

Service Provision Evolution In Mobile Communication Networks

Vangelis Gazis¹, Nikos Houssos¹, Anastasia Alonistioti¹, Lazaros Merakos¹,

¹ Communication Networks Laboratory, Department of Informatics & Telecommunications
University of Athens, Panepistimioupolis Ilisia, 15784, Athens, Greece
e-mail: {gazis, nhoussos, nancy, merakos}@di.uoa.gr

Abstract - Following the commercial deployment and worldwide success of 2nd generation mobile telecommunication systems, standardization bodies, industry partners and regulatory fora from around the globe joined forces in producing the standards for 3rd generation mobile telecommunication systems. In parallel, the academic research community has been gradually shifting its focus in defining the scope of 4th generation systems. The present paper highlights fundamental concepts and key market developments in 4th generation mobile networks, outlining their major high level requirements and identifying the fundamental building blocks of 4th generation system architectures in terms of technological solution sets, proposing also a set of priorities for the 4th generation research agenda.

I. INTRODUCTION

In the European Union, the debate about 4th generation systems has taken place mostly within the context of the IST Framework Programme activities [1] [2]. Out of this – still ongoing – process has spawned the vision of a system that enables an “always on, always best connected” mode of communication [3]. This widely accepted vision sketches a heterogeneous communication landscape comprising different wireless access systems in a complementary manner, where the user, supported by his/her personal intelligent agent(s), enjoys untethered connectivity and ubiquitous access to applications over the most efficient combination of wireless systems available (Figure 1).

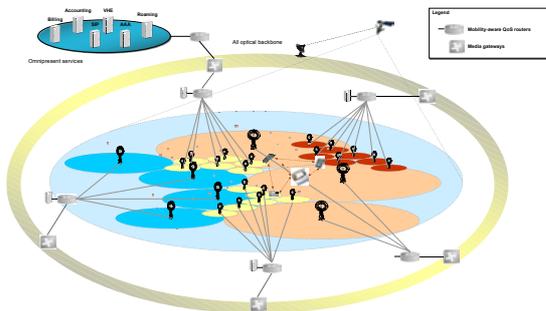


Figure 1. The 4th generation mobile network.

Future mobile communication systems will be heterogeneous in nature, forming an integrated network environment that comprises various wireless technologies and access systems in a complementary manner. These individual access networks will interface to core network elements over the IP protocol, the lingua franca of networking technology. Regardless of their internal technical details (e.g.,

licensed or unlicensed frequency band, signaling protocols, air interface standard, etc), heterogeneous wireless access networks are expected to have the following in common:

1. A dynamic address assignment mechanism (e.g., DHCP, SLP, GPRS/UMTS) that is capable of associating a short-lived or long-lived IP address to the respective wireless interface at the mobile terminal.
2. A transparent IP forwarding service that is accessible over the logical termination of the IP layer at the mobile terminal and one or more gateways (e.g., GGSN, Mobile IP-aware router) at the wireless access network infrastructure. The IP forwarding service is established by employing signaling procedures (e.g., PDP context signaling in the UMTS case) specific to the technical architecture of each wireless access network.

Recent research contributions on the similarities of wireless network architectures [4] clearly support these assumptions.

The rest of the present paper is structured as follows; Section II presents the fundamental high-level requirements of 4th generation systems. Section III highlights important value chain impacts in 4th generation mobile communication environments and identifies the pivotal roles engaged in the service provision process along with its critical aspects. Section IV identifies the key enabling technologies to address the aforementioned requirements. Section V concludes the paper and proposes a set of directions for the 4th generation research agenda.

II. HIGH LEVEL REQUIREMENTS

A. User requirements

In the forthcoming 4th generation era, the mobile user will expect to be able to access the same services from a variety of terminals with highly diverse characteristics (e.g., cellular phone, PDA, laptop computer) in terms of processing power, availability of persistent storage and content visualization capabilities. In addition, when browsing for available applications and value-added services, he/she will expect to be presented with a list that matches not only the current capabilities of the mobile terminal equipment in use, but also his/her personal preferences (e.g., language). These technically demanding expectations raise the need for a flexible service provision process that accommodates these disparate concerns and undertakes all necessary interactions with the

underlying network infrastructure and the application-hosting servers [5].

B. Service provision requirements

It is understandable that a flexible transitional roadmap that will smoothly transition existing infrastructure towards the 4th generation will be necessary. Since novel technological capabilities will be introduced in the existing communication infrastructure in an incremental manner, rollout of (value-added) services and other applications over a mosaic of different mobile networks will be a task of significant perplexity that raises the bar on service management requirements. Important issues include the rapid and automated deployment of disparate services over diverse technological environments and the availability of a one-stop shop for mobile users that will orchestrate the personalized discovery and optimal provision of services via a universal user interface, thereby assisting the user in finding his/her way within the space of available service options.

C. Quality of service & billing requirements

The provision of quality of service guarantees for user traffic streams is intrinsically bound to the application of differentiated pricing and billing schemes. The multitude of players in the upcoming 4th generation value chain and the plethora of possible pricing and billing combinations suggests that the underlying pricing and billing architecture should adopt a segregated structure where access and transport related charges are computed independently from service and session related ones [6]. The development of innovative applications marketed in the context of multi-party business arrangements is expected to further raise the bar on candidate charging & billing architectures by requiring support for advanced and sophisticated features (e.g., hot-billing, reverse charging, multi-party charging, etc). Therefore, the service provision process must interact dynamically with the charging process in order to provide notification mechanisms that communicate to each user the portion of the current – and overall – charges that attributes to each domain (i.e., transport or application) for each particular session.

D. Service and application requirements

Service independence: To achieve mass scale deployment over millions of mobile terminals with disparate technical characteristics, application architectures should promote service independence and adopt the “write once, run anywhere” paradigm. Virtual machines and interpreted languages lend themselves nicely to the nature of mobile devices, as their restricted capabilities may not suffice for a full compilation of a downloaded application. Thus, independent service providers will be relieved from the burden of designing, developing, supporting and maintaining multiple versions of their applications for each possible client architecture – along with all

the related benefits in terms of cost savings and improved application quality.

Dynamic service deployment & management: Typically, service deployment and management is a cumbersome and ad-hoc process imposed with various restrictions regarding the type of supported applications (e.g., transport technologies used, programming languages and external interfaces) and applicable end systems (e.g., desktop computer, etc). Ideally, however, service deployment should be applicable without requiring expert knowledge about the implementation details of each particular service or the exact location where an application will run. Rapid service deployment is one of the most significant requirements emerging together with 4th generation mobile networks.

Independence from network technologies: To be widely applicable and commercially viable, the implementation artifacts that realize or orchestrate the service provision process should bear as few technological dependencies on underlying network architectures as possible. Naturally, there will be some functionality subsets (e.g., QoS provisioning) that will have inherently network dependent aspects and may thus require enhancements with specific network components (e.g., QoS traffic classifiers); but nonetheless, such technological dependencies should be the exception rather than the norm in future mobile communication systems.

Service adaptability: Due to the volatility and unpredictability of the exact context an application will run within, service adaptability mechanisms are particularly useful, especially in the light of the technologically heterogeneity of 4th generation mobile environments. As an additional requirement, services should be able to communicate the quality of service requirements of their traffic flows to the system, either statically (e.g., via a service profile) and/or dynamically (e.g., using an appropriate API).

III. SERVICE PROVISION ASPECTS

A. Value chain impacts

With regard to service provision matters and consequent market developments, existing 3rd generation mobile telecommunication systems constitute a paradigm shift from the inflexible subscription-based offerings of their vertically architected predecessors. The emerging era of mobile communications will be characterized by the participation of multiple players in the value chain, thereby reaping the economical and developmental benefits of competition, namely diversified service offerings and sustainable technological evolution. These new players will typically come in the form of value added service providers, application providers, content providers and content aggregators – to name but a few – that will contribute additional value to the overall service provision process and compete alongside the mobile network operator for the lion’s share of user revenue (Figure 2).

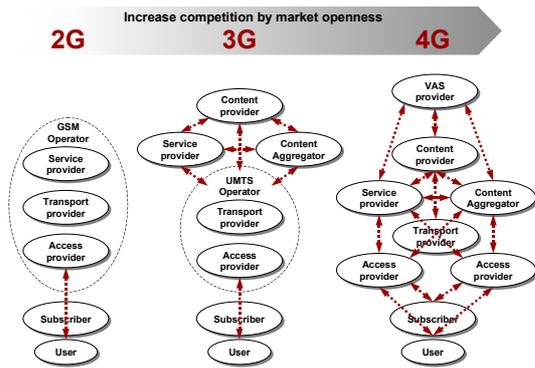


Figure 2. Evolution of the mobile value chain.

B. Analyzing the vision

Over the technologically heterogeneous 4th generation mobile network infrastructure, a plethora of disparate services and multimedia applications will have to be deployed in an efficient yet flexible manner, thus raising the bar on service management requirements [7]. On the other hand, mobile users will expect seamless global roaming across these different wireless access systems and ubiquitous access to personalized applications and content via a universal user-friendly interface.

In studying the implications of the heralded “always best connected” vision of 4th generation systems, one can readily identify the notion of utility, implicitly embedded in the “best” adjective. Utility is a cornerstone concept in microeconomic theory that concerns a – typically – continuous function representation for the consumer’s preference relation over a set of commodities [8]. The following section elaborates on user utility issues.

C. User utility issues

Users consume services and applications to realize various objective and subjective benefits that, perceptually, overbalance the respective charges. Communication-based services and applications depend on the timely and orderly provision of network transport services to realize their functionality. Inasmuch as the network is unable to provide the required levels of service, user-perceived benefits of these applications will remain elusive, thus leading to a degraded user experience.

Communication-based applications make use of transport services to exchange application-specific signaling between remote application entities and to transport various forms of user information (e.g., image, video, data, etc) between communicating endpoints. Overall application performance is dependent on the accommodation of quality of service requirements for their native signaling as well as for the transport of arbitrary user information. From a network viewpoint, these factors translate to traffic flows with different quality of service requirements that will levy different charges – at least in principle – thus decreasing user satisfaction. Consequently, ensuring an adequate performance level for communication-based applications so as to

maximize user satisfaction, requires honoring the quality of service requirements of their traffic flows while minimizing the overall charges incurred, i.e., solving the user’s utility maximization problem. Notably, transport service providers face the dual problem, i.e., maximizing revenue and minimizing network resource usage levels whilst satisfying quality of service requirements for all traffic flows.

Therefore, we claim that the notion of user utility and underlying preferences must necessarily constitute a fundamental building block of 4th generation system architectures, and an issue that should always be accounted for in the respective service provision process.

D. Service provision in future mobile networks

Given the multitude and diversity in the product offerings of the potential value chain participants, the technological complexity of heterogeneous communication environments and the IT illiteracy of a major consumer segment, we understand that most users will be unwilling – or even unable – to engage and coordinate service provision all by themselves. Therefore, in 4th generation systems, some kind of intelligent mediation, as part of the service provision process, will be necessary to efficiently address the utility-related aspects revealed by the “always best connected” vision. Research efforts in intelligent mediation issues highlight the huge potential of service provision platforms for the flexible yet efficient service deployment of value-added services and multimedia applications over heterogeneous network environments. The alternative option is to impose several bilateral customer relationships between the user and all kinds of wireless access network operators he/she wishes to use when accessing available services and applications. However, that significantly complicates service provision matters by mandating the resolution of all technical issues (e.g., deployment mechanisms, activation preconditions, pricing structures, etc) on a per service provider-mobile network operator basis for each particular user – an approach that is obviously non-scalable.

E. Intelligent mediation

The evolving radical developments in the mobile communication market position intelligent mediation as a new role in the value chain, besides the existing roles of network operators and (value added) service providers. That new entrant will maintain the customer relationship with the subscriber and provides him/her with a universal roaming and service access capability whilst accommodating his/her personal preferences, regardless of the access network(s) and terminal equipment in use. Regardless of whether this new role will emerge from existing VHE or MVNO approaches, one of its major tasks will be to provide billing services for its customers by collecting related charging information from other players engaged in the service provision

process, correlating it and issuing a single itemized bill to the customer, thus fulfilling user requirements for one-stop billing. In parallel to the billing process, it will also act as a clearinghouse, realizing accounting procedures that apportion overall revenue between the interested players according to bilateral or multilateral accounting agreements. That would relieve the burden of associating multiple bills with one or more access sessions from the user and significantly improve his/her system experience. Such accounting models have known wide acceptance in the credit card business; thus, we feel they could be successfully applied in the mobile telecommunication sector as well.

The described scenario imposes a paradigm change in terms of the network operator role, as well as of the economics of telecommunications. Although it may seem contentious, or even heretic, the authors see no apparent reason why such business combinations should not prosper under mutually satisfactory reciprocity agreements. Notably enough, other authors have embraced similar far-looking visions of future market developments in the mobile communications sector [9].

F. Reconfigurability

Alongside this line of reasoning, reconfigurability is a radical concept that has been increasingly drawing the interest of the academic and industrial research community over the last decade, as a critical technological enabler for the seamless integration of diverse technologies into a coherent whole. Reconfigurability concerns the dynamic arrangement of functional entities in the communication and computing infrastructure that supports service provision for the purpose of improving and enriching the overall user's experience by enabling emerging services and substantially enhancing existing ones.

IV. 4TH GENERATION ENABLING TECHNOLOGIES

A. Adaptable & intelligent service provision

Knowledge of mobile terminal capabilities will be essential for an efficient service provision process, so that the user is consistently presented with a list of services that can be supported by the mobile device he/she is currently using for service access. Given that wireless access systems differ in their strengths and limitations in terms of coverage area and supported bandwidth, mobile terminal and access network capabilities (e.g., traffic load, availability of performance enhancing proxies and content transcoding facilities) should also be considered along with other important flavors of context information (e.g., user location, mobility patterns) for the efficient adaptation of the service provision process to each particular environment and situation. The implementation of such capability negotiation schemes requires flexible representation formats and announcement procedures along with sophisticated user profiling mechanisms.

Application developers and value-added service providers will benefit from such a capability negotiation mechanism, since they will be relieved from the burden of accommodating the technical details for each possible target system architecture in their application versioning scheme. Mobile network operators will also benefit through the resulting economies in the scarce and expensive spectrum bandwidth, thereby facilitating service for a larger number of mobile terminals per radio coverage area.

B. Transparent mobility and universal roaming

The variety of wireless access technologies that will coexist in the 4th generation network environment raises the bar on the technical as well as the regulatory aspects of roaming. Seamless user mobility across various wireless access technologies (e.g., WLAN, UMTS) with minimal or zero user intervention must be supported by efficient inter-system mobility management and handover procedures. To reduce signaling load, micro-mobility should be handled by the specific mobility management mechanisms of each wireless access network, while macro-mobility and roaming should be built on standard industry protocols (e.g., Mobile IPv6) and architectures (e.g., AAA).

Important research issues that require further investigation are generalized inter-system signaling and handover procedures and incorporation of information from higher layers in mobility management procedures (e.g., user preferences, quality of service requirements of traffic flows, pricing information). These requirements arise from the need to consider pricing information alongside with quality of service levels in the decision stages of the service provision process so as to solve the user's utility maximization problem.

C. Automated reconfiguration mechanisms

In a 4th generation environment, the plethora of available – but disparate – applications combined with the existence of multiple wireless access systems, and the need for a user profile-driven decision process, suggests that there may be multiple alternative configurations for the accommodation of the same set of services. For instance, transporting a media stream could be accomplished by means of either a wireless LAN or a UMTS bearer service – provided that the mobile terminal is within the service area of these systems. However, the decision regarding which particular system to use depends on a number of factors, such as the respective cost of service, availability of network resources, radio link quality and user preferences. Taking into account that using different systems will result in accruing different charges, user pricing preferences should be allowed to bias such decisions in 4th generation mobile communication networks.

Therefore, the respective system and protocol configuration procedures should be dynamic and automated to the highest degree possible so as to

facilitate higher-layer control over transport services. Currently, policy-based management presents the most promising approach to accomplish the aforementioned tasks. Policy-based management demarcates between enforcer entities and decision entities in the infrastructure, thereby allowing the realization of a flexible management architecture that spans across multiple administrative domains. Policy protocols support an outsourcing and a provisioning model of operation, and supported by standard information models that record the features and mechanisms of the infrastructure in a unified manner, policy-based management possesses great potential for managing the entire service provision process in heterogeneous network environments.

D. Interoperable inter-domain QoS management

Currently, the prevalent QoS models for the IP protocol are Integrated Services (IntServ) [10] and Differentiated Services (DiffServ) [11]. Despite the differences in the scope of their architectural assumptions, control model and tradeoffs between accuracy and scalability, IntServ and DiffServ share a common subset of functionality, e.g., the classification elements in the user traffic forwarding plane. Consequently, it is possible to regard the components of the aforementioned QoS architectures as instances of a generic information model for network elements capable of providing QoS-aware treatment to forwarded IP packets. That, in turn, would constitute part of a generic information model for an entire network infrastructure that provides an aggregate IP forwarding service.

Combined with policy-based management approaches, information models provide a universal view of network element functionality and facilitate its dynamic configuration regardless of the particular technologies on which that functionality is built upon. Provided an open API (e.g., IDL) that exposes the quality of service mechanisms of the network infrastructure in a technologically opaque fashion and allows trusted third parties to exercise application-level control over the underlying network services is available, the realization of distributed inter-domain QoS management is greatly simplified.

E. Flexible pricing and billing mechanisms

The clear demarcation between the network and the service/application domains that has been architected in existing 3rd generation systems suggests that any future-proof charging and billing architecture should portrait similar – if not greater – flexibility in its design. As a minimum requirement, network-related pricing models and policies should be completely independent from service-related ones, with regard to both formulation and application matters. Once again, information models [12] and policy-based management approaches can contribute greatly to the interoperable specification of pricing models for specific domains (e.g., volume-based or QoS-based charging for the network domain) and to

the configuration of appropriate components in the network infrastructure (Figure 3).

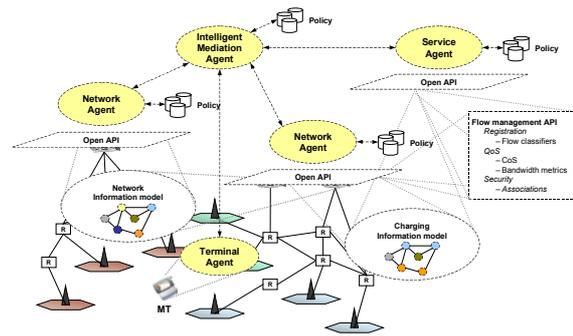


Figure 3. Architectural abstractions 4th generation mobile network.

In addition, the configuration and dynamic adaptation of the charging process to the business model and specific user-service combination requires a flexible and configurable charging infrastructure. The utilization of standardized policies for the expression of conditions and related configuration actions is a critical issue, since it allows the configuration of heterogeneous infrastructures in a universally applicable manner.

F. Mobile execution environment architectures

The execution environment at the mobile terminal should shield applications from mobility-induced events (e.g., change of IP address) in the underlying protocol stacks while enabling a rich interaction between applications and the network infrastructure to take place. Therefore, any API that exports the transport services of the execution environment to the applications should avoid using network or link related information (e.g., IP address, port numbers) in its class and method definitions. Such information should be managed internally by the execution environment in order to shield applications from undesirable network-related events (e.g., loss of transport socket connection). These events should be handled by the execution environment that will send an appropriate notification to concerned applications, allowing for their graceful termination, if necessary.

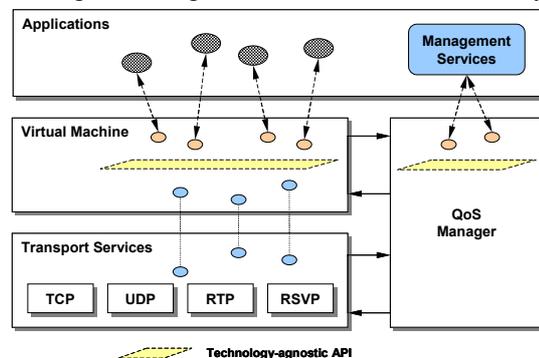


Figure 4. Architectural components of fourth generation mobile execution environments.

In addition, the mobile terminal must provide a local QoS Manager that interacts with the active applications, accepting requests for the allocation of

resources for their traffic flows and forwarding them to the intelligent mediation agent if the available resources at the mobile terminal suffice. Thus, the QoS signaling procedure should proceed in a two-stage admission control – one at the mobile terminal (since it is a restricted device in terms of processing capabilities and power resources), and one at the intelligent mediation agent that will aggregate the admission control decisions from the respective wireless access networks. Undoubtedly, there will be cases where the mobile terminal rather than the wireless interface is in scarcity of resources (e.g., drained power supply), so this approach can potentially minimize unnecessary signaling over the radio interface and preserve valuable bandwidth.

V. CONCLUSIONS

Advances in mobile communication technologies have been rapid and their effects have frequently manifested themselves in ways and places far beyond the ones imagined by their inventors. With regard to 4th generation mobile communication systems, it is paramount to grasp an understanding on important market developments as well as critical enabling technologies before proceeding to the architecture specification phase. Hierarchical mobility management, policy-based management, flexible pricing and billing approaches, capability negotiation frameworks, information model concepts, reconfigurability technologies and open, technology-independent API frameworks are all fundamental building blocks of 4th generation mobile network architectures.

REFERENCES

- [1] J. Pereira, "Beyond third generation", Wireless Personal Mobile Communications (WPMC) 1999, September 22, 1999, Amsterdam, The Netherlands.
- [2] J. Pereira, "Fourth generation – Now it is personal", 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 18 – 21 September 2000, London, United Kingdom.
- [3] J. Pereira, "Fourth generation – Beyond the hype, a new paradigm", IEE 3G Mobile Communication Technologies, March 28, 2001, London, United Kingdom.
- [4] M. Annoni, R. Hancock, T. Paila, E. Scarrone, R. Toenjes, L. Dell'Uomo, D. Wisely, "Radio access networks beyond 3rd generation: A first comparison of architectures of 4 IST projects", Mobile Communications Summit 2001, 10 – 12 September 2001, Barcelona, Spain.
- [5] N. Alonistioti, V. Gazis, N. Houssos, S. Panagiotakis, "An Application platform for downloadable VASs provision to mobile users", IST Mobile Communication Summit 2000, 1 – 4 October 2000, Galway, Ireland.
- [6] V. Gazis, M. Koutsopoulou, C. Farmakis, A. Kaloxylos., "A flexible charging and billing approach for the emerging UMTS network operator role", Applied Telecommunication Symposium, 22 – 26 April 2001, Seattle, Washington, United States.
- [7] N. Houssos, V. Gazis, S. Panagiotakis, S. Gessler, A. Schuelke, S. Quesnel, "Value added service management in 3G networks", Network Operations and Management Symposium (NOMS) 2001, 15 – 19 April 2002, Florence, Italy.
- [8] A. Mas-Collel, M. D. Whinston and J. R. Green, "Microeconomic Theory", Oxford University Press, ISBN 0-19-510268-1.
- [9] H. Luediger, S. Zeisberg, "User and business perspectives on an open mobile access standard", IEEE Communications Magazine, September 2000, pp. 160 – 163.
- [10] RFC 1633, "Integrated Services in the Internet Architecture: an Overview".
- [11] RFC 2475, "An Architecture for Differentiated Services".
- [12] F. Hartanto, G. Carle, "Policy-based billing architecture for Internet Differentiated Services", IFIP Fifth International Conference on Broadband Communications, Hong-Kong, 10 – 12 November 1999.