CASE STUDY – PART I

Operator-governed, end-to-end, autonomic, joint network and service management

Abstract

A Network Operator (NO) wants to deploy new services and/or accommodate new traffic on top of its multi-vendor and multi-technology infrastructure involving both Radio Access Networks (RANs) and backhaul/core segments. The automation and achievement of coordinated, end-to-end performance in such a process is both desirable and challenging, especially considering the problems dealing with the different manifestations of heterogeneity in the considered underlying networks i.e. heterogeneity in the used technologies and domains, in the equipment coming from different vendors and of course in the management tools/systems.

Although there have been efforts towards integrating and automating such service deployment processes, they are not completely successful in delivering a convincing, end-to-end solution. In addition, even though they may be elaborated enough, they exhibit loose or no integration and partial or no automation at all, having humans with special skills and expertise still being highly involved into the processes.

This case study reports on the above problems and calls for solutions that will provide a unified, goal-based, autonomic management system for the service deployment and/or new traffic accommodation on top of heterogeneous networks encompassing both OFDM-based RANs and MPLS-based backhaul/core segments.

This document presents a brief description of the case study, its methods, concepts and expected innovation. The specific functional, non-functional requirements and the associated problems of this case study were presented in the deliverable D4.1 [1]. The prioritization of the problems and functional requirements were presented in deliverable D4.2 [2].

Date of release

17/09/2012

CONTENT

STORY LINE

In a continuously alterable telecommunications ecosystem, Network Operators (NOs) ask for the capability of dynamically introducing new load (specific services and user classes) to their networks. This includes the dynamic service deployment of a new service/application or introduction of new traffic for a given service (or real time modification of service characteristics) on top of its multi-vendor, multi-segment and multi-technology infrastructure and the required fulfilment and resource provisioning so as to preserve the requested Quality of Service (QoS) and Quality of Experience (QoE) levels. For tackling such a situation today, operators would rely on processes that are not as flexible as they need and can be, and therefore, impose costs. In general, the solution of the problem relies on (in very high level terms): (i) planning and deployment (rollout); (ii) optimization and maintenance.

On the one hand, planning is an essential phase in the engineering of telecommunication systems. Nevertheless, telecommunication systems face changing situations, due to the time variant traffic demand, the occurrence of faults and mobility and radio conditions, in case of wireless access. As a result, handling all the potential situations, based only on planning, means that the worst (most demanding) scenario has to be considered as the reference one, against which the network has to be planned. This leads to unnecessary over-provisioning of resources (e.g., elements, bandwidth, etc.), which negatively impacts the cost (Capital Expenditures – CAPEX). In this framework, the dynamically adaptation of the network to the encountered situation, through management functionality, is the solution to the problem.

On the other hand, management relies on processes that are elaborate but not fully automated, as they have to deal with heterogeneous technologies, which are not adequately integrated as shown in Figure 1. More specifically, the management processes/systems of an operator will typically adhere to specific standards. In general, these systems are heterogeneous, depending on the technology and on the vendor of the technology. This means that, in principle, the management systems of a wireless and wireline access technology will be different. Moreover, the management systems for a specific technology, obtained by two different vendors, will be different. The heterogeneity means that there is little or no integration between the management processes/systems. This negatively impacts the time required for (re-) configuring the infrastructure. Moreover, it means that human intervention is required in the process that leads to cross-technology configurations. This can cause, apart from delays, errors and inconsistencies. Finally, loose or no integration means that the information available to the different systems cannot be readily exploited for the purpose of optimizing the operation of the infrastructure as a whole.

Heterogeneity as a barrier to end-to-end service delivery

- *Loose or no integration*
- *Partial or no automation*
- *Manual translations*
- *Over-provisioned static configurations due to worst-case planning*
- *Increased time-to-market*
- *Increased CAPEX/OPEX*

In addition service management and customer relation management rely on processes that are not fully automated. Several aspects (phases of the overall process) often require manual intervention and/or the use of heterogeneous systems that are not integrated. This increases the cost of managing the customer relations.

Last but not least, manual configuration of network devices that requires strong technical expertise of at least one specialist per network segment is a standard situation that leads to an increase in the OPEX (Operational Expenditures).

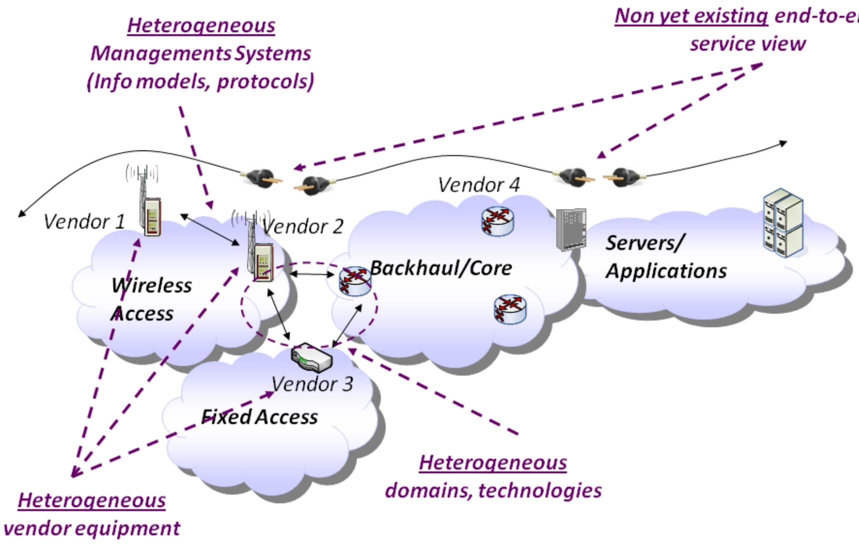

Figure 1. Network architecture composed of heterogeneous technologies

The above unveil great challenges in achieving the required end-to-end, dynamic service delivery, which the operators must address so as to have a competitive advantage over those who do not. The question is how can the operator escape from his cumbersome tactics and ensure that the customer order will be satisfied in a fast, reliable and cost-efficient manner. The goal of this study is exactly to provide an answer to that question.

So, it starts by assuming a dynamic service deployment scenario, where a (mobile network) operator receives an urgent request concerning a real time, video-based application (e.g. video-streaming of a programmed event), a specific geographical region and time period and with specific QoS/QoE requirements provided in a form of a dynamic Service Level Agreement (SLA). The operator owns a multi-vendor, multi-technology infrastructure encompassing OFDMA-based (e.g. LTE) RAN and IP/MPLS based Backhaul/Core Network. The case study aims at finding solutions that will assist operators in automating service deployment and/or new traffic accommodation on top of their infrastructure and in particular by:

- Enabling operators to describe their goals and objectives through high-level means, and govern their network.
- Achieving policy-based operation of RAN (OFDMA-based) and Backhaul/Core Network (IP/MPLS-based) segments, which is optimized with respect to QoE/QoS efficiency, taking into account metrics and knowledge derived in network nodes and end-user devices that are aligned with the operator objectives.
- Achieving coherence between these segments through cooperation, negotiation and federation

Network phases

This case study covers all the potential network phases spanning Network Deployment, Network Operation and Optimization. Moreover, Figure 2 depicts the areas of TMF Forum’s eTOMap [3] that are impacted by the case study.

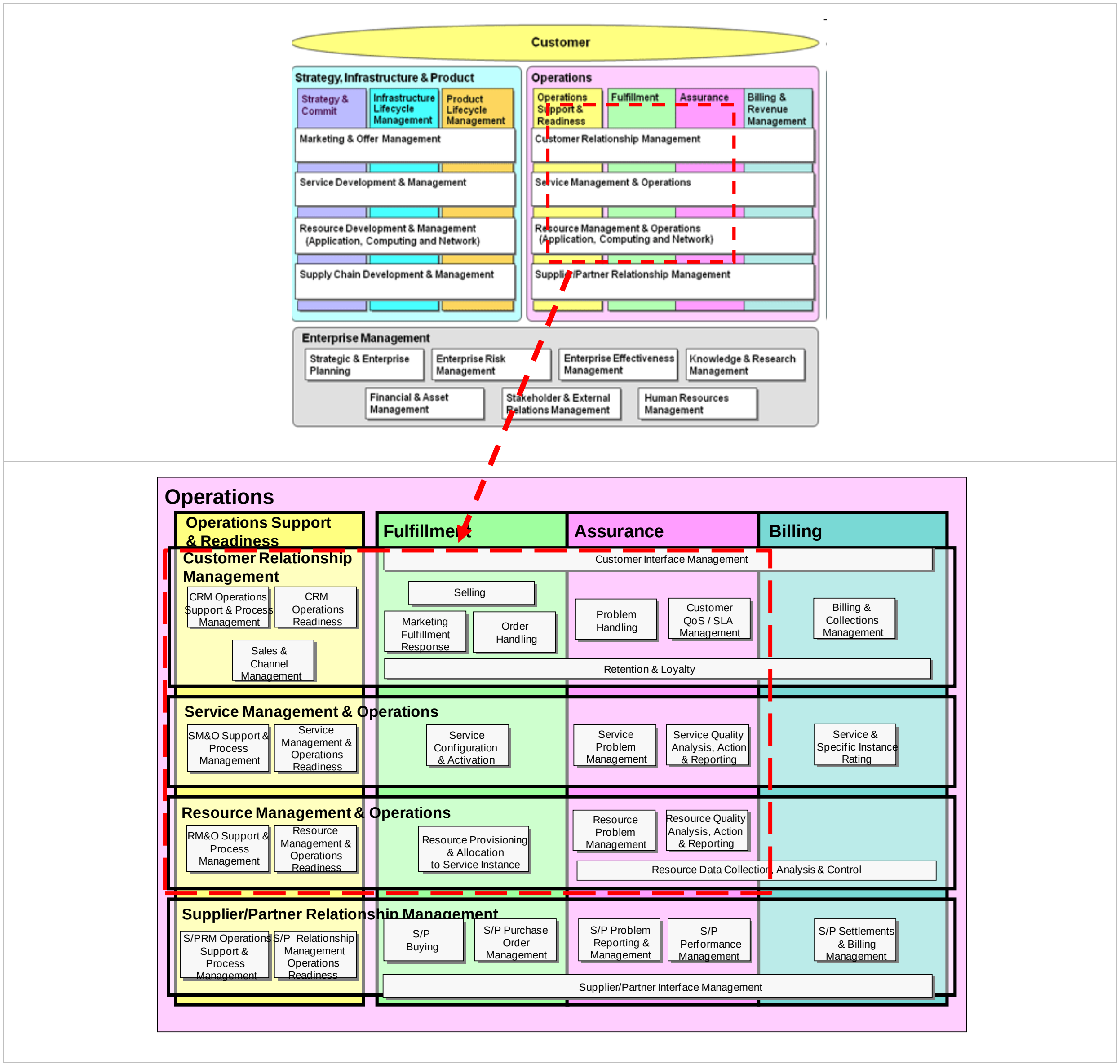

Figure 2. Mapping to eTOMap (two levels of granularity)

PROBLEM STATEMENT

The main problem that it is tackled in this case study is the holistic *operator-governed end–to-end autonomic management of heterogeneous networks and service management according to high-level operator goals.* However, the complexity calls for proper decomposition into a set of (sub) problems, which are presented in the following:

Problem: ***Setting the business goal***

According to the trigger, the operator defines business goals/policies, in high-level terms. Policies are then derived according to the higher level goals, to provide constraints and priorities, in respect of the new accrued conditions. The derived policies are assessed against existing goals/policies so as to be identified and resolved potential conflicts (in fact, conflicts can arise if the defined goal/objective/policy is conflicting with previous goals or the impact of these goals on already deployed services).

Problem: ***Analysing the business request***

The inputs/requirements embodied in policies, derived from the business entry, need to be elaborated and correlated together with pertinent knowledge stemming from user & service raw data in order to be derived technology (network) specific requirements.

Problem: ***Determination of candidate solutions***

This problem concerns the determination (reasoning with) of the candidate solutions (networks) that can satisfy the derived network requirements. The candidate technologies and networks, which can contribute to this satisfaction, need to be discovered, taking into account existing knowledge that was extracted from raw network data.

Problem: ***Invocation of RAN***

This problem deals with the invocation of the selected RANs and the request for an offer in terms of the quality, which the RAN can provide. RAN (actually the respective management/control entities) investigates ways to accommodate the request (anticipated load). In OFDMA-based (LTE) case this may result in solution by means of radio resource allocation, scheduling, relay positioning in case of multi-hop networks, link selection, compensation by means of SON mechanisms etc.

Problem: ***Invocation of backhaul/core segment***

The backhaul/core segment is triggered and the network morphing is activated. Network morphing, consisting of a variety of different methods, methodologies and techniques, aims at finding the ideal configuration, which corresponds to the network configuration that provides the best possible performance, in order to support the solution (offer) provided previously by RAN, while maximizing network utility, relatively to a network operator predefined cost function. The outcome of these processes may involve Label Switched Path (LSP) configuration in IP/MPLS case, as well as Gateway (GW) (e.g., Service GW (SGW), Packet Data Network GW (PDN-GW)) (re) selection/configuration, GW migration/dimensioning.

Problem: ***Achievement of coherence***

The problem here is to resolve possible incompatibilities between the offered QoS from RANs and backhaul/core segments, respectively. Such incompatibilities can also occur in the case of overlaying/virtual networks that are built on top of the segments and have different QoS objectives than the segments they rely upon. For that reason, some sort of negotiation and cooperation between segments is needed that will be used to fine-tune the resulting offers from the underlying segments, in order to achieve coherence.

Problem: ***Configuration***

At this phase, the problem is to proceed with the actual application of the configuration decisions in the RAN and core/backhaul network nodes. First, it’s necessary to identify concerned equipment and request each of them to perform the appropriate reconfiguration actions. Then, each of the targeted equipment has to translate and enforce the decision e.g. activation of technology specific, lowest level policies and device-commands, while taking into consideration the appropriate protocol and configuration parameters.

Problem: *Assurance*

Having configured the network, continuous monitoring is needed for collecting measurements (i.e. Performance Measurements and/or UE measurements) in order to ensure that the desired QoS level is guaranteed during the operational phase of the service. Also reactive actions such as admission/congestion control and reconfiguration actions can be triggered in order to adjust the network configuration parameters following the traffic and network conditions.

It must be noted that the elaboration of the above problem resulted in a rich set of requirements that have been grouped within Deliverable 4.1 “Synthesis of use case requirements - release 1” [1] and were further elaborated and prioritized in Deliverable 4.2 “Synthesis of use case requirements - release 2” [2].

MODELLING

In this section, a modelling of the case study is presented. The case study lifecycle is decomposed in several phases that each of them corresponds to a specific task that is performed within the case study. Triggers that enable the case study functionality and the actor, who interacts with the functionalities of this case study, are also identified.

Actor(s)

In this case study there is a single actor i.e. the Network Operator (NO). The NO is the one that initiates the lifecycle of the case study and the one that receives the feedback from it.

The NO initiates the case study lifecycle by using the human to network interface to express business level goals/requests (highest level policies) that the network has to meet.

The NO receives notifications for taking actions e.g. to adjust manually the network configuration parameters following the traffic and network conditions, when such adjustments cannot be autonomously made.

Triggers

The triggers that derive from/aim to the NO, in this case study, are the following:

- New business level goals/requests (e.g. the deployment of a new service) from the NO, which need to be met by the network.
- In the opposite direction, a notification is sent during the Assurance phase so as to inform the operator about the adjustments of the network configuration parameters made autonomously by the system or to request to perform further actions.

Phases

The lifecycle of this case study can be decomposed in several tasks, each of which corresponds to a phase of the case study. These phases have been already introduced in the previous section in the form of the problems that are tackled in this case study. There is a one-to-one relation between the problems, already introduced, and the phases of this case study. The sequence of the phases into the case study lifecycle is described below.

Initially the NO provides the business level goals/requests that the network has to meet through the human to network interface. This triggers Phase A, which derives conflict-free policies based on the new NO's business level goals.

Then Phase B is triggered, where network level/technology specific requirements (e.g. accurate estimation of traffic bandwidth needs) are derived based on the policies. Derivation mechanisms in Phase B utilize variant type of information related with the concrete network, e.g. information related to the existing application, the user class profile and the Service Level Agreements (SLAs). These network level/ technology specific requirements trigger Phase C.

In phase C, the candidate technologies and networks in the RAN and in the Backhaul/Core segments, which can contribute to the satisfaction of the service requirements, are identified. This identification takes into account existing knowledge that is already extracted from raw network data, comprising for example, information of the available network elements in the RAN segment and the current network resource availability. The set of candidate networks of the RAN and the corresponding network level requirements (for example, anticipated load related information and status information of the candidate networks of the RAN segment) trigger Phase D.

In phase D, the invocation of the selected RANs is performed, in order to accommodate the request (anticipated load). The traffic offer from the RAN triggers the next phase, which is Phase E.

Powerful Governance interface

Autonomic/Knowledge-based, resource provisioning

End-to-end optimization (RAN /Core)

In phase E, the Backhaul/Core finds the best configuration (utilizing algorithms and techniques for optimization of protocol parameters and taking into consideration the different possible reconfigurations and the related costs) and accordingly makes an offer of quality, so as to support the solution (offer) provided previously by RAN. The traffic offers, from the RAN and the Backhaul/Core segment respectively, trigger Phase F.

Possible incompatibilities between the offered QoS from RAN and backhaul/core segments are resolved in phase F. The end-to-end offers after the achievement of coherence trigger Phase G.

In phase G, the actual configuration of the RAN and Core Network nodes is applied (comprising changes in protocol and configuration parameters), according to the configuration required, in an autonomic way.

Finally, phase H is triggered by the end-to-end configuration already applied to the network segments. In this phase continuous monitoring and reactive actions are performed in order to ensure that the desired QoS level is guaranteed during the operational phase of the service (achievement of SLA compliance). This phase is also responsible for notifying the NO, via the human- to-network interface, in case actions are required e.g. manual adjustment of network configuration.

The case study actor, the triggers and the phase sequence are presented in Figure 3 using the concept map modelling proposed in [4].

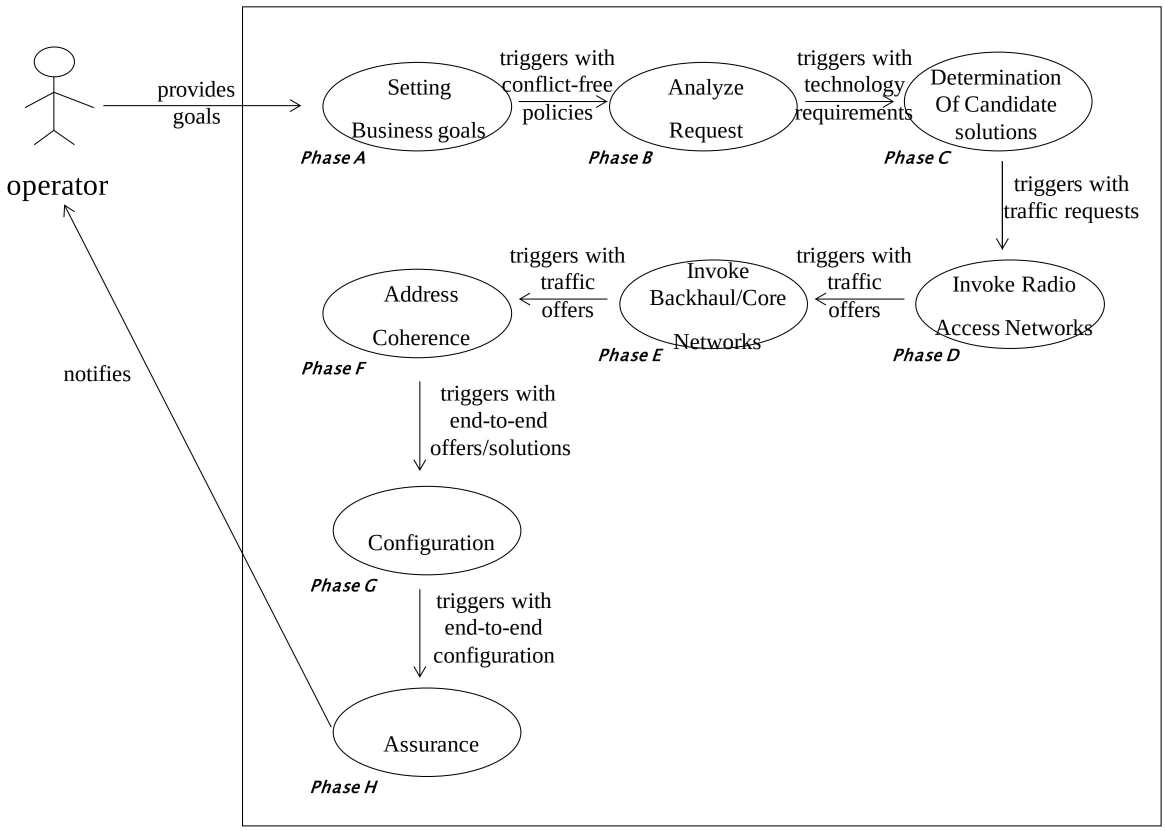

Figure 3. Concept map for the case study

INNOVATION

Enabling concepts and mechanisms

The main innovation in this case study lies in the design of a unified goal based management system for service deployment on heterogeneous networks encompassing both RANs and backhaul/core segments. In the preceding analysis, a set of problems and challenges associated with bringing such innovation were identified and discussed. More importantly, they can be used to designate the design and specification of the solution to the above problems. Specifically, the case study identifies the need to gradually upgrade the whole management chain with systems that will be able to:

- provide the operator the means to express their goals and govern (control) their possibly self-x capable network and this can be only done through a robust policy based framework
- achieve end-to-end integration, federation of wireless/wired access and core/backhaul network segments and their associated management systems, and all these under the "auspices" of operator policies
- maintain and exploit always-up-to date inventories and knowledge (possibly derived through incorporated learning mechanisms) in all the situations above, thus increasing the reliability and adaptability of management decisions, and contributing to autonomicity
- provide an optimized resource provisioning for RAN and Core (backhaul) segments, based on policies and knowledge

Figure 4 gives a high level view of the envisaged solution and its ingredients.

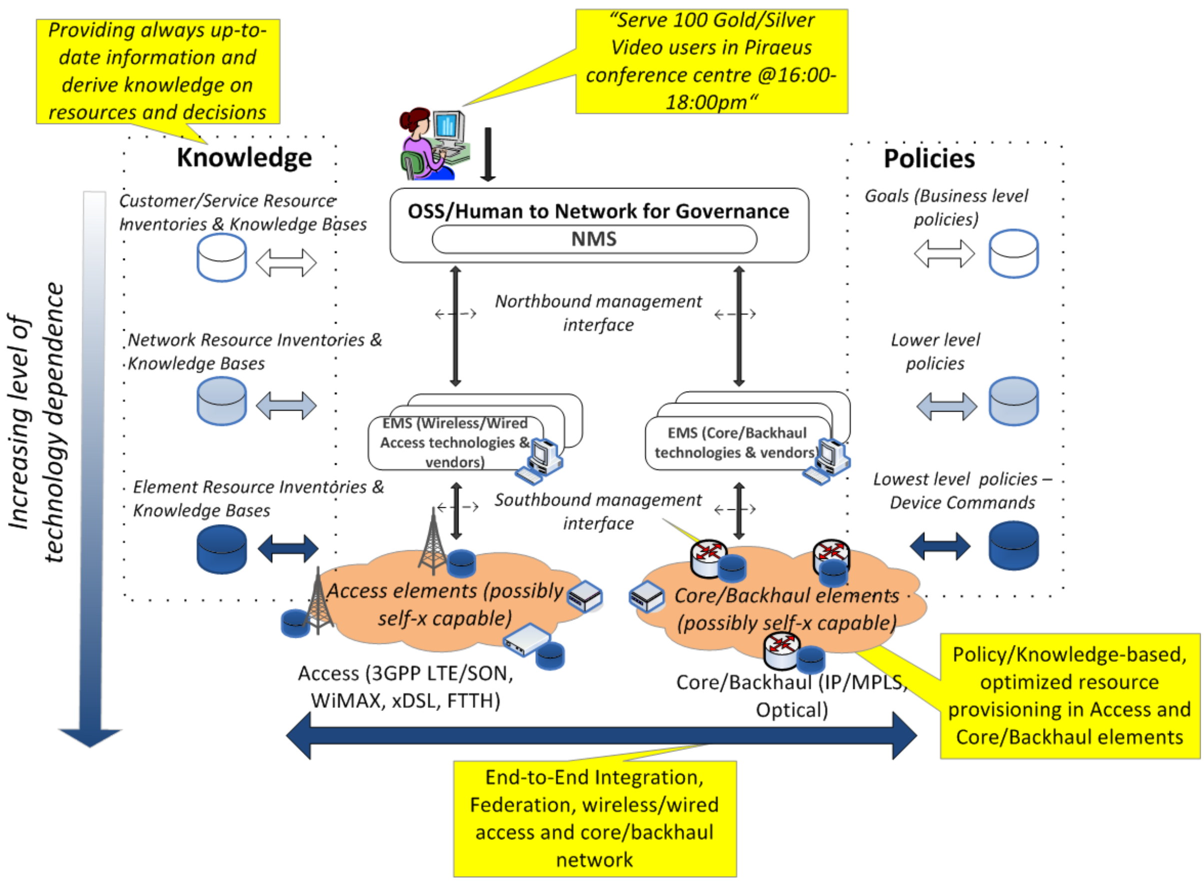

Figure 4. High level view of the designated solution [5]

Differentiation from the state of the art

The novelty of the case study is reflected through the following prominent keywords: automation, knowledge-based operation, end-to-end flexible service and resource management and joint service and network management, which are all aligned with the project objectives. More specifically, the innovative approaches that are designated by this case study are the following:

- The human-to-network interface, which does not exist in current systems and will be used to introduce the business level goals in high level terms, and leave the system to autonomously resolve the situation and meet the objectives.
- The joint management of services and networks and the translation of service requirements into network requirements are performed in an autonomic manner. As of today, services and networks are managed separately and such mapping of service requirements into network requirements is mainly decided and deployed by humans with the aid of some partially automated tools and scripts. As a general practice, people in charge of development and planning will consider these needs and make traffic estimations and assumptions regarding the related SLAs and will define the related network requirements.
- The selection of candidate technologies and networks, which will contribute to the satisfaction of network requirements, is defined by taking into account existing knowledge that was extracted from raw network data. The legacy situation is that network requirements are sent to the operational teams in charge of the network segments (planning, assurance) who finally select the candidate networks. These operations are in general manually done, with static processes and without considering any accumulated knowledge.
- The invocation of the selected RANs is done in an autonomous way as the RAN investigates way to accommodate the request (anticipated load) and it also makes an offer in terms of the quality that it can provide. In current systems, the RAN invocation phase is carried out by the RAN administrator who is responsible for the management of the targeted RANs and RAN elements and has to make an estimate of the available network resources, considering a given operating point of the network, typically the worst case.
- The backhaul/core invocation is autonomously done also, and the network morphing provides a dynamic adaptation of the connectivity to support/accommodate the request provided previously by RAN. Furthermore, in this framework, new traffic engineering optimization techniques are applied. Taking into account core/backhaul network aspects while allocating traffic and QoS into RAN is not the legacy situation. As in the case of RAN, the description of traffic demands is sent to the operational teams who are in charge who interact with elements/network in the core/backhaul segments by using CLI-based remote logins or SNMP-based request/responses in order to obtain the current status that will be used for their offline calculations/estimations.
- The negotiation and cooperation between both the RAN and Backhaul/Core segments to fine-tune the resulting offers of anticipated load, in order to achieve coherence in the performance between segments. As of today, interaction of admin/management domains (physical interaction between administrators) is definitely needed for achieving coherence between the offers from both RAN and backhaul/access networks, however such synchronization of all the participating segments takes too much time.
- The assurance that the desired QoS level is guaranteed during the operational phase of the service. This involves continuous monitoring and actions for adjusting the network configuration (reconfiguration) parameters following the traffic and network conditions. Currently the performance analysis is done by periodically activating collection of measurements. The results are elaborated offline and manual changes of the network configuration are performed. Deciding what to monitor and eventually report and visualize in the Human-to-Network governance GUI is very important when considering the autonomic nature of infrastructure. It must be noted that in scenarios like the assumed one, where the service could be up and down and/or even modified in the time scale of minutes, service fulfilment/provisioning are actually intertwined with the service assurance [6].

Impacts and benefits

It is obvious that the autonomous operations that are described in this case study will significantly impact the transaction costs of the system. The deployment of such autonomic self-management procedures using the network knowledge reduces the need for human intervention and subsequently lowers OPEX. This is possible as the configurations of devices, which were performed manually by technical experts so far, are performed in an autonomic manner. Furthermore the unified management of the heterogeneous technologies and the integration of the management processes imply a reduction of the demanded time and of required efforts (e.g. reconfigurations) for optimally reconfiguring the infrastructure. Moreover, they will help guaranteeing the requested QoS or even increasing it, while improving the capacity of the operator network, both of them offering the operator more revenues, a larger market share and a good reputation in the market. And also this process does not require several different experts, in order to perform the reconfiguration. As a result of this, it is obvious that there is a surplus reduction of the OPEX. At the same time, network knowledge allows the dynamic allocation and the optimal usage of network resources (CAPEX reduction), as well as the reduction of the average number of resources across the different layers and segments (PowerEx – Power Expenditure – reduction), for example through “cognitive” traffic engineering decisions. This avoids the static provision of resources that derives from the planning-based allocation, and in many cases leads to an over-provisioning (which negatively impacts the CAPEX), due to the fact that the planning phase often relies on worst-case estimates. A reduced time-to-market is also envisaged due to the reduction of expenditures (OPEX and CAPEX).

TO BE CONTINUED

This piece of work is the first part of a case study called "Operator-governed, end-to-end, autonomic, joint network and service management" that is published in the context of this project and aims at introducing to the reader to the story line, the problem statement, the associated requirements and challenges and the innovations proposed. The next part of this case study (Part 2, expected October 2012) will be published in the sequel and will focus on the detailed design, specifications, as well as performance evaluation of a solution to the problems described herewith. The series will be completed with Part 3 (expected Summer 2013) that will focus on assessing the impact of the fully validated (evaluation, feasibility, proof-of-concept) solution.

REFERENCES

[1] Univerself Deliverable 4.1 “Synthesis of use case requirements, release 1”, July 2011, publicly available at: http://www.univerself-project.eu/news/new-report-available-synthesis-use-case-requirements-release-1-deliverable-d41

[2] Univerself Deliverable 4.2 “Synthesis of use case requirements, release 2", April 2012, publicly available at: http://www.univerself-project.eu/news/new-report-available-synthesis-use-case-requirements-release-2-deliverable-d42

[3] TMForum: eTOM - Telecom Operators Map (2006), http://www.tmforum.org/BusinessProcessFramework/1647/home.html

[4] Univerself Milestone 28, “Method for graphical representation of use-case”, 2011, publicly available at: http://wiki.univerself-project.eu/ms28

[5] K.Tsagkaris, P.Demestichas, G.Nguengang, I.G.Ben Yahia, P.Peloso, "Unveiling technical challenges for the governance of end-to-end service delivery and autonomic infrastructures", to appear in Proc. *IEEE Global Communications Conference (GLOBECOM 2012),* Anaheim, California, USA, 3-7 Dec, 2012

[6] D.Thomas, "Closing the Loop between Assurance and Fulfillment", Pipeline Publishing, Volume 4, Issue 7, online available at: www.pipelinepub.com

[7] Jacobson Ivar, Christerson M., Jonsson P., Övergaard G., Object-Oriented Software Engineering - A Use Case Driven Approach, Addison-Wesley, 1992

[8] Cockburn, Alistair. Writing Effective Use Cases. Addison-Wesley, 2001.

CONTACT INFORMATION

For additional information, please contact: Kostas Tsagkaris (ktsagk@unipi.gr) or Panagiotis Demestichas (pdemest@unipi.gr); or consult www.univerself-project.eu

UNIVERSELF CONSORTIUM