Wireless Personal Communications (2021) 119:97–116
https://doi.org/10.1007/s11277-021-08200-0

On Threats to the 5G Service Based Architecture

Geir M. Køien[1]

Accepted: 29 January 2021 / Published online: 19 February 2021

Abstract
The 3GPP-based 5G System marks a clear departure form the previous generations. There is a new radio system and a complete overhaul of the core network design. The core network is redesigned both on the control plane parts and the transport plane. The control plane signalling within the core network is now largely based on the service based architecture (SBA) design, featuring Web-based technologies and the associated security solutions. In this paper we conduct a preliminary generic survey of threats to the SBA.

Keywords 5G security · Core network · Service based architecture · N32 roaming interface · Threats · Risk · Robustness

1 Introduction

The modern mobile systems are critical infrastructures and provide essential services to citizens and businesses alike. A mobile system consists of two major parts: the core network and the access network. The core network is a set of servers that provide macro-mobility services and hosts home network databases, etc. The access networks are comprised of radio access networks (RANs) and the tranceivers (base-stations, access points).

This paper is mainly concerned with system signaling within the core network, and in particular we investigate the basics of the so-called Service Based Architecture found in the 5G core networks. The emphasis here is on security and reliability of the chosen technologies.

1.1 The First Generation (1G)

The mobile systems have come a long way since the first generation systems, which became operational during the early 1980ies. Core network (CN) signalling was based on Signalling System No.7 (SS7). There were few threats to 1G CN signaling and there were no security protection mechanisms in place.

✉ Geir M. Køien
geir.koien@usn.no

[1] University of South-Eastern Norway (USN), Campus Vestfold, Norway

1.2 The Second Generation (2G)

The original GSM system did have security features, but these were oriented towards- and limited to securing the over-the-air interface. The SS7 system has never had any built-in security, and this vulnerability was known early on [1]. In GSM, mobility signaling was handled by the SS7-based MAP protocol (Phase 1 – TS 09.02 [2], latest TS 29.002 [3]). The MAP protocol is based on transaction handling and use of remote operations functionality, and it served its purpose very well. Early on, GSM was mostly a European system and there were usually only one mobile operator per country. Trust was not considered much of an issue and SS7 did not support any security features anyway. The roaming model therefore assumed that the home network would have complete trust in the visited network. GSM became operational around 1991.

1.3 The Third Generation (3G)

The 3G systems are direct successors of the 2G systems. The 3G systems were designed to be digital and to support broadband IP-based data traffic. Security-wise, one extended and improved on what was already there. One also made a profile of the IPsec protocols, called NDS/IP, in order to secure the IP-based GPRS Tunneling Protocol (GTP) system signaling protocol [4–6]. The GTP protocol, together with the MAP protocol, makes up the core network mobility handling protocols in the early versions of UMTS. The use of NDS/IP [7] was however optional, and in the end this has meant that the roaming signalling was rarely protected. Eventually, one also had the option of using the AAA framework [8] and a dedicated application for performing the MAP functionality [9]. The application was in effect a transcoded version of the MAP protocol.

For the 3G system, the roaming situation was becoming a lot more complex. There were multiple operators per country, and the systems were in global use. It became too cumbersome and complex for many of the smaller operators to work out roaming agreements with all possible roaming partners. Thus, the need for roaming brokers were born. Originally these brokers would be called GPRS Roaming Exchange (GRX) operators [10]. Later, they would be known as IP exchange (IPX) operators [11]. The GRX/IPX operators commonly inspects and sometimes rearrange signalling exchanges. This is ostensibly to provide improved signalling compatibility between the operators. The UMTS system became operational around 2001.

1.4 The Fourth Generation (4G)

The 4G system architecture no longer retains any SS7 components, and a greenfield 4G system would be an All-IP system. However, most operators deploy mixed generation systems, and so SS7 lives happily on also in a 4G context. Security in 4G is a lot better than in 3G, but many of the problems that stem from the original mobility model are still with us. Roaming is based on the AAA-based application [9], but MAP is still with us for backwards compatibility reasons. The IPX regime still exists, and use of NDS/IP is still only an option. The 4G system became operational approximately 10 years subsequent to 3G, and generally coexists with the 2G/3G systems.

1.5 On the Scope of the Preliminary Threat Analysis

Since this is a preliminary threat analysis, we shall not attempt a complete analysis. Only the high-level aspects will be covered, and we will only focus on threats related to the 5G Service Based Architecture (SBA).

2 Features of the 3GPP-based 5G System Architecture

The overall 5G system architecture is specified in TS 23.501 "System architecture for the 5G System (5GS)" [12]. The primary security specification is TS 33.501 "System architecture for the 5G System (5GS)" [13].

2.1 5G New Radio

The new 5G radio system, dubbed "New Radio (NR)", is a main feature of the 5G system. The Service Based Architecture (SBA) and the subscriber roaming aspects are not influenced by radio system design. We shall therefore not delve deeper into NR and its characteristics in this paper.

2.2 The SDN/NFV and MANO Aspects of the Core Network

Software Defined Networking (SDN) and the associated Network Function Virtualization (NFV) are very important aspects of 5G deployment. They are considered fundamental to achieving scalability and efficient operation of future network, and they are standardized for 5G [14]. The SBA components are agnostic as to how the lower layers of the networking is realized, and SBA is functionally independent of SDN/NFV (and MANO). We shall therefore not delve deeper into SDN/NFV or MANO functionality in this paper.

2.3 The Web-based Technological Foundation

2.3.1 Representational State Transfer (REST) Technologies

The SBA and the core network signaling is to a large extent based on Representational State Transfer (REST) technologies. The REST API scheme was originally conceived by Roy T. Fielding in his PhD dissertation (Chapter 5 in [15]). Concerning the 5G architecture as depicted in Fig. 1, all the *Nxxx* interfaces (functions), including the N32 interface, are based on REST technologies. This means that web technologies dominates the mobility handling.

2.3.2 JavaScript Object Notation (JSON)

The information transfer associated with REST is normally encoded in the "Javascript object notation" (JSON). JSON is no longer tied to Javascript, although it still reflects the Javascript types and object model. What JSON does is that it provides a text-based data

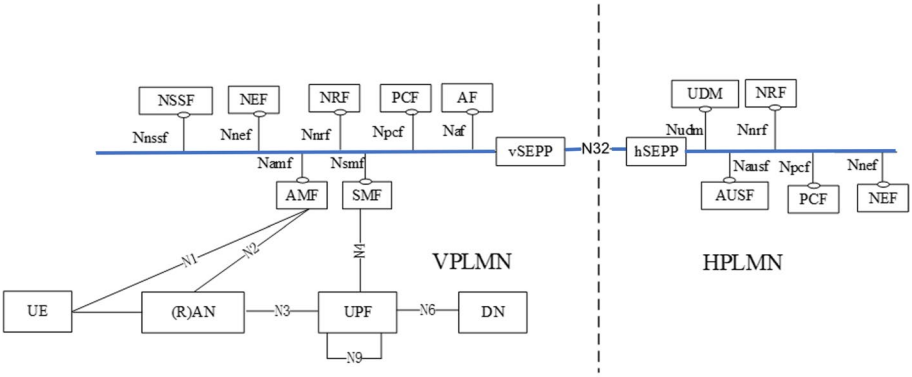

Figure 4.2.4-1 Roaming 5G System architecture- local breakout scenario in service-based interface representation

Fig. 1 The 5G Roaming architecture (Transposed and extended from TS 23.501 [12])

exchange format for encoding different kinds of data objects. In 2G-4G, data encoding was by means of the Abstract Syntax Notation.1 (ASN.1) [16] and for AAA/DIAMETER, the AVP encoding.

2.3.3 Transport Layer Security (TLS)

Web-technologies will normally rely on the Transport Layer Security (TLS) schemes for security protection (TLS 1.2 [17] or TLS 1.3 [18]). The 3GPP facilitates and promotes the use of TLS for the Service Based Interfaces (SBI). However, while support for TLS is mandatory (vendor), it is not mandatory to use TLS for the operators (See Section 13.1.0 in [13]).

2.3.4 Javascript Object Signing and Encryption (JOSE)

The roaming model in 5G includes the IPX roaming brokers, and this means that end-to-end security for IPX-based roaming traffic is infeasible. To respond to this problem, one has defined a separate security scheme for use over the N32 roaming interface. Here, one intends to use the Javascript Object Signing and Encryption (JOSE) scheme for securing the JSON encoded payloads [19, 20]. The scheme permits two IPX entities to become part of the N32 signalling, and the IPX brokers are permitted to modify the messages.

2.3.5 Authorization and Delegation Using OAuth 2

Authorization among network servers was never been much of an issue in previous generations of the 3GPP system. The MAP model with dedicated interfaces implied (and assumed) authorization per interface. It was not a sound assumption, and with the service bus in SBA, implied authorization is not a tenable proposition. The SBA will therefore use the OAuth 2 technology to facilitate access control [13, 21]. Oauth 2 is generally considered to be sound.

2.4 Roaming and the Service Based Architecture

2.4.1 The Service Based Interfaces

The SBA, as depicted in Fig. 1, encompasses both intra-network and inter-network signaling (the roaming part). The SBA communications is over a service bus (depicted in blue). A feature of the message bus approach is that it is agnostic concerning the specific interface.

2.4.2 The Security Edge Protection Proxy (SEPP)

The Security Edge Protection Proxy (SEPP) will act as a network separator. It will be performing filtering and authorization checks on SBA traffic that enters/exits the network. It thus handles access control for roaming traffic.

2.4.3 The SEPP and N32 Roaming Interface

The roaming case features both a visited network and home network version of the SEPP. Figure 2 shows the N32-interface in more detail. Note that while the SEPP, as depicted in Fig. 2, has got new prefixes (c - consumer, p - producer), there are no real changes. The JSON Web Encryption (JWE) [20] and JSON Web Signature (JWS) [19] parts are used, and it involves both the SEPP entities and up to two IPX entities.

3 The Service Based Architecture at Large

This section provides high-level information about the characteristics of SBA.

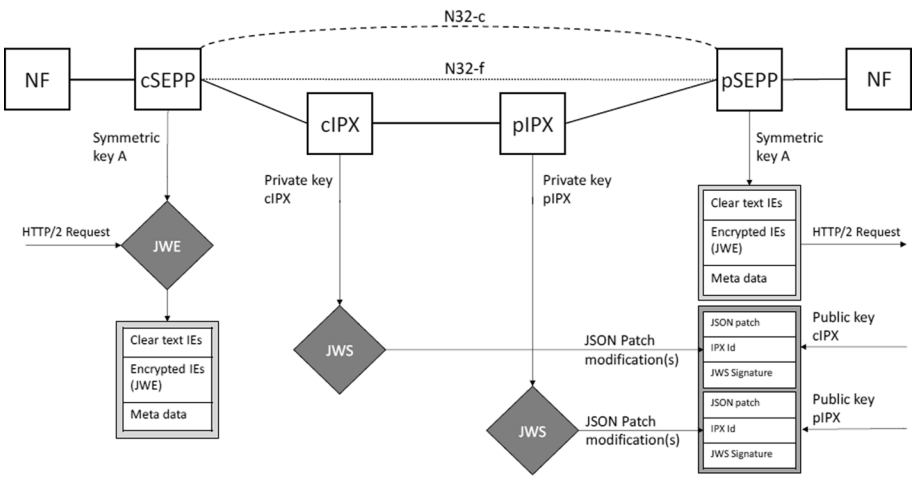

Fig. 2 The N32 roaming interface (transposed from TS 33.501 [13])

3.1 The Critical Infrastructure Perspective

3.1.1 Impact Potential

The SBA is to provide crucial signaling services to a critical infrastructure. There are strong requirements on reliability and availability. There are a lot of software inter-dependencies, with associated version and compatibility bindings. This can potentially make the system brittle and fragile.

3.1.2 Inherent Risks of Complex Systems

The 5G system complexity is reminiscent of the situation described by Perrow [22]. Perrow describes a very different system, namely the Three Mile Island nuclear reactors, which experienced a partial meltdown in 1979. There are many significant differences, but the systemic complexities and the apparent brittleness seems comparable. The term "Normal accidents" is used to denote system accidents that are deemed inevitable in extremely complex systems. There will be cases where multiple innocuous failures can interact with each other. Each of the individual failures may be inconsequential and appear trivial, but may lead to chaotic and unpredictable cascading throughout the system. Perrow identified three conditions that make a system susceptible to these kinds of failure modes. These are:

- The system is complex
- The system is tightly coupled
- The system has catastrophic potential

The 5G system architecture, including the SBA, seems to exhibit all three characteristics. The 5G systems certainly are highly complex systems. There is a lot of for tight coupling, and the design does not appear to try to mitigate this. Given that the 5G systems are poised to become the primary ICT public infrastructure, it is clear that the breakdown of 5G services could lead to a catastrophic outcomes. Many countries also plan to use 4G/5G systems for emergency services communications. This will serve to amplify the catastrophe potential.

3.2 Data Exchange in SBA

Data exchange in SBA is by-and-large by means of JSON encoded data. Almost all modern programming languages has a JSON module, which permits conversion from the internal representation into the JSON text representation.

3.3 The Risks of Using JSON

3.3.1 Abstract Syntax Notation No.1 (ASN.1)

The ASN.1 notation, which is captured by ITU-T recommendation (X.680 series), is used in many of the 2G and 3G protocols. The ASN.1 model for data representation is based on the Type-Length-Value (TLV) scheme. You always start by identifying the data type, then proceed to give the length (in octets) and then the actual data. Everything is explicitly

given, and there is (almost) no room for ambiguity. However, ASN.1 is somewhat cumbersome to use and some of the encodings are less than optimal (that also applies to JSON). But, it works and it is robust.

3.3.2 JSON Text Based Data

JSON uses a text based representation of data and is almost exactly the opposite of ASN.1 with respect to conciseness and rigor. This is not a limitation of text based representations per se, but is a consequence of poor specification quality. In an extreme act of ignorance, the original JSON specification was not given any version information. The author assumed that it would be complete, and therefore that no modification would ever be necessary. The lack of version information is a deliberate break with Principle 10 in [23] (the "explicitness" principle). And, it was very soon evident that the description was neither complete nor consistent.

The case now is that there are several competing, different and incompatible "standard" specifications for JSON. Amongst these are the original "visit card" specification, IETF RFC 4627 "The application/json Media Type for JavaScript Object Notation (JSON)", the ECMAScript 262 (section 15.12) standard (specifies Javascipt and JSON), and the ECMA 404 "The JSON Data Interchange Syntax" (two versions, from 2013 and 2017 respectively). Additionally, the IETF has specified the data interface format (IETF 5158, IETF 7158 and IETF 8259). In addition to this, the 3GPP mandates that the IETF RFC 6902 "JavaScript Object Notation (JSON) Patch" be used.

To add insult to injury, many of the JSON modules and libraries also fail to indicate which version of JSON they are supposedly compliant with. As if this isn't bad enough, it has also been shown that the JSON modules and librates are fairly error prone. Of course, with ambiguous and incomplete specifications that is to be expected. However, it also seems that there are quite a few additional implementation related errors in the libraries. Many of the libraries also seem to make assumptions based on the language that they support. At the core of the problem is that JSON is weak on type description. What is an integer and how is it supposed to be represented? In modern versions of C, you would probably use a `stdint.h` definition, which explicitly defines the integer type (signed or unsigned) and the number of bytes used in the representation. But, there is no easy way to distinguish between different integer types in, like `uint_16` and `int32_t`, in JSON. The problem is pervasive.

In "JSON Parsing Considered Harmful" [24], the author dissects a number of JSON implementations. He analyses many of the more complex scenarios, and the results are fairly disappointing. To quote from the conclusion:

> As a final word, I keep on wondering why "fragile" formats such as HTML, CSS and JSON, or "dangerous" languages such as PHP or JavaScript became so immensely popular. This is probably because they are easy to start with by tweaking contents in a text editor, because of too liberal parsers or interpreters, and seemingly simple specifications. But sometimes, simple specifications just mean hidden complexity.

However, while the results were disappointing, the outcome of the publication has also been that several of the flaws have been fixed. But, while the authors of the modules and libraries can fix implementation flaws, they cannot be expected to fix inherent design problems.

3.4 Challenges to using OAuth 2.0

Authorization and access control is needed for the SBA architecture. The OAuth 2 standard has been chosen as the means to facilitate this. The OAuth 2.0 standard is unfortunately relatively complex and it does introduce new network functions, etc [21, 25]. There is therefore clearly a complexity cost to introducing OAuth 2.0. And, OAuth 2.0 is not without its flaws [26–28]. Some of the reported problems are implementation related, but given that there are few implementation, the problem is nevertheless a real one. Of course, it tends to be much easier to fix implementation flaws than design errors. Deployment and operation of a reasonably complex scheme like OAuth 2.0 can be challenging, and there are a number of issues that can arise from wrongful or incorrect usage [25]. Still, if used properly, the OAuth 2 standard seems to be sound and reasonably robust.

3.5 The Web-based Development and Deployment Model

The Telecom world has a culture which is different form the Web world. This affects how designs are done, how implementations are carried out, the deployment model and the operational regime. In legacy systems, there tend to be long cycles. The deployed software is generally well tested, but fixes and changes to systems in the field can take quite a long time. In the web-model there are frequent releases, and these are dispatched online and often (server-side) in real-time. New functions can be added dynamically. Security updates, etc. can be installed in the same way. From a security point of view, we note that the systems are typically highly exposed, and new vulnerabilities and flaws are routinely reported. But, it is also the case that security fixes and counter-measures can be made available almost instantly. Done properly, "web" security should potentially be better than that for the legacy systems.

3.5.1 Design

During 1G, 2G and partially 3G, the mobile systems were standardized according to very strict regime. The standards were designed with limited input from implementers, and the designs took considerable time to complete. The design bodies where typically CCITT (now ITU-T), CEPT (now ETSI) and others. The 3GPP was formed partially in response to this. One still specifies designs, but with the mantra "adopt, adapt, improve". In particular, IETF standards are being liberally adopted. The 3G and 4G standards shows this quite clearly. For 5G one has gone one step further, and many more of the designs incorporate industry standards. At the same time, the scope has been broadened significantly. That means that the 3GPP now designs and composes architectures, and employs predefined parts and components where possible. To some extent this makes for a patchwork, but it would be infeasible to complete the 5G standards without adopting and adapting other standards.

3.5.2 Implementation (and Testing)

The telecom vendors used to implement most of the systems themselves. This was certainly true for 1G and 2G systems, and to a large extent also for 3G systems. Later, one

had to use externally developed components, frameworks, etc. This became important with 3G, and even more so with 4G. However, it was still a reasonably static regime and the release cycle was fairly long.

With 5G, which is highly "softwarized", there will be substantial changes. Implementation of many of the core components are now oriented around a web-development cycle. This means that there is a different software cycle, more akin to the agile software development model [29]. The web development paradigm will encompass the "RESTful web services" development model, which also focuses strongly on rapid development [30]. One will likely also adopt elements of the "Continuous Integration, Delivery and Deployment" practices [31]. The functionality will be continuously updated, and thus continuously be in flux. This has a huge impact on the implementation regime, and it obviously also affects the deployment and operational phases.

3.5.3 Deployment and Operations

For the previous generations of mobile systems, there was a sharp distinction between deployment (including commissioning) and operations. Deployment was a major event. Operations was what took place when you actively served the subscribers. However, in a web-regime, scheduled fixes and added functionally are released online, and there can be daily updates if one wants. A regime of rapid and frequent changes will naturally require more automation for these processes. The mobile systems are under centralized control, and it is therefore feasible to tailor the whole process to the operational needs. Traditionally, this has meant that new functionality and non-critical security fixes are scheduled on a relatively infrequent basis. The telecom industry will have to change in this respect. Frequent updates is part and parcel of the web development regime, and if one don't adopt it fully one may end up with the worst of both worlds (high exposure and slow response).

4 The N32 Roaming Interface

Figure 2 depicts the N32 roaming model. The N32-interface requirements are primarily covered in clause 13 in [13].

4.1 Overview over the Roaming Model

The roaming model allows for up to two active interconnect operators (IPX) between the cSEPP and the vSEPP. Figure 2 depicts the N32 interface. The trust assumptions between the SEPP and IPX operators follows the business relationships:

- cSEPP has as a business relationship with cIPX.
- pSEPP has as a business relationship with pIPX.

One IPX may be both cIPX and a pIPX in the same roaming exchange. One may also make do without any IPX. The SEPP's will then be directly interacting with each other.

4.2 The N32-interface

The N32-interface comes in two flavors: N32-c and N32-f. The N32-c interface is for control plane messages, while the N32-f is for data traffic. The rules for communications over the N32 interface are quite complex, and involves key exchange, security policies, renegotiations, etc. This encompasses how the messages are encrypted, using the JWE scheme [20]. There are also rules for how the IPX'es may modify the contents of the messages. This is done with the aid of the JWS scheme [19]. The IPX will need to see the contents that they may need to modify.

The security policy rules for of N32-interface is a bit on the complex side. The use of JOSE is intended to be the way one enforces the policies. We note that modifications made by the IPX'es will be visible to the SEPPs, which is an improvement over previous roaming schemes. The parties will know that a change was suggested, and one will know that the suggesting party is authorized to propose the change. However, the rules for handling these cases are quite complex and this could inadvertently introduce inconsistencies. The actual rights may be hard to discern.

4.3 The Risks of Using JSON in an Inter-Vendor Environment

We shall not repeat the arguments in Sect. 3.3, but suffice to say that the problem will potentially be a lot more severe for the roaming interface. This will be an environment with multiple parties, and probably also different implementations of the JSON handling software. For N32, an operator will face all its roaming partners. There will be IPX'es present, and those may filter and mediate, but the risk of incompatibilities will surely be there.

4.4 Security by Ignorance?

The JOSE standard is a relatively new standard. It is also a standard with many cryptographic elements, and it is mainly designed by people without deep cryptographic competence [32]. There is thus some criticism of the design of JOSE and there is also considerable criticism of the JOSE implementations [33–36]. Others have a more balanced view, while noting that the standard is too complex and easily allows for inappropriate use [37]. This does suggest that the goals of JOSE-based security may be hard to achieve in practice. It will be interesting to see if this affects the willingness to deploy JOSE based solutions. Of course, with time, JOSE and current best practices may mature sufficiently to alleviate these worries.

5 High-Level Threats and Risks of 5G

5.1 Risk Assessment on Cybersecurity in 5G networks

The EU and the European Union Agency for Cybersecurity (ENISA) has recognized the criticality of the 5G systems and their role as a major critical infrastructure, and published a report on the risk assessment on security in 5G networks [38]. The identified

threats are mostly very high-level threats, and many threats are more related to 5G dependent technologies than 5G per se. Other aspects noted are:

- The risks associated with the key innovations (like softwarization).
- The role of suppliers and associated dependencies (supply chain).
- Increased exposure to attacks and more potential entry points.
- Increased sensitivity of base stations and management functions.
- Threats to availability and integrity (larger societal dependency).

We fully accept and endorse these concerns.

5.2 The 5G Threat Landscape

5.2.1 The ENISA 5G Threat Landscape Report

ENISA has published an extensive threat landscape report for 5G networks [39]. This report is more comprehensive than [38] and it also investigates the 5G architecture at large. Of particular relevance to this paper is chapter 3, which investigates the 5G use-cases and deployment scenarios. Some aspects receive a lot of attention, while other aspects are only superficially investigated.

- Threats associated with the SDN/NFV and MANO paradigms.
- Network slicing is investigated.
- The individual network functions is briefly investigated.
- The new 5G RAN is investigated.
- Multi-access Edge Computing described.
- The security architecture is briefly explained.
- The 5G physical infrastructure described.

5.2.2 A Noteworthy Omission

Conspicuously, SBA is only briefly mentioned in the ENISA report. It is mentioned that network functions utilize SBA, but without any further investigation. In Annex A, which depicts an "Asset map", SBA is identified as a leaf of the *data* item. There is basically therefore no threat assessment per se of SBA in the ENISA 5G threat landscape report. This is a curious omission, and very peculiar in that the actual performance of the core network signalling so crucially depends on the SBA. Furthermore, it is conspicuous that role of SBA in roaming is not investigated. Actually, it is noteworthy how little is said about roaming at all.

5.3 High-level SBA Threats

The following is a short-list of some high-level threats to the SBA.

- *SBA is a new design* This implies that there are undetected design flaws and shortcomings.
- *A new design implies new implementations* New implementations are highly likely to introduce new implementation flaws.

- *Use of JSON is a liability* Different implementations will use different JSON libraries. There is a considerable chance that there will be inconsistencies, and these *may* lead to security problems.
- *Authorization and OAuth 2.0* Use of authorization is new to the 3GPP core network signaling system. There is therefore a considerable chance that there will be wrongful or inappropriate use. This affects both the design requirements and the realization of the requirements. Furthermore, it is well known that there are problems with some of the OAuth 2.0 implementations.
- *The Telecom industry is not prepared for a web regime* There are undoubtedly Telecom vendors and operators that are well prepared. However, we can safely assume that many will be less than well prepared.
- *Unsecured connections* Appropriate use of TLS will solve this problem. But, use of TLS is up to the discretion of the operator, and the Telecom industry has a poor track record in this area. Many operators will likely get their act together and set up TLS, but even this may not be sufficient. Digital certificates must be managed properly and the TLS implementation must be kept up-to-date. Some operators may opt to not deploy TLS. This will save them money and operational complexity, albeit at the cost of being exposed to all kinds of intrusions.

5.4 High-level N32 Threats

The following is a short-list of some high-level threats associated with the N32 roaming interface.

- *The N32 trust model* The corresponding roaming partner must be inevitably be trusted to some extent. The IPX operators must be fully trusted.
- *Complexity and clarity* The N32 interface is a quite complex. The language is informal, and the structure of TS 33.501 section 13 is somewhat ad-hoc [13]. It is hard to ascertain consistency and completeness when there is no formally specified requirements.
- *JSON inconsistencies* Inherent incompatibilities in a JSON library could be exposed in processing in a multi-operator environment, which would be a typical scenario for the N32 interface.
- *JOSE* The JOSE standard is quite complex, with plenty of opportunities to use it in an inappropriate way. The standard may also have additional problems, which is being alluded to by several pundits. That being said, no actual flaw or exploit has been demonstrated.

6 Web Threat Categories and the Associated Security Measures

What are the high-level threats to SBA and what are the security measures that prevent and mitigate those threats? In this section, we briefly investigates the categories of threats that one should consider. We also provide some suggestions as to security counter-measures.

6.1 Web Application Vulnerabilities

There are several companies/organizations that makes “top 10 vulnerabilities” types of list. The most well-known is probably the OWASP Top 10 list, but it has not been updated

since 2017 [40]. The company Acunetix, which makes security scanner software, publishes annual findings (The 2020 edition is found here [41]). Other similar publications can also be found (e.g. [42]), but the general message is more-or-less the same. That is, missing or inadequate authentication, data integrity problem, data validity and data mishandling problems dominate the picture. The Common Weakness Enumeration (CWE) Top 25, organized by MITRE, is also a well-known list [43]. It is not specifically geared towards web applications, but at software at large. Weaknesses concerning input handling features prominently on this list.

6.2 Threat Categories

Based on the actual problems uncovered by OWASP, Acunetix, MITRE and others, we have come up with the following threat categories. This is by no means an exhaustive list, but should serve well to capture the most serious threats to the SBA services and the areas in which they occur.

1. *Masquerade* These threats would encompass aspects such as identification and entity authentication. For signalling, it also involves message origin authentication aspects. Related to: *entity authentication* and *message origin authentication*.
2. *Eavesdropping* Threats concertizing eavesdropping is about unauthorized disclosure. Related to: *data confidentiality*.
3. *Integrity violations* Threats on integrity entails wilful unauthorized modifications to the messages. This includes insertion and deletion of messages. It also can include message replay and/or message reflections. Related to: *data integrity*, *message integrity* and *message sequence integrity*.
4. *Message data validity* Many attacks on web-based systems are exploiting vulnerability in the parsing and handling of data. Microsoft, in the Security Development Lifecycle (SDL) framework, puts strong emphasis on validation and sanitation of all inputs [44]. Related to: *data input validation*, etc.
5. *Authorization and access rights* Threats towards authorization and access rights includes access violation and illicit privilege elevation. Definition of consistent and complete security policies is a prerequisite. Related to: *Access control*, *authorization* and *security policy*.
6. *Availability and reliability* This encompasses threats to service availability and timely responses. Related to: *Denial-of-Service resilience*, redundancy, real-time properties, offline backups, etc.
7. *Accountability and attributability* This typically includes threats where a party attempts to deny sending or receiving messages. Related to: *log-handling*, *secure audit trail*, *digital forensics*, etc.

One could have added threats such as tracking of transactions, etc., but we have chosen not to include these kind of threats.

6.3 Associated Security Measures (Non-Roaming)

We briefly discuss the security measures that is defined, or is needed, to prevent or mitigate the above defined threat categories. Platform threats and platform security has not been evaluated.

It is noted that IPsec provides security on the network layer. Entity authentication with IPsec is thus related to network layer end-points (which is distinct from web-layer end-points). The IPsec protection will cover all higher-layers, but entities on the higher layers will not be authenticated as such and they will not know whether or not IPsec is used.

6.3.1 Masquerade

The SBA entities will generally be relying on TLS for providing entity authentication. We expect the SBA entity identifers to be included in a TLS certificate. This is an adequate security measure.

However, TLS is not strictly mandatory for use. One alternative, is to use IPsec. The IPsec profile for 3GPP-based systems will either be using a digital certificate or a pre-shared secret to authenticate the end-points [7]. An IPsec end-point is the tuple (IP-address, port number), which can only be indirectly associated with an SBA application entity. As such, the IPsec solution is solid, but not fully appropriate for authenticating an SBA entity.

The standard also permits operators to view the interfaces as trusted as-is, and then not deploy TLS or IPsec.

Status 1 Masquerade is only fully mitigated when TLS is used and when it is used with an SBA entity specific certificate identifier. IPsec is not a fully appropriate solution for SBA entity authentication.

6.3.2 Eavesdropping

The SBA entities will generally be relying on TLS for providing data confidentiality. It is also possible to use IPsec for providing data confidentiality. The drawback to this, is that the SBA application will not be aware of whether IPsec is applied or not. The standard also permits operators to view the interfaces as trusted as-is, and then not deploy TLS or IPsec.

Status 2 Eavesdropping is fully mitigated when TLS or IPsec is used.
However, the SBA enties will not know if IPsec is used.

6.3.3 Integrity Violations

The SBA entities will generally be relying on TLS for providing integrity services. It is also possible to use IPsec for providing the integrity services. The drawback to this, is that the SBA application will not be aware of whether IPsec is applied or not. The standard also permits operators to view the interfaces as trusted as-is, and then not deploy TLS or IPsec.

Status 3 Integrity Violations is fully mitigated when TLS or IPsec is used. However, the SBA enties will not know if IPsec is used.

6.3.4 Message Data Validity

Message data validity cannot be mitigated with the use of TLS or IPsec.

To ensure data validity, one needs to precisely define the data type, the data length and the value. For security protocols, it has long been recognized that these aspects must

be explicitly defined [23]. It is of course, good practice to define the permitted data as accurately as is practically possible for all types of protocols. Indeed, in programming languages, it has long since been recognized that providing type systems will improve the safety, reliability and security of the programs [45]. The theory and practice of defining a "contract" for input/output data dates at least back to the "Design by contract" ideas behind th Eiffel language [46]. Type-systems will help enforcing the contracts.

The ASN.1 encoding scheme [16] provides the Type-Length-Value scheme, and is thus able to provide some help for validating the input data. However, ASN.1 is not used in SBA. Rather, one uses JSON, and it falls short of proving TLV assistance to the receiving entity. One may write parsers that validate the input data for each specific data element used in SBA, and this can probably alleviate the problem. Many data elements would have a well-known representation, implicitly providing type information.

Status 4 Message data validity checking is not facilitated by JSON, but input validation is still possible. One will need verifiers to be defined for each data element used in the SBA exchanges.

6.3.5 Authorization and Access Rights.

Section 13.4 in TS 33.501 [13] details the use of OAuth 2 in SBA. We have not analyzed the specifics of OAuth 2 use in SBA, but if used properly, there is reason to assume that OAuth 2 will be adequate for handling access rights. We note that the standard says that "The authorization framework described in clause 13.4.1 is mandatory to support for NRF and NF.". That is, it is not mandatory for the operators to actually use OAuth 2.

Status 5 Authorization and access rights is handled with OAuth 2. Properly used, this should be an adequate solution. One must ensure that OAuth 2 is actually used.

6.3.6 Availability and Reliability

One may envisage Denial-of-Service (DoS) attacks on OAuth 2 servers and on SBA entities. It is possible to combat these threats. These would be standard techniques for DoS protection, and we shall assume that prudent operators will address these threats.

Status 6 Availability and reliability should be addressed by standard methods. These would largely be non-SBA specific.

6.3.7 Accountability and Attributability

SBA includes a "Network Data Analytics Function" (NWDAF) component [12]. The NWDAF is not a mandatory element of SBA, but it is highly recommended as it will enable logging capabilities. Attributability is partially supported by NWDAF functionality. Additional information may be needed, but this would be outside the scope of SBA.

Status 7 Accountability and attributability, with respect to logging and network analytics, is supported by NWDAF. However, one must ensure that NWDAF is actually used, and that it is use appropriately.

6.4 Associated Security Measures (Roaming)

We note that the N32-interface is fairly complex, and that the many options makes it hard to say exactly what kind of security is actually achieved. And, to a certain extent, that is point of its own. Section 13.5 in TS 33.501 [13] defines the types of security one expects:

- *The so-called PRINS solution (use of JOSE)* This scheme, using JOSE and permitting upto two IPX'es, is defined in Section 13.2 in TS 33.501 [13]. The N32-c (control plane) is to be handled with TLS protection. The actual security of PRINS will vary, suffice to say that there must be a high level of trust in the IPX operators. It will also be difficult to guarantee the level of threat protection actually provided.
- *Use of TLS* This is for the cases where one has a one-to-one connection between operators. OAuth 2 is not used for roaming cases, and it is the SEPPs that needs to handle (roaming-related) policy aspects in this case.
- *"Reserved"*

The observations concerning the high-level threat therefore stands.

Status 8 N32-interface: PRINS case. The threats towards the N32- interface *may* be mitigated by proper use of PRINS (for IPX cases). The actual level achieved will be subject to the negotiations between the involved parties. There is a substantial risk that the achieved level will be wanting.

Status 9 N32-interface: TLS case. The threats towards the N32-interface can be mitigated by proper use of TLS. The SEPP must handle authorization and access control aspects. This solution will potentially be adequate.

6.5 Security Assurance

The 3GPP has developed a set of security testing specifications under the heading "Security Assurance Specification" (SCAS). These specifications are useful in that they defined a minimum level of conformance testing that a 3GPP system should be subjected to. The 5G relevant SCAS specifications are found in the TS 33.5xx series. There are generally one standard per network function, for instance TS 33.517 for the SEPP [47].

The GSM Association (GSMA) has incorporated the SCAS specifications in the so-called Network Equipment Security Assurance Specifications (NESAS). Within the NESAS concept, the GSMA has arranged for security auditor functions, test laboratories, etc. That is, GSMA has provided a framework for third party security testing by certified security testers.

Status 10 Security assurance Full compliance with the SCAS/NESAS specifications should be a requirement for the prudent operator.

7 Discussion

A number of high-level threats towards the SBA and the N32-interface have been identified. These are not particularly novel or unexpected per se, but awareness of these threats are important goal in itself. In particular, given that ENISA's 5G threat landscape report

[39] barely mentioned SBA and roaming, we argue that the awareness of threats to SBA and roaming is key first step towards to mitigating and resolving these threats.

In addition to this, we have identifer a set of threat categories towards the SBA. We note that many threats to SBA can be effectively mitigated by the measures proposed in the standard, but we also note a many of these are not mandatory for use. The JOSE data encoding format is fairly brittle and that it does not facilitate data validity. Data validity concerns therefore become as serious worry for SBA.

Recommendations:

- TLS must be deployed and used.
- The TLS certificates should indicate the actual SBA node type.
- Message data validation schemes *must* be developed and deployed to alleviate JSON's many shortcomings.
- OAuth 2 must be deployed and used. Policy decision must be applied.
- NWDAF must be deployed and used. Logging must be used.
- N32: TLS (or IPsec) must be used for direct inter-operator signalling
- N32: PRINS profiles must be verified for cases with IPX operators.
- N32: Policy profiles (and logging) must be developed (both TLS and PRINS).

Threat modeling should be conducted regularly (operator specific). There is a need for extensive security testing, including, but not limited to, SCAS and NESAS. Penetration testing must be conducted Finally, one must schedule relatively frequent revisions of the operator specific threat model.

8 Summary and Concluding Remarks

In this paper we have conducted a first run of a threat analysis for the 3GPP 5G Service Based Architecture and the N32 roaming interface. As is natural for a preliminary analysis, we have taken a breadth-first approach. The findings are not unexpected, but needs to be heeded if the security level is to be satisfactorily.

8.1 Concluding Remarks

To paraphrase Nicolas Seriot in his conclusion in [24]: "Why do we have to make these designs so brittle and fragile?". We need our designs to be a lot more robust and less error prone in order to as reliable as they need to be. This means adopting a strong stance on security-by-design and preparing for failure.

Funding Open access funding provided by University Of South-Eastern Norway.

References

1. Moore, T., Kosloff, T., Keller, J., Manes, G., & Shenoi, S. (2002). *Signaling system 7 (ss7) network security. In The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.* (Vol. 3, pp. III–III).https://doi.org/10.1109/MWSCAS.2002.1187082.
2. ETSI, TS 09.02. Mobile Application Part (MAP) Specification. TS 09.02, ETSI, France (1995).
3. 3GPP, TS 29.002. Mobile Application Part (MAP) specification. TS 29.002 15.5.0, 3GPP, France, 06 2019.
4. 3GPP, TS 29.060. General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface. TS 29.060 15.5.0, 3GPP, France, 06 2019.
5. 3GPP, TS 29.274. 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3. TS 29.274 16.0.0, 3GPP, France, 06 2019.
6. 3GPP, TS 29.281. General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U). TS 29.281 15.5.0, 3GPP, France, 12 2018.
7. 3GPP, TS 33.210. 3G security; Network Domain Security (NDS); IP network layer security. TS 33.210 16.2.0, 3GPP, France, 06 2019.
8. Fajardo, V. (ed). (2012). Diameter Base Protocol. RFC 6733, IETF, 10.
9. 3GPP, TS 29.272. Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol. TS 29.272 15.8.0, 3GPP, France, 06 2019.
10. GSMA. Inter-Service Provider IP Backbone Guidelines, Version 7.0, 23 January 2012. IR 34, 01 2012.
11. GSMA.(2018). Guidelines for IPX Provider networks (Previously Inter-Service Provider IP Backbone Guidelines) v14.0. IR 34, 08 2018.
12. 3GPP, TS 23.501. System architecture for the 5G System (5GS). TS 23.501 16.1.0, 3GPP, France, 06 2019.
13. 3GPP, TS 33.501. Security architecture and procedures for 5G System. TS 33.501 17.0.0, 3GPP, France, 12 2020.
14. Bonfim, M. S., Dias, K. L., & Fernandes, S. F. L. (2019). Integrated NFV/SDN architectures: A systematic literature review. *ACM Computing Surveys (CSUR), 51*(6), 114.
15. Fielding, R.T., & Taylor, R. N. (2000).*Architectural styles and the design of network-based software architectures* (Vol. 7. University of California, Irvine Doctoral dissertation).
16. ITU-T. Information technology-Abstract Syntax Notation One (ASN.1): Specification of basic notation. Recommendation X.680, ITU-T, 08 2015.
17. Dierks,T., & Rescorla, E.(2008). The Transport Layer Security (TLS) Protocol; Version 1.2. RFC 5246, IETF, 08.
18. Rescorla, E. (2018). The Transport Layer Security (TLS) Protocol; Version 1.3. RFC 8446, IETF, 08.
19. Sakimura, N., Jones, M., & Bradley. J. (2015). JSON Web Signature (JWS). RFC 7515, IETF, 05.
20. Hildebrand, J., & Jones, M. (2015). JSON Web Encryption (JWE). RFC 7516, IETF, 05.
21. Hardt, D., (ed). (2012). The OAuth 2.0 Authorization Framework. RFC 6749, IETF, 10.
22. Perrow, C. (1999). *Normal accidents*. New Jersey: Princeton University Press.
23. Abadi, M., & Needham, R. (1996). Prudent engineering practice for cryptographic protocols. *IEEE Transactions on Software Engineering, 22*(1), 6–15.
24. Nicolas, S. (2018). JSON Parsing Considered Harmful. Technical report, Toulouse Hacking Convention, 03 2018.
25. Justin, R., Antonio, S., & Ian, G. (2017). OAuth 2 in Action. Manning Publications Shelter Island.
26. Daniel, F., Ralf, K., & Guido, S. (2016). A comprehensive formal security analysis of OAuth 2.0. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1204–1215. ACM.

27. San-Tsai, S., & Konstantin, B. (2012). The devil is in the (implementation) details: an empirical analysis of OAuth SSO systems. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 378–390). ACM.
28. Ethan, S., Henry, C., Dave, T., Patrick, T., & Kevin, B. (2015). More Guidelines Than Rules: CSRF Vulnerabilities from Noncompliant OAuth 2.0 Implementations. In Magnus Almgren, Vincenzo Gulisano, and Federico Maggi, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. 239–260). Cham, Springer International Publishing. ISBN 978-3-319-20550-2.
29. Janus, A. (2012). Towards a common agile software development model (ASDM). *ACM SIGSOFT Software Engineering Notes, 37*(4), 1–8.
30. Zolotas, C., Diamantopoulos, T., Chatzidimitriou, K. C., & Symeonidis, A. L. (2017). From requirements to source code: A Model-Driven Engineering approach for RESTful web services. *Automated Software Engineering, 24*(4), 791–838.
31. Shahin, M., Babar, M. A., & Zhu, L. (2017). Continuous integration, delivery and deployment: A systematic review on approaches, tools, challenges and practices. *IEEE Access, 5,* 3909–3943.
32. Thomas, P. Comments on "JWT is a Bad Standard That Everyone Should Avoid". Lobste.rs: https://lobste.rs/s/r4lv76/jwt_is_bad_standard_everyone_should_avoid, 03 2017.
33. Arciszewski, S. No Way, JOSE! Javascript Object Signing and Encryption is a Bad Standard That Everyone Should Avoid. Paragon Initiative.https://paragonie.com/blog/2017/03, 03 2017.
34. Madden, N. Should you use jwt/jose? Neil Madden Blog: https://neilmadden.blog/2017/03/15/should-you-use-jwt-jose/, 03 2017.
35. Fraser, T. No way, jose! lessons for authors and implementers of open standards. The 2018 Pass the SALT conference.
36. Dennis, D., Juraj, S., Christian, M., Vladislav, M., & Jörg S. (2017). On the (in-) security of javascript object signing and encryption. In *Proceedings of the 1st Reversing and Offensive-oriented Trends Symposium*. ACM.
37. Tozny. Cryptography and abstractions: Why all the jose hate? Tozny.com: https://tozny.com/blog/jose-cryptography-and-abstraction/, 03 2017.
38. NIS Coordination Group. EU coordinated risk assessment of the cybersecurity of 5G networks. Report, European Commission and ENISA, Brussel, Belgium, 10 2019.
39. ENISA. ENISA Threat Landscape for 5G Networks. TRL 2019-5G, European Union Agency for Network and Information Security (ENISA), 11 2019.
40. OWASP Foundation. OWASP Top Ten; Top 10 Web Application Security Risks (2017). https://owasp.org/www-project-top-ten/#, 2017.
41. Acunetix. Web Application Vulnerability Report 2020. https://www.acunetix.com/resources/report/Acunetix_2020_Web_Application_Vulnerability_Report.pdf, 05 2020.
42. ptsecurity.com. Web application vulnerabilities and threats: statistics for 2019. https://www.ptsecurity.com/upload/corporate/ww-en/analytics/web-vulnerabilities-2020-eng.pdf, 05 2020.
43. MITRE Common Weakness Enumeration. 2020 CWE Top 25 Most Dangerous Software Weaknesses. https://cwe.mitre.org/top25/archive/2020/2020_cwe_top25.html, 05 2020.
44. Microsoft. Develop secure applications on Azure. https://docs.microsoft.com/en-us/azure/security/develop/secure-develop, 12 2019.
45. Dennis, V. & Geoffrey, S. (1997). A type-based approach to program security. In Michel Bidoit and Max Dauchet, editors, *TAPSOFT '97: Theory and Practice of Software Development*, pages 607–621, Berlin, Heidelberg, Springer Berlin Heidelberg. ISBN 978-3-540-68517-3.
46. Meyer, B. (1992). Applying'design by contract'. *Computer, 25*(10), 40–51.
47. 3GPP, TS 33.517. 5G Security Assurance Specification (SCAS) for the Security Edge Protection Proxy (SEPP) network product class. TS 33.517 16.1.0, 3GPP, France, 12 2019.

Geir M. Køien received his PhD from Aalborg University in 2008. Before that he had worked for many years in industry, including LM Ericsson Norway and Telenor R&D. During these years he worked extensively with mobile systems and with security and privacy. He has also worked with the Norwegian Defence Research Establishment (FFI) and with Norwegian Communications Authority (NKom) on various security and communications related projects. Currently, he is a professor with the University of South-Eastern Norway (USN).