

Context-Aware Service Composition and Delivery in NGSONs over SDN

Federica Paganelli, Mehmet Ulema, and Barbara Martini

ABSTRACT

NGSON is an IEEE sponsored effort to standardize an open service ecosystem where different stakeholders may cooperate to satisfy users' ever changing requirements by composing and delivering services dynamically across different service providers and network operators' domains. A main distinguishing feature of NGSON is context awareness, that is, the capability of collecting, processing, and using context information to drive the interactions of the NGSON overlay nodes with service and network operators' infrastructures for guaranteeing an optimal user experience. Recent advances in network control and management technologies, such as SDN, NFV, and service-oriented networking paradigms, are expected to make the network effectively able to cope with application service requirements in a more flexible, timely, and agile manner. In this article we discuss how NGSON, which already incorporates several aspects of virtualization, could take advantage of network virtualization mechanisms and programmable traffic steering capabilities to enhance and effectively support context-aware service composition along with adaptive service delivery. We propose an architectural model for NGSON that leverages these novel network service control capabilities and illustrate through a use case how this model could effectively be put into operation.

INTRODUCTION

Today's consumers are constantly looking for better, faster, and immersive networked experiences. This end users' demand for interactive, context-aware, and content-oriented services is imposing increasingly stringent requirements on the delivery capabilities of underlying networks. In a context where relevant investments on infrastructures are needed to support the increasing bandwidth requirement, revenues are decreasing or, in the best cases, increasing at a slower rate. This means that network operators and service providers need to strengthen their collaboration to expand their service offerings, while optimizing revenues and limiting infra-

structure investments. This cooperative vision leads to the concept of an open service ecosystem where new valued-added applications may be provisioned by composing heterogeneous services dynamically across different providers' and network operators' domains.

The Next Generation Service Overlay Network (NGSON) is an IEEE sponsored effort to standardize such an ecosystem for the benefit of network operators, service and content providers, and end users [1–3].

NGSON has been standardized by IEEE [1] as a reference framework for collaborative and customer-centric service delivery. By logically separating service-related from transport-related functions, NGSON organizes the application services offered by various networks on an overlay that allows service providers to offer highly integrated and collaborative services and provide the optimal user experience. Thanks to these capabilities, the NGSON overlay may act as an intermediate layer that supports the provisioning of value-added context-aware application services by leveraging the delivery capabilities provided by heterogeneous underlying networks.

However, the NGSON intermediation role is not trivial due to the divergent interests of over-the-top (OTT) service providers and network operators in terms of business and service provisioning models. Indeed, while OTT providers can benefit from rapid innovation cycles, network operators have many difficulties in rolling out innovative network-based services and benefitting from increasing revenues of the new digital economy. A relevant factor here is the inflexibility of the network, which is static and costly to change. Moreover, the deployment of middleboxes exacerbates the rigidity of service deployment processes by imposing static chains of network appliances without considering specific service delivery requirements [4]. Middleboxes, although essential, are expensive, complex, and difficult to manage, and are prone to failures due to misconfiguration and overload. These facts prevent network operators from offering differentiated end-to-end services elastically, and thus from effectively monetizing innovative approaches to service delivery [5].

Recent efforts in research on promising net-

F. Paganelli and B. Martini are with the National Interuniversity Consortium for Telecommunications.

M. Ulema is with Manhattan College.

NGSON CONTEXT-AWARE SERVICE OVERLAY

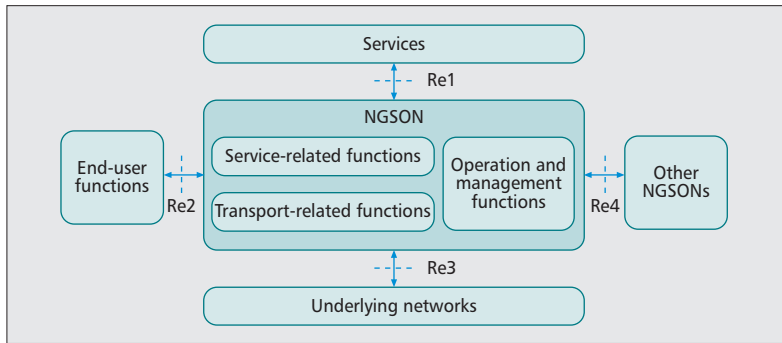


Figure 1. NGSON logical architecture.

work technologies, that is, software defined networking (SDN) [6], network function virtualization (NFV) [7], and service-oriented networking paradigms [8], are going in the direction of developing network virtualization and abstraction models as well as programmable service delivery, thus addressing the aforementioned flexibility, agility, and timeliness requirements. In fact, this technological shift has some similarities with recent paradigms in the computing domain (service oriented architecture, cloud computing, etc.), and is thus expected to promote a dynamic landscape of network services that can be discovered, negotiated, and composed to meet applications' requirements [9].

We argue that the NGSON overlay, as an intermediate layer between application and network, could take advantage of these emerging networking paradigms to enhance and effectively support context-aware service composition along with adaptive service delivery. This is a hot topic in current NGSON standardization activities. As a matter of fact, liaisons have been purposely set up with the International Telecommunication Union — Telecommunication Standards Sector (ITU-T) Joint Coordination Activity (JCA) on SDN, the European Telecommunications Standards Institute (ETSI) Industry Specification Group for NFV, and the TeleManagement Forum (TMF) Digital Services Reference Architecture (DSRA) project.

In this article we discuss how NGSON specifications, which already incorporate several aspects of virtualization, could evolve to fully take advantage of network virtualization mechanisms (i.e., NFV) and programmable and granular traffic steering capabilities (i.e., SDN) for negotiating proper delivery paths and network services on behalf of application services. First, we provide an overview of the NGSON context-aware service overlay. Second, we briefly introduce the main concepts of the SDN and NFV solutions underpinning the realization of an adaptive network-aware service composition model in NGSON. Finally, we propose an extension to the architectural model for NGSON that leverages the envisioned service-oriented abstractions and programming capabilities at the network control layer to optimize the delivery process. Then we illustrate through a use case how this extended NGSON composition model could operate to accomplish comprehensive context-aware service composition and delivery.

The NGSON architecture is composed of an overlay set of functionalities between services operated by service providers and underlying network infrastructures to provide collaborative and customer-centric services to users.

As shown in Fig. 1, the NGSON architecture includes service-related, transport-related, and operation and management functions [1]. Service-related functions include service control capabilities, such as service discovery, service registry, service composition, and service routing. Transport-related functions support service delivery, thus including context-aware network routing, traffic optimization capabilities, as well as interworking with different underlying networks. Operation and management functions provide overlay management and service assurance capabilities.

The NGSON architecture also specifies four reference points, Re1, Re2, Re3, and Re4, for interacting with external systems. Specifically, Re1 allows providers to publish their services to NGSON, and Re2 enables end-user functions to access NGSON service and content delivery capabilities. NGSON should support existing protocols (e.g., SIP, Diameter, and ALTO) to interwork with underlying heterogeneous networks (Re3). A federation of NGSONs can be established through Re4.

Basic and composite services can be searched for and delivered over different networks to optimize the users' quality of experience (QoE) as well as network and computing resource consumption. Cloud technologies for virtualization of network and computing resources are also expected to enforce this vision.

Context awareness is a feature that distinguishes NGSON from other related standardization activities in the areas of service delivery platforms and service overlay networks [3]. Indeed, adaptation to context is a strategy for driving service composition to maximize user satisfaction and service continuity while optimizing the use of services and network resources. Moreover, implementation of context-aware policies in NGSONs is empowered by the fact that the NGSON capability of interworking with different network operators and service providers allows acquiring context information from plenty of sources.

In NGSON, four classes of context have been defined: *service, user, network, and device context*. Such heterogeneous context items may be gathered by leveraging ad hoc or standard protocols. In Table 1 we mention some relevant examples for each context class.

Context-aware adaptation may be enforced at different levels, as shown in Fig. 2. At the application and service level, context parameters may be used for driving the dynamic discovery, composition and selection of services. For instance, this may include discovery mechanisms for retrieving those service profiles that match with context information (e.g., a user's preferences and current activity), the runtime selection and invocation of the service endpoints that are closer to a user's location (context-aware routing),

Context classes	Definition	Context source examples
Service context	Information that characterizes a service, such as service QoS (i.e., execution cost, response time, availability, reputation), service roaming state, service network endpoint, etc.	Service context information, such as target quality parameters and dynamic monitored values, may be acquired by using the WS-Agreement language and protocol.
User context	Information that characterizes the situation of a user, such as location, presence, current activity, social relations, and preferences.	User presence information can be modeled and exchanged according to two IETF specifications: SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE), and the XML-based XMPP messaging and presence specifications. APIs provided by public and private social networks (e.g., OpenSocial, OpenGraph) may also offer valuable information on users' social relationships and current activities.
Network context	Information characterizing underlying networks, such as topology, bandwidth, and traffic performance (e.g. delay, packet loss rate).	Network information can be acquired through the Application-Layer Traffic Optimization (ALTO) protocol, which is currently under specification by the IETF. ALTO is aimed at providing network information (e.g., basic network location structure and preferences of network paths) to applications with the goal of improving application performance by properly adapting network resources consumption.
Device context	Device hardware and software configuration and dynamic information (e.g., battery power, memory consumption, received signal strength of available access networks).	Device capabilities may be represented according to the Composite Capability/Preference Profiles W3C recommendation, which provides an RDF-based model and core vocabulary for describing device capabilities and user preferences.

Table 1. Context classes in NGSON and examples of context sources.

the dynamic substitution of a basic service with another one due to faults or requirement changes (adaptive service composition), as well as the selection of content caches or sources according to user location and guaranteed quality of service (QoS) (context-aware content routing).

At the network level, NGSON tries to optimize data delivery between an end user and a service endpoint or among service endpoints by negotiating QoS requirements in the underlying networks while enforcing proper service data forwarding across the network according to the current context (e.g., a user's current activity, location, and device capabilities) and required QoS.

BRIDGING THE GAP BETWEEN APPLICATIONS AND NETWORKS

Context-aware, dynamically adaptive, and self-organizing networking capabilities make NGSON the ideal candidate for the provisioning of highly integrated services addressing network-aware adaptations while satisfying the ever changing requirements of users. In this regard, recently emerging paradigms such as SDN [6] and initiatives such as NFV [7] promise to profoundly innovate network service provisioning models while effectively addressing dynamic QoS adaptation within NGSON [10].

SDN is about decoupling the software-based control plane (e.g., routing functions) from the hardware-based data plane (i.e., packet forwarding engine) while abstracting the underlying network infrastructure and moving the network intelligence to a centralized network controller. Such separation allows for more agile and cost-effective network operation thanks to full programmability of data forwarding, enhanced decision making capabilities based on a global

abstract view of the network status, and, finally, increased portability of the controller and future-proofing of investments ensured by the abstraction process. NFV is one of the most innovative manifestations of virtualization in networking, enabling network appliances, such as firewalls, deep packet inspection (DPI), and intrusion detection systems (IDS), to be deployed in software as virtualized components (i.e., virtual machines) provisioned in general-purpose hardware systems such as commodity servers.

Currently, such network appliances are provisioned in costly and specialized hardware systems (i.e., middleboxes), which are cabled in the required sequence and individually configured through time-consuming and error-prone procedures due to proprietary and complex command syntax. Such a practice leads to inefficient network service deployments. First, since service data transfer generally increases over time and depends on the number of users, building a service chain with such a burdensome effort requires inevitably resorting to overprovisioning to support occasional maximum levels of demand. Second, the sequence of network functions is decided at the initial network design time and statically provisioned as stationary network service chains without considering specific packet flow requirements. The result is today's widespread practice of directing excess traffic through unnecessary middleboxes or setting up cumbersome tunnels to avoid them. Instead, dynamically establishing chains of network appliances, which depend on QoS requirements of service data flows (e.g., delay constraints) as well as specific context information (e.g., available processing capabilities), would be beneficial [4]. For instance, it could be beneficial in the case of a long-lived multimedia data flow to avoid a deep packet inspection appliance in order to confine delays and save processing resources.

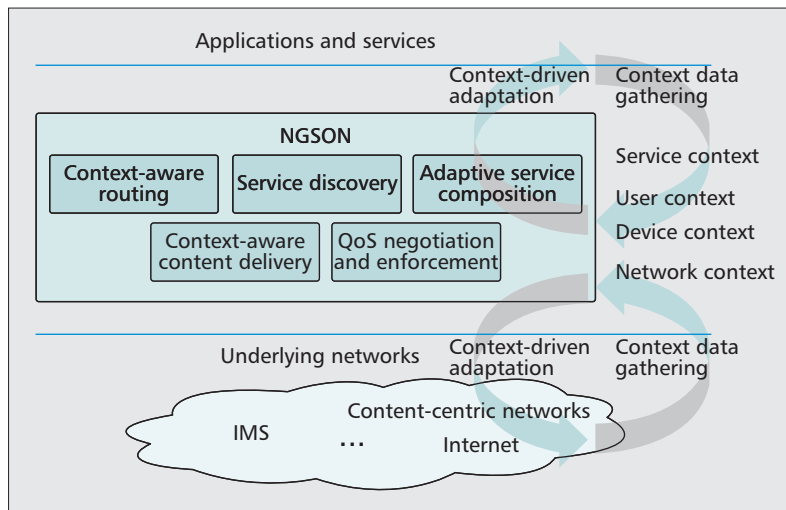


Figure 2. Context awareness in NGSON.

Definitely, the NFV initiative and SDN paradigms can significantly contribute to achieve greater elasticity in network service deployments and accelerate prominent innovation in NGSON service provisioning and delivery functions. Indeed, NFV and SDN pave the way for efficiently and optimally combining a series of network appliances while effectively providing QoS adaptive and self-organizing delivery capabilities across networks. On one hand, NFV opens up a promising technological ground to handle network appliances as independent services that can be flexibly composed to provide dynamic network service chains according to service-oriented principles. Moreover, NFV allows for agile scaling in/out as well as effective and opportunistic placement and replacement of network services thanks to their deployment as virtual machines across servers distributed in multiple facilities of network operators (i.e., data centers). On the other hand, SDN offers the capability to programmatically enforce traffic forwarding rules across network nodes on a per-flow or per-tenant basis. Such capability can be exploited to selectively deliver service data through the dynamically established sequence of network service endpoints, thereby accomplishing the deployment of a network service chain [11].

Ultimately, NFV and SDN can be promisingly employed within NGSON to support the orchestration, establishment, and management of network service chains and enabling QoS adaptive and self-organizing network service provisioning. Leveraging these technological advances, the setup of a dynamic and fluid service landscape is enabled in NGSON, where end-to-end service delivery chains can be elastically provisioned by properly combining network, computational, and storage resources, and dynamically configured and scaled depending on user demand and service requirements [5].

In NGSON, a scenario can be envisioned where service providers may lease a particular service chain with given delivery guarantees from network operators to offer value-added services to users. Requirements posed by the application on network infrastructures may also vary with

context changes (e.g., user location, user access patterns, external events), and adaptation of the network service chain may be accordingly required. Although service providers could in principle directly negotiate network services to optimize the QoS of their exposed services, this approach may not be sufficient in an open service ecosystem where the final service delivered to the end user is based on the composition of services from different providers. Instead, our proposed extension of the NGSON overlay offers the capabilities of discovering, negotiating, and establishing network services on behalf of application-level services across different providers while adapting service QoS to changing requirements, resource availability, or context information.

CONTEXT-AWARE SERVICE COMPOSITION AND DELIVERY

Context-aware composite service provisioning can be divided into three main tasks: service composition specification, service selection and service delivery.

Service composition specification consists in defining a composite service as a flow of functional capabilities (i.e., *abstract services*). The invocation flow can be predefined at design time or specified at runtime via dynamic composition approaches. The output is given in the form of *abstract plans*, and the binding to concrete endpoints is not specified. Context-aware adaptation rules can be applied to modify the service composition logic according to the current situation.

Service selection consists of selecting a service endpoint among available candidates for each abstract service. Selection can be performed by applying context-aware rules and algorithms optimizing a QoS-based utility function for a given composition plan.

Service delivery concerns the use of transport functions provided by the underlying network infrastructure to guarantee end-to-end service provisioning. Different types of service may impose different delivery requirements. In addition to QoS requirements, service delivery guarantees may include the provision of network appliances that extend the network forwarding capability with customizable packet processing features. For instance, the traffic directed to a web server may be processed by a network address translator (NAT), a firewall, an IDS, and a load balancer, while multimedia content delivery may pass through a video optimizer, a firewall, and a NAT. By adopting NFV terminology, hereafter we refer to these network appliances as network functions (NFs), and we suppose they are implemented as virtual NFs (VNFs), although physical legacy network functions provided by dedicated hardware could be considered as well.

NETWORK-AWARE NGSON FUNCTIONAL ARCHITECTURE

Figure 3 shows the functional architecture of NGSON, distinguishing the functional entities (FEs), which are mainly involved in service control and delivery, and interact with the underlying

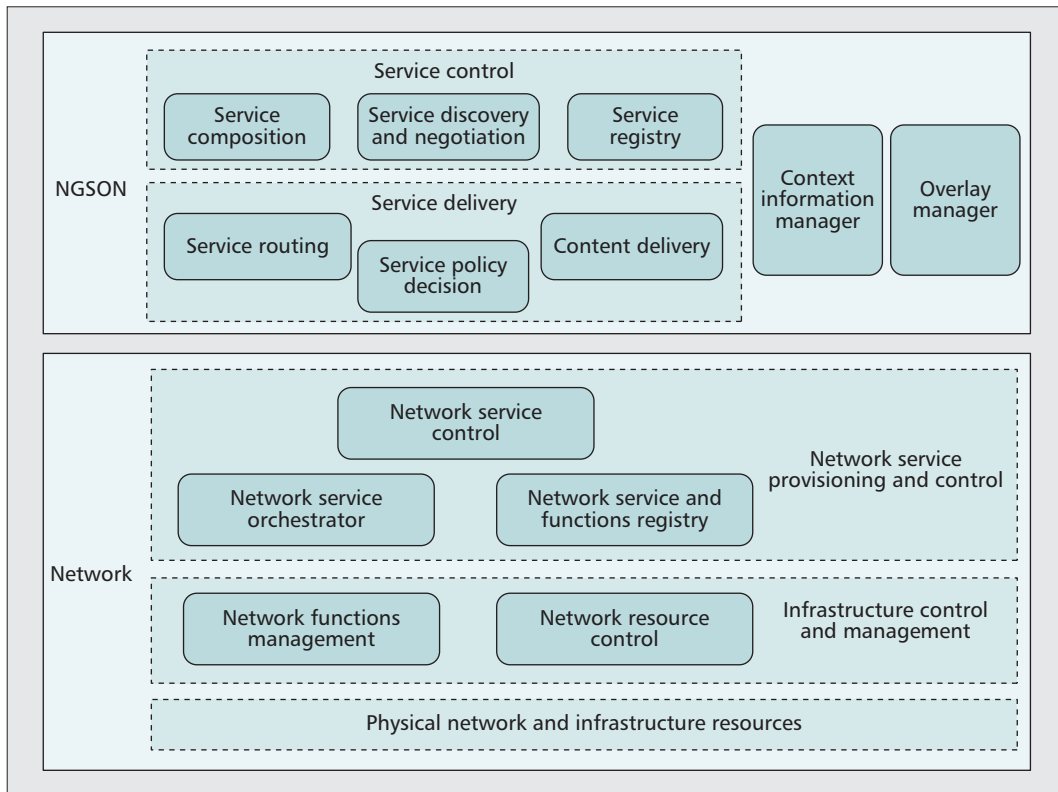


Figure 3. NGSON network-aware functional architecture for adaptive service composition and delivery.

The Service Composition FE handles predefined and dynamic service composition plans. It interprets and controls the runtime execution of the flow of service invocations that addresses a user's request. It can also adapt the composite service specification and adjust its runtime execution according to available context and QoS information.

ing networks to adapt service QoS to context-specific information. Specifically, service QoS requirements are translated into transport-level QoS requirements; accordingly, a delivery path is negotiated with the underlying network decision function able to address specified QoS guarantees. To this purpose, we also provide a model of the underlying network layer that highlights the main functional entities enabling the envisioned NGSON-triggered composite service delivery.

The *service registry* (SReg) FE maintains descriptive information about services. It stores both static and dynamic information, such as service interface description, prerequisites, binding address, roaming status, service priority, and non-functional information (e.g., availability, costs, and reputation). The service profile might also be extended with delivery requirements (i.e., QoS guarantees and network services in the delivery path) in accordance with contracts agreed on with network operators. In this way, each basic service of a composite service is modeled in its turn as a composition of the final application service and the network services required to handle the end-to-end delivery process.

The *service composition* (SCo) FE handles predefined and dynamic service composition plans. It interprets and controls the runtime execution of the flow of service invocations that address a user's request. It can also adapt the composite service specification and adjust its runtime execution according to available context and QoS information.

The *service discovery and negotiation* (SDNeg) FE provides service selection and negotiation

capabilities. It interacts with the SReg to select the service endpoints that match user preferences, context information, or other selection criteria.

The *service routing* (SRou) FE interacts with both the end users and service providers to receive service requests and route them to the appropriate service endpoint. Routing decisions may be optimized according to context and QoS information. It also acts as a proxy between the SCo and the basic services to be invoked in a composite service execution.

The *context information management* (CIM) FE handles and stores context information. It acquires raw information from various sources and provides processing capabilities (filtering, aggregation, and reasoning) to obtain high-level context information that can be consumed by other NGSON FEs and external clients.

The *content delivery* (CDe) FE supports adaptive content delivery by offering cache and forwarding capabilities, content transformation, and merging and synchronization features.

The *service policy decision* (SPD) FE is in charge of QoS negotiation and assurance at both the service and transport levels. The SPD receives requests for QoS assurance issued by the SRou FE, in accordance with the user profile and the service information retrieved in the SReg. The requests are analyzed to derive the corresponding service-related QoS and transport-related QoS.

Service-related QoS requirements are typically specified in terms of service cost, response time, reliability, and reputation metrics. These metrics can be monitored by the SPD or third-party monitoring services when available. Based

If the underlying network cannot meet the target QoS, the SPD may choose among the following alternatives: rejecting the service request, accepting the QoS or best effort service provided by the network, or, finally, negotiating an acceptable QoS.

on the service QoS profiles acquired from the SReg, the SPD FE decides whether or not a QoS requirement can be met. If the service request cannot be satisfied, it is rejected. If more than one service provider can satisfy the QoS requirements, the SPD can apply an optimization algorithm to select the service endpoints that guarantee overall optimal or suboptimal quality.

Transport-related QoS may include bandwidth, end-to-end delay, packet loss rate, and jitter specifications. The SPD FE may interact with the underlying networks to enforce the desired QoS values. If the underlying network cannot meet the target QoS, the SPD may choose among the following alternatives: rejecting the service request, accepting the QoS or best effort service provided by the network, or, finally, negotiating an acceptable QoS.

In this work we extend the role of the SPD FE as defined in the standard specifications with the capability of negotiating a delivery path endowed with the proper NFs deployed as VNF chains. Hereafter, by transport-related QoS we mean an extended QoS requirement also including specifications on the NFs required to adequately process and deliver service data between users and service endpoints. As outlined in the previous section, such QoS requirements can be addressed effectively by leveraging NFV and SDN capabilities. On the basis of these enabling network technologies, we represent the underlying network layer in terms of functional blocks providing *network service provisioning* and *control* functions, and *infrastructure control* and *management* functions, on top of physical network and resource infrastructures.

The network service provisioning and control functions support the dynamic establishment of data delivery channels between service endpoints or between the user and a service endpoint while ensuring adequate transport-related QoS as specified by SDP.

The *network service control* FE is in charge of coordinating the provision of delivery paths through a series of VNFs. It offers a north-bound interface to applications (i.e., NGSON SDP FE) for requesting the selection of delivery paths with specified transport-related QoS. It leverages the *network service orchestrator* FE for selecting the proper chain of VNFs while addressing the QoS delivery requirements. Finally, it gives instructions to the *network resource control* FE to enforce the steering of data flows through a specified chain of VNFs. It is worth noticing that the delivery request issued by the SDP may be targeted to an anycast service address (e.g., multiple service locations are available). In this case, the *network service control* FE will contribute to the final selection of the service location based on the availability of an adequate delivery path.

The *network service orchestrator* coordinates the instantiation, operation, and connection of VNFs to satisfy obligations with service providers, and manage adaptation mechanisms for coping with service providers' requirements and network context changes. This process runs transparent to the NGSON layer.

The *network service & functions registry* FE maintains the information base required by the

network service orchestrator for performing its decision and orchestration tasks. Specifically, it handles descriptive, operational and monitoring information on individual VNFs, VNF chains and delivery paths connecting them. In particular, monitoring information collected through the network function management and network resource control FEs includes up-to-date performance data (e.g., delay) on VNF chains and related delivery paths in order to warn the network service control about the need for cooperative renegotiation with the NGSON SPD.

The *infrastructure control* and *management* functions are responsible for the proper transfer of data across network nodes and VNFs.

The *network function management* FE manages the life cycle of VNFs, in accordance with the instructions received by the network service orchestrator. It also collects monitoring information on the operation of VNF and notifies the service & functions registry when events of interest occur (e.g., unavailability or fault of a VNF).

Network resource control is in charge of the coordinated and automated programming of individual network nodes, leveraging low-level directives for monitoring, discovery, and provisioning on a per-flow basis. Specifically, it handles the requests originating from the network orchestrator FE for connecting a set of newly instantiated VNFs for new or modified service contracts. Moreover, it handles requests from the network service control FE to consistently enforce forwarding rules across network nodes in order to steer service data flows through the desired VNF chain. In this regard, IETF efforts in the design of a comprehensive SDN controller, application-based network operations (ABNO) architecture, provide instrumental guidelines and operational workflow specifications related to the network service control and network resource control functions and interactions, thanks to the utilization of standard protocols and components [12].

Figure 4 shows the NGSON service composition workflow extended with the proposed QoS-adaptive service delivery capabilities.

USE CASE

In this subsection we show a simple use case that illustrates how NGSON FEs cooperate with envisioned network layer FEs. The use case, called *support on site*, consists of a composite service delivered to the employees of an enterprise, according to the QoS parameters that the enterprise negotiated with the service provider.

A service technician is at a customer's site to fix a problem in the evening. He needs help to go on with the intervention. He thus accesses the enterprise service portal to invoke the support on site service, and specifies the topic on which he is seeking support (e.g., "biomedical device X presenting a malfunction of type Y"). The support on site composite service logic first identifies a list of appropriate experts. The expert selection is performed according to context parameters (e.g., an expert's location and current activity). Then, depending on context information, a suitable communication channel between the technician and the expert is established (e.g., a voice call if the expert is available or a chat if the expert is at work

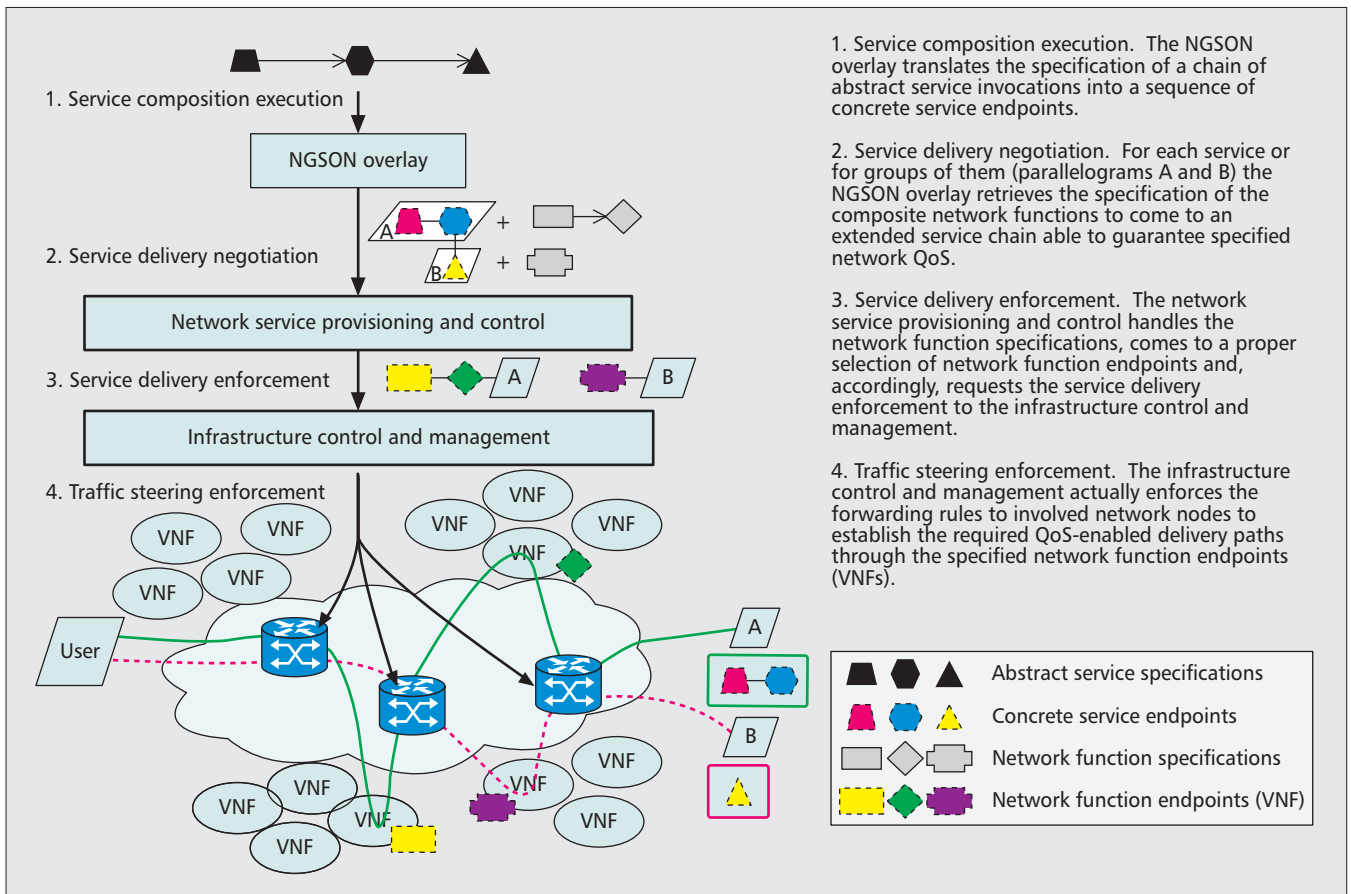


Figure 4. NGSON service composition workflow extended with adaptive service delivery.

but involved in a meeting). If no experts are available at that moment, the composite service logic switches to the invocation of a video tutorial search service providing access to the enterprise knowledge base. In the latter case, the provided response is thus a list of suitable experts with their availability status and a list of matching videos. Since no experts are available at that time, the user decides to watch one of the video tutorials that shows how the malfunction should be handled.

The support on site composite service is specified as a flow of invocations to different types of information and content provisioning services. Hereafter, we detail the case where the identified experts are not available, and the user decides to watch the video tutorial. The main interactions among the FEs are shown in Fig. 5 and described hereafter.

The SRou FE forwards the request received by the user to the SCo FE. This entity retrieves the composite service specification, which includes the flow of abstract services to be invoked, and sends a discovery request to the SDNeg FE to obtain the corresponding concrete service endpoints. The endpoints may be selected by the SDNeg FE among available candidates (e.g., service replicas) depending on context information. For example, the service instances could be selected based on topological proximity to user location or on service context parameters (e.g., availability and response time).

The SCo FE triggers the invocation of the first service (Expert Finder). This is an informa-

tion-providing service that might be exposed as a SOAP or REST web service to retrieve the list of appropriate experts. Analogously, the SCo requests the invocation of the Video Search service. Then the next main step consists in the delivery of a video tutorial to the end-user device. For each service, the SRou FE obtains from the SReg FE dynamic information on the service functional profile and QoS delivery requirements. For instance, the traffic directed to the server exposing the Expert Finder service has to pass through a sequence of NAT, firewall, IDS, and load balancer functions, while the traffic for the video content delivery should be steered through a video optimizer, a firewall, and a NAT function. Moreover, the chain of network services can be modified according to context information (e.g., a transcoding function could be added to adapt the content to the video formats supported by the user device).

For each service, the SRou asks the SPD to enforce the required QoS with the underlying network. Thus, the SPD negotiates with the network service control function the service delivery with the desired QoS. The network service control FE obtains the VNF chain that best guarantees the desired QoS and triggers proper traffic steering actions by leveraging the network resource control FE, which will program the involved network nodes to enforce the proper traffic flow forwarding rules across network nodes. If the path passing through the desired VNFs is not available, or is available with differ-

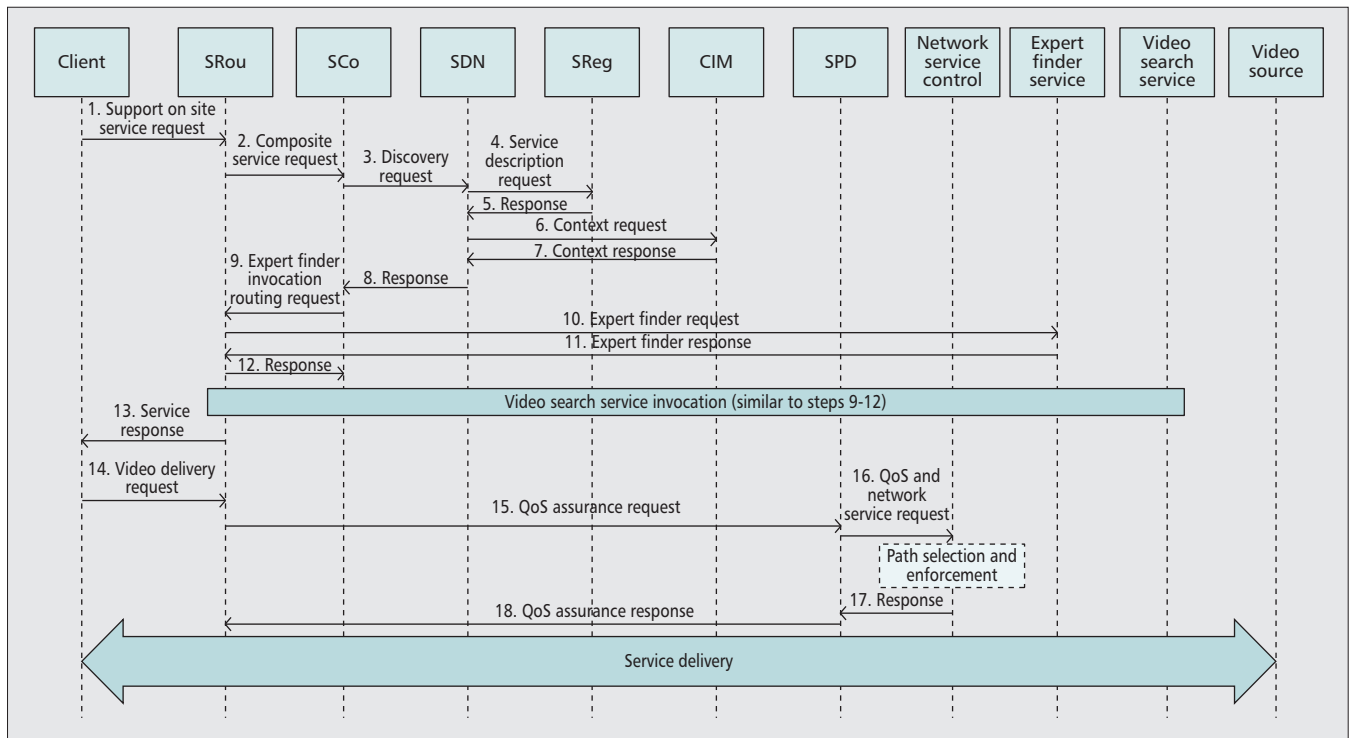


Figure 5. Sequence diagram for adaptive service delivery: main functional interactions for the support on site service use case.

ent QoS guarantees, the SPD is notified through network service control and decides to accept or reject the delivery offer.

CONCLUSIONS

In this article, we present a proposal for making NGSON evolve to take full advantage of recent trends in the research on SDN and NFV network technologies. We introduce a network-aware service composition and delivery model that leverages network virtualization mechanisms and programmable traffic steering capabilities. This model is based on the extension of the NGSON functional architecture to accommodate proper negotiation mechanisms with underlying network service control capabilities. To this purpose, we extend the concept of transport-related QoS to also include the specifications on network functions required to adequately process service data in the delivery path between users and service endpoints. Consequently, we also enhance the role of the SPD FE with the capability of negotiating this extended transport-related QoS with the network level. Finally, we have shown through a use case how this model can be put into operation. As a result, this confirms the central role that NGSON can play in enabling and facilitating the conception and deployment of value-added application services in future network scenarios.

For its part, the proposed evolution of NGSON specifications poses a set of challenges for SDN and NFV research areas. This includes the need for scalable mechanisms for the dynamic orchestration of NFV instances, their optimal placement and replacement for supporting efficient scaling mechanisms, and increasingly dynamic and granular traffic steering capabilities scaling at the level of per-tenant and per-application flows.

Finally, this article proposes one of the possible approaches of interoperation among NGSON, SDN, and NFV initiatives. In fact, official liaisons have recently been set up, and we expect that further joint research directions will emerge within the official standardization activities. For instance, it would be interesting to investigate how NGSON overlay nodes could be elastically deployed as VNF themselves in fully virtualized infrastructures.

REFERENCES

- [1] IEEE Std 1903-2011, "The Functional Architecture of Next Generation Service Overlay Networks," Oct. 2011.
- [2] M. Ulema et al., "Next Generation Service Overlay Networks (NGSON)," Guest Editorial, *IEEE Commun. Mag.*, vol. 50 no. 1, 2012, pp. 52–53.
- [3] S.-I. Lee and S.-G. Kang, "NGSON: Features, State of the Art, and Realization," *IEEE Commun. Mag.*, vol.50, no.1, Jan. 2012, pp. 54–61.
- [4] W. John et al., "Research Directions in Network Service Chaining," *Proc. 2013 IEEE SDN for Future Networks and Services*, 11–13 Nov. 2013, Trento, Italy, pp. 1–7.
- [5] S. Latre et al., "The Fluid Internet: Service-Centric Management of a Virtualized Future Internet," *IEEE Commun. Mag.*, vol.52, no. 1, Jan. 2014, pp. 140–48.
- [6] B. Nunes et al., "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," *IEEE Commun. Surveys & Tutorials*, pp. 1–18.
- [7] ETSI GS NFV 002 V1.1.1, "Network Functions Virtualisation (NFV); Architectural Framework," Oct. 2013.
- [8] Q. Duan, Y. Yan, and A. V. Vasilakos, "A Survey on Service-Oriented Network Virtualization Toward Convergence of Networking and Cloud Computing," *IEEE Trans. Network Service Management*, vol. 9, no. 4, 2012, pp. 373–92.
- [9] D. Matsubara et al., "Toward Future Networks: A Viewpoint from ITU-T," *IEEE Commun. Mag.*, vol. 51, no. 3, March 2013, pp. 112–18.
- [10] S. Davy et al., "Challenges to Support Edge-as-a-Service," *IEEE Commun. Mag.*, vol. 52 no. 1, 2014, pp. 132–39.
- [11] Y. Zhang et al., "StEERING: A Software-Defined Networking for Inline Service Chaining," *Proc. 2013 21st IEEE Int. Conf. on Network Protocols*, Goettingen, Germany, Oct. 2013.

[12] D. King and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations," IETF Internet-Draft, draft-farrkingel-pce-abno-architecture-07, 13 Feb. 2014.

BIOGRAPHIES

FEDERICA PAGANELLI [M'07] (federica.paganelli@unifi.it) received her Ph.D. degree in telematics and information society from the University of Florence, Italy, in 2004. She is now a senior researcher at the National Interuniversity Consortium for Telecommunications (CNIT), Italy. Her research interests include context-aware systems, ambient intelligence, service-oriented computing, and next generation networks. She is also involved in standardization activities in the IEEE NGSON Working Group.

MEHMET ULEMA (m.ulema@ieee.org) is a professor of computer information systems at Manhattan College, New

York. Previously, he was with AT&T Bell Labs and Bellcore. He has been on the editorial boards of a number of journals, including *IEEE Transactions on Network and Service Management* and *Springer Journal of Network and Services Management*. He has served as General Chair and Program Chair for a number of IEEE conferences, including ICC and GLOBECOM. He received his M.S. and Ph.D. from Polytechnic Institute of New York University, and B.S. and M.S. degrees from Istanbul Technical University.

BARBARA MARTINI (barbara.martini@cnit.it) received her Master's degree in electronic engineering from the University of Florence, Italy, in April 1999. She is currently a senior researcher at CNIT National Laboratory of Photonics Networks, Pisa, Italy. Her main research interests include service platforms for next generation networks and smart cities, software-defined networks for clouds and data centers, and security for multi-domain networks.