
RESTful APIs for the 5G Service Based Architecture

Georg Mayer

Chairman of 3GPP TSG-CT, Huawei, Vienna, Austria
E-mail: georg.mayer.huawei@gmx.com

Received 31 March 2018;
Accepted 3 May 2018

Abstract

5G sets out to be the global connectivity and integration platform for a broad variety of industries in the upcoming decade. In order to do so it not only needs to fulfil the requirements of these industries but must also ensure its tight integration into the digital infrastructure of the 2020s by embracing key technologies. This article shows how one of these key technologies, the RESTful design of Application Programming Interfaces (APIs), is used in the 5G Service Based Architecture (SBA). The basic principles of modern API development are explained and it is shown how those integrate into the specific needs of the 5G Core Network.

Keywords: 5G, REST, application programming interfaces, 3GPP, service based architecture, SBA, northbound APIs, NAPS, HTTP.

1 Introduction

5G aims to become the global connectivity enabler and service platform for a broad variety of industries in the upcoming decade and beyond. To achieve this goal the 5G system not only opens up towards a whole new group of customers, the so-called verticals, it also embraces the technologies which are and will be used by these verticals and more commonly in the digital landscape of the 2020s.

Journal of ICT, Vol. 6-1&2, 101–116. River Publishers

doi: 10.13052/jicts2245-800X.617

This is an Open Access publication. © 2018 the Author(s). All rights reserved.

Technologies and concepts such as virtualization, cloud computing, internet of things, functionality exposure and self-organizing networks are central building blocks of 5G and will allow seamless communication as well as enable synergies between different industries.

For the purpose of exposure of functionality to 3rd parties as well as other types of system internal communication 3GPP chose to make use of the widely established REST architecture design paradigm, which describes the design of distributed applications and more specifically of Application Programming Interfaces (APIs). This article explains why the REST paradigm was chosen for certain aspects of the 5G system and how the different REST principles are applied in the 5G Service Based Architecture.

The so-called RESTful APIs can be understood as an example of 3GPP's commitment to tightly integrate 5G in the current and upcoming digital ecosystem of their customer.

2 New Communication for the Mobile Core

Capability exposure, i.e. making 5G Core Network functionalities available to 3rd parties such as service providers and vertical industries outside the operator's domain, is provided by the Network Exposure Function (NEF). The interface provided by the NEF to 3rd parties can be regarded as one of the essential membranes through which 5G communicates more closely towards vertical industries than mobile networks of earlier generations did. It was therefore a key requirement that 3GPP defines this interface in way that it would fully align with widely accepted and future proof principles for the design of such exposure interfaces.

Exposure of functionality is a common concept used by modern software design, especially for web-services which are offered over the internet. The use of APIs for this purpose is practically without competition and is applied from simple temperature sensors in automated home environments to large-scale cloud providers enabling near real-time content access. All these different APIs are defined along a number of common principles which are referred to as REST architectural style, which will be described in the following sections.

3GPP decided that 5G service exposure by the NEF should be based on RESTful APIs, as shown in Figure 1.

Nevertheless, the APIs offered to 3rd parties, also known as "northbound APIs" are only applicable to a single interface of the 5G system, whilst the NEF is one of many Network Functions within the completely redesigned 5G Core

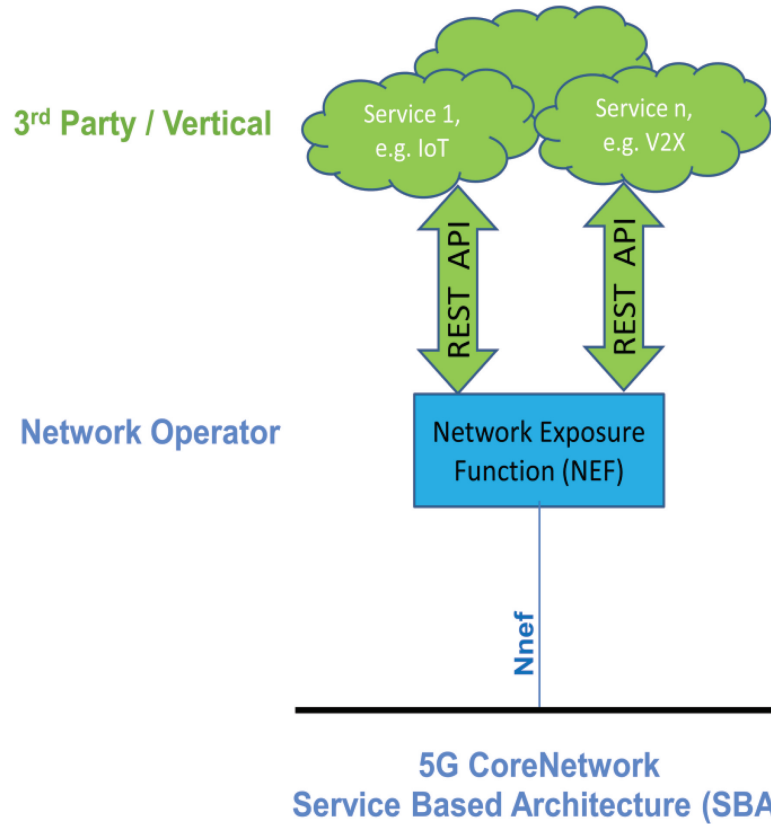


Figure 1 NEF (as part of the 5G SBA) providing services to 3rd parties via RESTful APIs.

Network. This redesigned core is the architectural and technical realization of the service-based change design of the 5G system and therefore was named Service Based Architecture (SBA). The Network Functions (NFs) forming the SBA communicate with each other via Service Based Interfaces (SBI), as shown in Figure 2.

3GPP took the forward-looking decision to use RESTful APIs not only for 3rd party functionality exposure but also for via the SBIs. Therefore the 5G Core Network internal communication obeys the same principles as the functional exposure, thus allowing a harmonized and holistic technological approach of the complete 5G system, fully in-line with the progressive paradigms which are at the heart of a wide range of services used by end-customers as well as for the automation of whole industries.

In chapter 6 of his dissertation Fielding describes in detail how the principles of REST can be used within the World Wide Web, i.e. by making use of Uniform Resource Indicators (URIs), the Hypertext Transfer Protocol (HTTP), different data description languages and how such technologies can be used in a “RESTful” way for real world deployments.

In the 18 years since the publication of Fielding’s dissertation, the REST paradigm has fundamentally re-shaped the way how software applications are designed, implemented and deployed. It is used throughout the IT industry, there exist countless tools as well as books, articles and web pages to support its use and a huge developer community is experienced with REST principles. The paradigm itself as well as the related technologies, protocols and tools are further developed not only by software companies and universities, but also by the open source community and by global standards organizations such as the W3C (World Wide Web Consortium) and IETF (Internet Engineering Task Force) [1].

REST has proven to be a reliable and future proof way for developing distributed applications. It’s therefore safe to say that there is thriving ecosystem which is built on the REST principles.

4 Example RESTful SBA Procedures

This section describes an example scenario consisting of three API calls within the 5G SBA which are meant to exemplify how RESTful principles are used by 3GPP. The given examples are not complete and were only chosen to give the reader an initial overview. Further information can be found in the given references. Section 5 will explain how the REST principles are applied in 5G SBA, based on the examples of this section.

4.1 Example Scenario

The functional split chosen for 5G SBA Network Functions includes e.g. an Access and Mobility Management Function (AMF), which serves as the single-entry point for a user equipment (UE) for all its communication. Once the user decides to use one of the services, e.g. to browse the web, the AMF needs to assign a Session Management Function (SMF) which manages the users session context. As in 5G virtual network functions (VNF) can be instantiated and deleted at any time, the AMF first needs to discover an available and suitable SMF, which is achieved via the Service Discovery procedure performed between the AMF and the Network Repository Function

(NRF). To allow for successful discovery, the SMF must have registered beforehand with the NRF.

Thus we look at a dedicated example, consisting of three different procedures, which are depicted in Figure 3.

1. *Service Registration*: the SMF registers the services it provides with the NRF.
2. *Service Discovery*: the AMF queries the NRF for a suitable SMF and in return receives the address of the SMF which registered in step 1.
3. *Session Establishment*: the requested session is established on the control plane level by the AMF via the SMF.

3GPP TS 23.501 [2] defines the roles of service consumer and service producer. The service consumer is the NF which requests the related service, whilst the NF which exposes the requested service is the service producer. Services exposed by a NF are further structured into service operations, as defined in 3GPP TS 23.501 [2] and 3GPP TS 29.501 [5].

The basic procedures as outline here are defined in 3GPP TS 23.502 [3], whilst the more detailed flows and protocol elements which are exchanged are specified in the related NF specifications, i.e. for the NRF in 3GPP TS 29.510 [7] and for the SMF in 3GPP TS 29.502 [6].

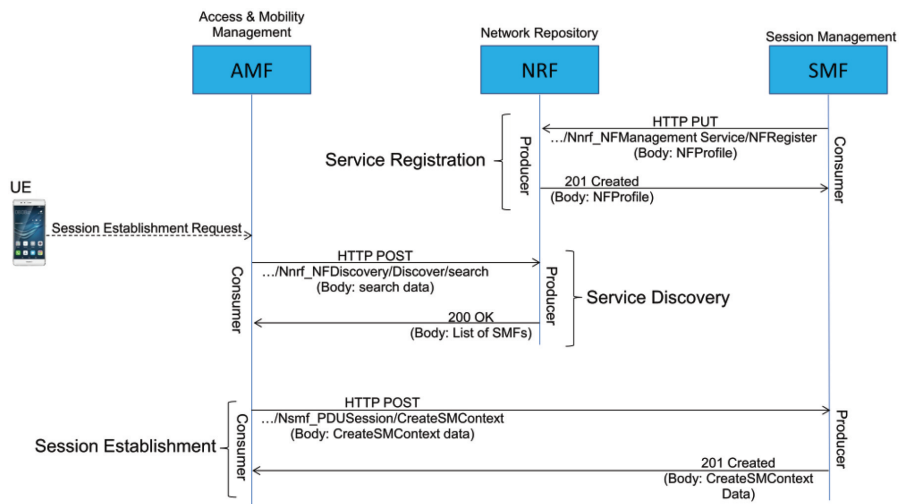


Figure 3 Simplified API calls for example SBA procedures.

4.2 Service Registration

During Service Registration the SMF takes the role of the service consumer and the NRF takes the role of the service producer, exposing the Nnrf_NFManagement service with the NFRegister service operation (see 3GPP TS 29.510 [7] and TS 23.501 [2]).

The SMF sends a HTTP PUT request towards the NRF. The Uniform Resource Locator (URI) addresses the profile information which is intended to be created at the NRF, e.g. “`https://nrf7.slice-v2x.opx.3gpp/nrf-nfm/v1/nf-instances/smf5-slicev2x`”. Note that the URI directly addresses the Service Profile of the SMF as it will be stored at the NRF, i.e. the URI is not just the address of the NRF but of the resource which it is requested to create and host.

Within the Service Registration procedure the SMF sends the NFProfile information, which is encoded as a structured data type in a JSON (short for JavaScript Object Notation, as defined by RFC 8259 [10] and ECMA-404 [11]) document in the body of a HTTP PUT request. This NFProfile document includes information about the SMF, such as

- the Type of NF (in this case the SMF);
- the network identification (PLMN ID) and slice identifiers (S-NASSI, NSI ID) to which the SMF belongs to;
- the address or addresses (IP Address or FQDN) of the NF (e.g. “`https://smf5.slice-v2x.opx.3gpp`”);
- a list of service names which the NF supports (e.g. “Nsmf_PDUSession” and “Nsmf_EventExposure” in the case of SMF);
- for each service name a list of the supported service operations (e.g. for the Nsmf_PDUSession the service operations “Create SM Context”, “Update SM Context”, “Release SM Context”, “Notify SM Context Status”, “Query SM Context”, etc).

Once the NRF receives the HTTP PUT request it authenticates and verifies it. Assuming the received HTTP PUT request passes all tests the NRF then uses the URI, i.e. the address to which the HTTP PUT request is sent to, to create a new local resource, including the SMF NFProfile.

It then acknowledges the creation of the resource by returning a HTTP 201 Created response, in which the NFProfile is included again in the body.

Due to the Service Registration procedure a new resource with the unique address “`https://nrf7.slice-v2x.opx.3gpp/nrf-nfm/v1/nf-instances/smf5-slicev2x`” is created at the NRF. This resource includes the above described structured information about the SMF.

4.3 Service Discovery

During Service Discovery the AMF takes the role of the service consumer and the NRF takes the role of the service producer, exposing the Nnrf_NFDDiscovery service with the NFDDiscover service operation (see 3GPP TS 29.510 [7]).

The AMF sends a HTTP POST request to the NRF, querying for a list of SMFs which registered with a specific set of supported services. The list of services which the SMF's need to support are included as a structured data type (JSON document) in the body of the HTTP POST request.

The NRF authenticates and validates the incoming HTTP POST request. Once it has passed all checks the NRF searches its local resources, i.e. the registered and stored profiles of NFs, for matches against the query list received from the AMF.

The list of matches is sent from the NRF to the AMF in the body of the 200 OK response to the HTTP POST request.

4.4 Session Establishment (AMF Creates Session Context at SMF)

Once the AMF receives the response it chooses “https://smf5.slice-v2x.opx.3gpp” as the SMF for the intended session establishment.

During Session Establishment the AMF takes the role of the service consumer and the SMF takes the role of the service producer, exposing the Nsmf_PDUSession service with the Create SM Context service operation (see 3GPP TS 29.502 [6]).

The AMF sends a HTTP POST request to the SMF, addressing the exposed service of Session Establishment which was found in the list of supported services, i.e. “https://smf5.slice-v2x.opx.3gpp/nsmf-pdusession/v1/sm-contexts”. This indicates that the AMF wants to create a Session Management context at the SMF. The details of the context under creation are included as the SmContextCreateData structured data type in a JSON document within the body of the HTTP POST request. The SmContextCreateData information includes all information necessary to create a Session Management Context, e.g. the IDs of the requesting UE, slice identifiers and so forth.

The SMF authenticates and validates the incoming HTTP POST request. Once it has passed all checks the SMF, if it has the necessary resources available, creates the requested Session Management Context.

All information related to the created Session Management Context is then returned by the SMF to the AMF in the body of the 201 Created response to the HTTP POST request.

Note that especially for the case of Session Establishment this article only treats the very first SBA interaction between the AMF and the SMF. It leaves out all subsequent interactions e.g. with the Unified Data Management (UDM), the Policy Control Function (PCF) and any other NF. It also does not describe interactions between the SMF and a User Plane Function (UPF). More details can be found in 3GPP TS 23.502 [3] Section 4.3.2.2.

5 REST Principles and Technologies Adopted by 5G SBA

At the core of RESTful service architecture design and service development as described by Fielding [1] lay six principles which all aim to make the creation and deployment of distributed services flexible, coherent and scalable.

This section first explains a number of essential terminologies used to describe REST principles, then details all six principles and evaluates how they are implemented in 5G SBA. Finally this section gives a short overview how HTTP is used within RESTful deployments and 5G SBA specifically.

5.1 Resource, Serialization and Representation

In the context of REST, a **resource**, e.g. the SMF Service Profile as used during Service Registration, can be stored in any form on a server (here: NRF) and can also take on different internal states. It only exists under a unique Uniform Resource Location (URI), which is used to address it for different purposes – e.g. creation, modification and deletion.

During these procedures it is often necessary to exchange either all or part of the information related to the resource between a client (here: SMF) and a server (here: NRF). To allow this exchange the resource gets **serialized**, i.e. all information related to it is written into a document. In the case of 5G SBA objects are serialized as JSON documents. These documents are individual **representations** of the resource, i.e. they are not the resource themselves, which still is only available via the URI, but only a snapshot of it.

5.2 Client/Server Principle

Client/Server is the REST principle which demands that a dedicated split of responsibilities between the entity which requests a resource, the client, and the entity which provides access to the resource, the server.

In the 5G SBA example Service Registration procedure outlined in the previous section the SMF acts as a client towards the NRF during Service

Registration, whilst the NRF acts as a server. 3GPP alternatively uses the terms service consumer for the client/SMF and service producer for the server/NRF.

Also the Service Discovery example shows the clear split between client (AMF, service consumer) and server (NRF, service producer).

5.3 Stateless Principle

Stateless is the REST principle which demands that the server does not keep any state related information concerning the communication with specific clients. This in turn means that the requests from the client need to include all necessary information which enables the server to perform the desired service. This not only frees resources at the server, but also allows for load balancing and resilience in a distributed environment, as requested services can be made available by a set of servers, to which load can be distributed on a per-request basis.

In the example scenarios both the NRF as well as the SMF, when acting as service producers, are simply returning the requested information to the respective clients, e.g. the AMF. Neither NRF nor SMF keep any additional state about the communication with the AMF.

Going further, in order to achieve full statelessness it is also necessary that all information that is assumed to be “local” on the server is indeed kept within a storage unit (database server), which is accessible from all servers offering a particular service. In the example scenarios the service profiles are still stored locally at the NRF after the example Service Registration procedure. This means that access to this Service Profile can only be provided via a single NRF, i.e. “nrf7.slice-v2x.opx.3gpp”, which acted upon the Service Registration request from the SMF. If this specific NRF would e.g. be shut down, the stored Service Profile would not be available any longer within the network.

In order to create a truly distributed system 3GPP therefore created a data storage architecture (see e.g. 3GPP TS 23.501 [2]), where structured data and unstructured data can be stored at central repositories. The Unstructured Data Storage Function (UDSF) is part of this data storage architecture and offers services for data storage, manipulation and retrieval to every NF within the 5G SBA.

In the current 3GPP specifications it is specifically foreseen that the UDSF is used for achieving a fully stateless AMF (see 3GPP TS 23.501 [2] section 5.21.2 and 3GPP TS 29.500 [4] section 6.5.2). Nevertheless the UDSF can be used by any other NF to store local data within the centralized repository,

so that services related to specific users or session can be handled not only by a single instance of a specific NF type, but also multiple instances, thus achieving higher reliability and load-distribution by statelessness.

5.4 Cacheable Principle

Cacheable is the REST principle which demands that clients get indication from servers whether the received information can locally be cached at the client, i.e. it can be re-used by the client at a later time. This of course is only possible if the information is not supposed to change. Therefore responses which contain serialized representations of resources must indicate whether the contained information can be cached by the client or not. 5G SBA follows this principle.

5.5 Layered System Principle

Layered system is the REST principle which demands that the client is unaware to which specific server it is connected to and if parts of the communication (e.g. authentication) are performed by different servers than other types of requests (e.g. mobility management).

Such a decoupling of services and servers which offer them is implemented into the 5G SBA e.g. via the NRF. As shown in the example procedures, the Service Registration and Service Discovery procedures enable a server, e.g. the AMF, to search for specific services provided within the system, without looking for dedicated servers.

More generally, the function split chosen by 3GPP for the 5G SBA was guided by the idea to create a truly layered system. This is visible in the different tasks assigned to the NFs. Whilst, for example, the AMF offers services access and mobility management, the SMF offers those for session management, the Network Slice Selection Function (NSSF) for network slice selection and so forth. All of these functionalities are required to make features like mobility, roaming or security seamlessly available throughout the global system to its users.

5.6 Code on Demand Principle

Code on demand is the REST principle which allows pieces of code to be downloaded from a server to a client to allow for much more dynamic and flexible service operation. This principle is currently not implemented by 5G SBA, as at currently the 5G SBA services do not require this flexibility.

5.7 Uniform Interface Principle

The uniform interface is a REST principle which encompasses four essential building blocks for the creation of distributed services to allow flexible and dynamic applications.

Resource identification in requests means that the resource or service needs to have a unique address, usually an URI. It is important to note that it is not a specific server which is addressed by the URI but the resource. As outlined above it is possible in distributed systems that more than one server can offer the same resource, in such a case the resource would have the same URI, regardless on which server it would be offered during a specific API call.

Within the example procedures we used the URI “https://nrf7.slice-v2x.opx.3gpp/nrf-nfm/v1/nf-instances/smf5-slicev2x” to address the Service Registration of the SMF at the NRF. This URI is constructed based on strict rules which are defined in 3GPP TS 29.510 [7] for NRF addresses. 3GPP defined related URI construction rules for every NF within the 5G SBA.

The URI includes the following elements:

- scheme (“https:”), i.e. the protocol used for message transport;
- the API route (“nrf7.slice-v2x.opx.3gpp”), i.e. the address of the NF to which this request is sent to;
- the api-name (“nrf-nfm”), i.e. the specific service which is triggered, in this case;
- the api-version (“v1”), as specified in the related 3GPP specifications;
- NF specific resource information (“nf-instances”), indicating that the conveyed information relates to a specific NF instance; and
- the instance ID of the NF which registers its profile (“smf5-slicev2x”).

Note that in the given examples fixed server addresses (e.g. “nrf7.slice-v2x.opx.3gpp”) were chosen, which is not in general necessary. 5G SBA allows that more generic API roots (e.g. “nrf.opx.3gpp”) can be used, as long as different NF instances all offer the same resources.

Resource manipulation through representation requires that a resource on a server can be modified or replaced by a client by sending a related request which includes a representation of either the whole or part of the resource.

5G SBA fully supports this principle. For example an already stored Service Profile of an SMF at an NRF can be modified with e.g. a HTTP PATCH request from that SMF. HTTP PATCH request needs to include a

JSON document which represents either the complete Service Profile or at least those parts of it, which are intended to be modified.

Self-descriptive messages are important for the client/server and statelessness concepts of REST. It demands that requests and response sent needs to include enough information indicating how the message should be processed. This was already outlined in section 5.1.3 above.

Hypermedia as the engine of application state (HATEOAS) allows the client to learn by means of hyperlinks included in the response which further actions can be taken towards the resource in the given situation. The server provides the hyperlinks in the response based on the situation a specific resource is in.

This principle is intended to be followed within 5G SBA, but so far has not been applied to any dedicated service.

A simple example for future use is the list communication means by which a user is reachable. A user registered from a UE to the 5G system is e.g. available for calls and for receiving short messages. If an NF needs to get aware of the available communication means for the specific user it will query e.g. the AMF. The AMF, acting as service producer or server, then returns the options (hyperlinks) within the body of the response to the querying NF. In the described case these would be one hyperlink pointing to the service for establishing a call and a second hyperlink pointing to the service for sending short messages to the user.

If the user would not be registered to receive voice calls, then the AMF should, in the body of the response, only return a single hyperlink, pointing to the service for sending short messages to the user.

5.8 HTTP

RESTful principles were right from the beginning meant to evolve the way services are developed and deployed in the world wide web [1]. At that time the Hypertext Transfer Protocol (HTTP) was available as version 1.1 (as defined in IETF RFC 2068 [8]) and had a number of shortcomings which complicated the deployment of distributed services. HTTP/1.1 was designed for straightforward web-browsing, i.e. in its basic operation a browser (client) would send a HTTP request to a server, identifying a web page in the URI of the request. All going well the server then would send back the related web page in the 200 OK response to the HTTP request.

Meanwhile HTTP/2 is available (defined in RFC 7540 [9]), which solves a number of problems and especially allows for higher performance than its

predecessor due to new features such as e.g. multiplexing and binary framing. The Internet Engineering Task Force (IETF) developed HTTP/2 to allow full compatibility with HTTP/1.1 but also to allow for better handling of (RESTful) API operations and general web-service development.

The HTTP methods or request types (sometimes also called “verbs”) trigger different types of functionalities within a RESTful environment, e.g.

- the HTTP POST request creates a new resource which can be addressed by the URI to which the request is sent;
- the HTTP GET request lists/retrieves the resource(s) addressed by the URI;
- the HTTP PUT request replaces the resource addressed by the URI with the representation contained in the request;
- the HTTP PATCH request updates a resources addressed with the representation contained in the request;
- the HTTP DELETE request deletes the resource addressed by the URI.

5G SBA makes use of all these different HTTP request types in the described manner. The example procedures showed how POST and PUT requests are used in 5G SBA.

6 Conclusion and Outlook

3GPP decided to not only base its northbound APIs on the principles of RESTful design, but also the 5G Core Network internal communication, i.e. to make use of RESTful APIs over all Service Based Interfaces. Already in the first 5G release, i.e. 3GPP Release 15, the principles of RESTful design have been followed strictly and thus guarantee a high level of flexibility in future service creation and enhancement.

Many details concerning the usage of RESTful APIs in 5G SBA could not be handled in this article in detail, e.g. the Richardson Maturity Levels, subscriptions to NF services, interactions with the transport layer protocol and also the openness of 5G SBA towards Remote Procedure Calls (RPC). All this would require a level of detail which cannot be covered by an article of this limited length.

Another aspect which was not handled here is the strong cooperation between 3GPP and IETF for ensuring 5G APIs to be aligned with the latest internet protocol and RESTful design developments. This coordination also allows 3GPP to push for IETF standardized solutions of the specific 5G requirements in these areas. There are many actors in both 3GPP and IETF

which work to enable the alignment and integration of 5G and internet technologies.

3GPP currently (May 2018) is in the process of completing the first 5G release. This is also the first 3GPP release which saw the serious implementation of RESTful design principles. By December 2019 the second 5G release will be completed and currently the discussions are ongoing on how to further evolve the service concept of the SBA and which additional functionalities could make use of RESTful design. But already with the decisions taken so far it is clear that RESTful APIs play a central role for the technical realization and integration of 5G.

References

- [1] R. Fielding, ‘Architectural Styles and the Design of Network-based Software Architectures’, Dissertation, University of California, Irvine, https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf, 2000.
- [2] 3GPP TS 23.501, ‘System Architecture for the 5G System; Stage 2 (Rel-15)’, March 2015.
- [3] 3GPP TS 23.502, ‘Procedures for the 5G System; stage 2 (Rel-15)’, March 2015.
- [4] 3GPP TS 29.500, ‘5G System; Technical Realization of Service Based Architecture; Stage 3’, March 2015.
- [5] 3GPP TS 29.501, ‘5G System; Principles and Guidelines for Services Definition; Stage 3’, March 2015.
- [6] 3GPP TS 29.502, ‘Session Management Services; Stage 3’, March 2015.
- [7] 3GPP TS 29.510, ‘Network Function Repository Services; Stage 3’, March 2015.
- [8] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, T. Berners-Lee, ‘RFC 2068 – Hypertext Transfer Protocol – HTTP/1.1’, January 1997.
- [9] M. Belshe, R. Peon, M. Thomson, ‘RFC 7540 - Hypertext Transfer Protocol Version 2 (HTTP/2)’, May 2015.
- [10] T. Bray, ‘RFC 8259 - The JavaScript Object Notation (JSON) Data Interchange Format’, December 2017.
- [11] ECMA-404, ‘The JSON Data Interchange Syntax’, December 2017.

Biography



Georg Mayer is the Chairman of the Core Network and Terminals, Technical Specification Group of 3GPP (TSG CT). His current focus is on the coordination of 5G related work inside and outside 3GPP. He participates in the IETF and works closely with several of the new stakeholders in 5G, such as public safety, railways, autonomous systems and IoT service providers.

He is a published author and has co-authored a book on IMS, amongst other titles.

Georg Mayer holds an MSc in Computer Science from the University of Hagen, Germany and he works for Huawei Technologies.