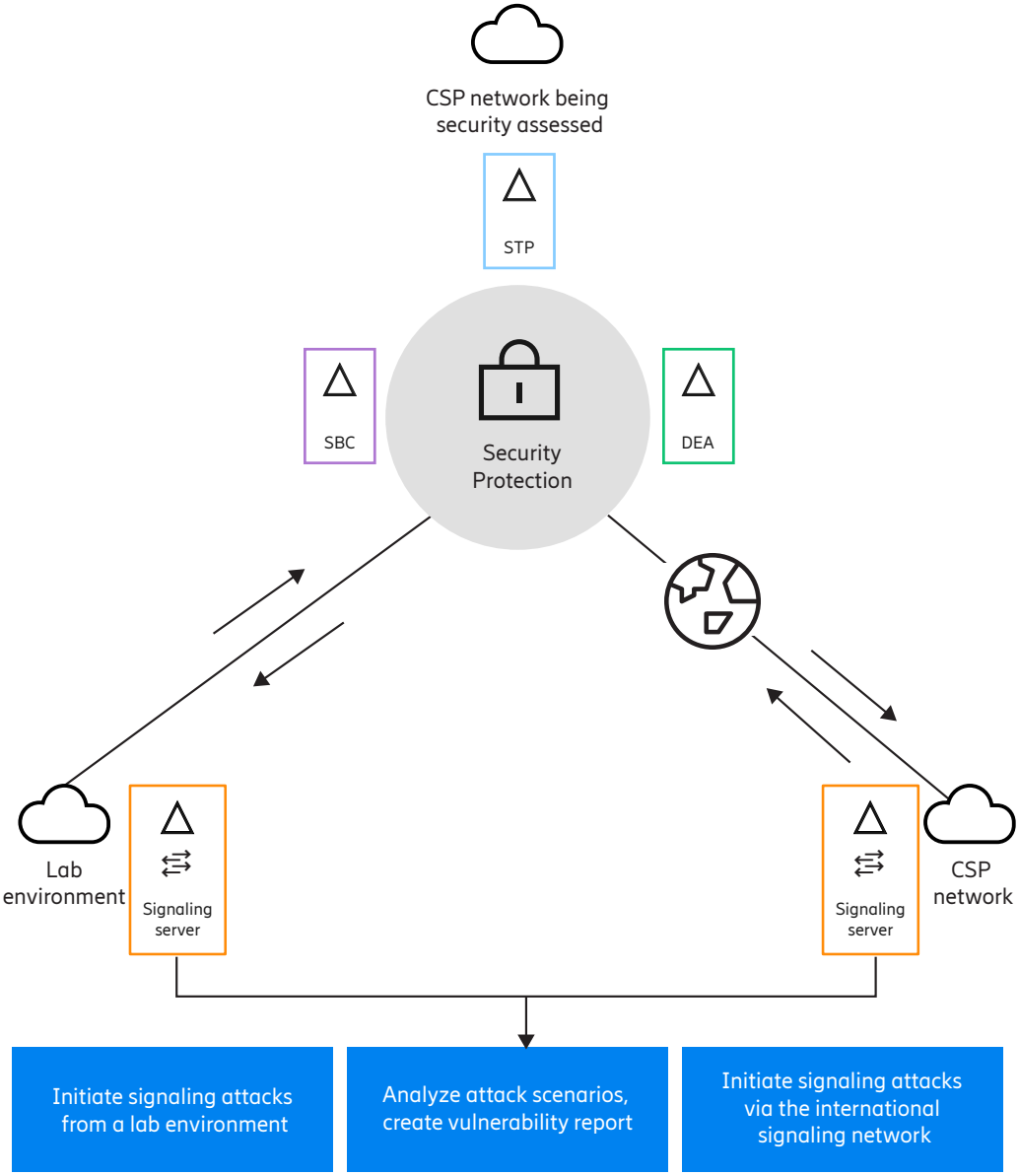

Figure 5. Active attack initiation

Over time, security assessments need to be adapted to the latest known security risk level. Once updated, a security assessment can be reapplied to a CSP network, verifying that the security measures are sufficient to protect the network from the newly identified security risks.

Conclusion

An innovative adaptive security strategy is required to protect CSP assets from a diverse range of security threats to the signaling network from the interception of private communications or location information to the takeover of user accounts to initiate money transfers, to denial-of-service attacks. The recommended approach consists of three steps:

- adopting a signaling security framework
- employing analytics and process automation
- carrying out regular security assessments

The first priority for a CSP is to prevent unauthorized access to the network entities and to block all known security attacks either at the network border or at targeted destination nodes. 5G networks have adopted well-proven security features that provide additional opportunities to protect networks from fraudulent traffic. Transition to Zero Trust approach can bring security to the next level if efficient trust level computations are supported by advanced analytics capabilities. Unknown and more sophisticated attacks can be detected by a unified security and fraud governance solution that provides end-to-end network knowledge to secure the CSP's assets by consolidating information from different network elements. Adversary behavior driven threat detection and machine intelligence powered data analytics can extract threat signatures from the data collected. This process allows a high level of automation and is highly relevant given the increased flexibility of CSPs' telecommunication networks and their migration to virtual network solutions. Finally, carrying out security assessments regularly ensures that the protection mechanisms for the threat signatures identified remain in place.

Glossary

ATT&CK	Adversarial Tactics, Techniques, and Common Knowledge
CSP	Communications service provider
DEA	Diameter Edge Agent
DESS	Diameter End-to-end Security
DTLS	Datagram Transport Layer Security
GPRS	General Packet Radio Service
HLR	Home Location Register
HSS	Home Subscriber Server
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPsec	Internet Protocol Security
MAP	Mobile Application Part
SS7	Signaling System 7
STP	Signaling Transfer Point
TLS	Transport Layer Security
USIM	Universal Subscriber Identity Module

References

1. Ericsson, Ericsson Mobility Report, June 2022, available at: https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf
2. Ericsson, IoT, available at: https://www.ericsson.com/en/internet-of-things/iot-security
3. A guide to 5G network security 2.0, available at: https://www.ericsson.com/en/security/a-guide-to-5g-network-security

Authors

Gergely Matefi is a System Architect at Security Solutions within Business Area Technologies and New Businesses. He has gained 21 years of experience at Ericsson ranging from packet QoS, media processing, over-the-air synchronization and cloud technologies, through his various system architecture design, technology exploration and standardization assignments. In his current position, he is responsible for the evolution of telecom security analytics architecture. His focus is on end-to-end automation of threat detection and mitigation loops. Matefi holds an M.Sc. in information technology from Budapest University of Technology and Economics, Hungary.

Michael Stief joined Ericsson in 1994 and has worked on system management and product management assignments for various wireless and wireline applications over the years. He is currently working as Technical Product manager for Signaling within Solution Line Communication Services, where he is technically responsible for Diameter, SS7 and HTTP signaling products, including Diameter Signaling Controller (DSC), IP-Signaling Transfer Point (IP-STP) and cloud native Signaling Controller (SC). Stief graduated from the Technical University of Dortmund, Germany with a degree in electronic engineering. Academy, an ITU affiliate that transfers knowledge to emerging markets through accredited academic institutions.