Network economy in service- and context-aware next generation mobile networks

Albayrak, Sahin; Bür, Kaan; Toker, Ahmet Cihat

Published in:
Proceedings of PIMRC - IEEE International Symposium on Personal, Indoor and Mobile Radio Communications

DOI:
10.1109/PIMRC.2007.4394534

2007

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Albayrak, S., Bür, K., & Toker, A. C. (2007). Network economy in service- and context-aware next generation mobile networks. In *Proceedings of PIMRC - IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* IEEE - Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/PIMRC.2007.4394534

Total number of authors:
3

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

NETWORK ECONOMY IN SERVICE- AND CONTEXT-AWARE NEXT GENERATION MOBILE NETWORKS

Sahin Albayrak, Kaan Bür, Ahmet Cihat Toker
DAI-Labor, Electrical Engineering and Computer Science Department
Technische Universität,10587 Berlin, Germany

ABSTRACT

The Internet, mobile telephony and portable computers let today's user meet novel applications provided by the service developers, who are motivated by the popular customer trends in technology. However, the supply-demand chain observed in telecommunications causes the complexity of both the user devices as well as the network to increase. The complexity is caused by the need to support quality of service, security and mobility for the new services throughout the heterogeneous communication network. These problems slow down the application development process significantly and cause the user demand not to be met in time. We believe that a new look at the network is necessary in order to address these issues. This article presents this new look, which we term network economy. It is a service-centric approach to next generation mobile networks, whereby the network is redefined as a collection of diverse resources. The aim of network economy is to optimise the allocation of these resources such that network efficiency is maximised. In order to achieve this objective, the intelligence level of the network is increased by decision engines, which continuously collect information from the network, classify and store it in a repository to be queried by their peers. The approach is accompanied by context-aware, self-organising user devices. Network economy provides a network-wide infrastructure to benefit users, service developers as well as network operators.

I. INTRODUCTION

The ever-growing popularity of the Internet and the rapid penetration of mobile telephony have shown us three important motivations for the citizens of today's society. Regarding the choices they make when they confront with novel technology, the contemporary citizens tend to opt for information, communication and mobility. The commercial success of portable computers proves that people want information to be accessible at all times. They also want to communicate with their friends, sometimes even with total strangers and a variety of information or entertainment services. Thus, it can be argued that recent developments in microelectronics and communication technologies stimulated a conceptual shift in their expectations from voice towards multimedia, from availability towards acceptable quality and from stand-alone towards group-oriented computing.

The conceptual shift explained above has a significant impact on today's networks in terms of the need for quality of service (QoS), which defines a guarantee given by the network to satisfy a set of service performance constraints for the user in terms of end-to-end delay, jitter, available bandwidth, and packet loss probability [1]. QoS support for multimedia applications is closely related to resource allocation, the objective of which is to decide how to reserve resources such that QoS requirements of all the applications can be satisfied [2]. However, it is a significant technical challenge to provide reliable, high-speed, QoS-incorporated end-to-end communications in heterogeneous next generation networks (NGN), due to a multiplicity of access technologies and the complexity of network management [3]. Thus, it is essential to utilise resources effectively in this environment.

One of the major concerns for mobile communications is the ability of the routing infrastructure to cope with the dynamics of node mobility. In order to maintain connectivity and support QoS with maximum possible accuracy and minimum overhead, adaptation to mobility is essential. Thus, the network should be able to draw conclusions from the mobility pattern of the user, make predictions regarding the destination and adapt its resource reservation policy. On the other hand, the need for security stems from being subject to a variety of malicious attacks as a result of being connected to the rest of the world. Many aspects of security have been addressed in the research literature until today. Nevertheless, the heterogeneity of NGN may result in inefficiency or even allow security breaches. Therefore, a deeper look at the potential interoperability and integration issues related to these aspects is necessary.

From the perspective of its different actors, the network should ideally be the underlying mechanism that provides the necessary functionality to achieve their respective goals. As explained above, this functionality has to contain QoS, security and mobility. The goal of the network operator is to accelerate the implementation of new technologies such that the throughput is maximised and the necessary management effort to maintain the network is minimised. The service provider, on the other hand, makes use of the state-of-the-art technology in order to minimise the development cycle of new applications such that added-value services can be deployed as quickly as possible. Finally, the user wishes to get the service that suits her needs for an acceptable price. It is the job of the network infrastructure to provide these actors with the means necessary to achieve these goals, which can only be done with new and intelligent node architectures.

As the number of beyond-third-generation (B3G) services spanning a heterogeneous network with wired, fixed-point wireless and mobile ad hoc domains increases, adaptation to mobility, self-organisation and self-management aspects become even more critical in the face of the interoperability issues expected to arise among various access technologies. *Autonomous communications* is a new term summarising the desired features of these networks. It should be noted that incremental changes on today's networks cannot efficiently address the critical issues mentioned above. As a well-known example, the original design assumptions and decisions of the Internet did not consider issues such as QoS, group-oriented

1-4244-1144-0/07/$25.00 ©2007 IEEE.

computing and adaptation to mobility [4]. These issues are still open as future research for the further development of the underlying protocols and network technologies. Moreover, enhancements on existing technologies such as the Internet protocol (IP) multimedia subsystem (IMS) increase the complexity of the network even further.

Yet, the near future seems to hold more challenges for the Internet architecture. B3G services expect the network to provide open interfaces that enable new software technologies and allow innovative, added-value services to be created, deployed and managed. Although the Internet architecture is flexible enough to allow new applications to be developed that can be easily adapted to the existing network, this flexibility is still limited and cannot easily be extended to the support of QoS, security and mobility. The success of NGN depends on new service platform technologies that increase both the availability as well as the number of added-value services in the network. This, in turn, can only be realised if service discovery, QoS, security and mobility are supported by the network. Thus, a more intelligent network is essential.

The aim of this article is to introduce and define a novel concept, network economy, which proposes enhancements to the well-known layered network architecture in order to address the challenges mentioned above. Our ultimate goal is to increase the speed of evolution for the NGN through the realisation of an intelligent communication infrastructure. Network economy sees the network as a collection of diverse resources and has the objective to optimise the allocation of these resources to users and services such that the overall network efficiency is maximised. To this end, the node intelligence is increased through the use of an information interpretation entity. Working independently within this entity, intelligent agents continuously sniff the control data arriving at each node, classify and store it in an information repository. Collectively, the intelligent agents form a learning engine, which draws conclusions from the notifications of a decision engine and uses them for further data processing. Decisions of the information interpretation entity yield to improvements in fault-tolerance, robustness, self-healing and, ultimately, the maximisation of network efficiency.

The rest of this article is organised as follows. Section II recalls the Internet architecture and summarises some recent research efforts towards new design alternatives. Section III proposes a context-aware, intelligent node architecture to achieve the goals of network economy. Section IV demonstrates a QoS-aware service access scenario based on this architecture. Section V concludes the article and gives future research directions.

II. TOWARDS THE NETWORK OF THE NEXT GENERATION

Since network economy addresses the requirements of NGN, and the Internet, as the ever-growing IP network of networks, has the function of a core network connecting all the others and this trend is likely to continue for some more time, it is necessary to start this section with a brief review of the design philosophy of the Internet. Then, some of the new generation design issues are investigated.

A. *Internet*

Clark summarises the motivations behind the design decisions that shaped the Internet with one fundamental and several second-level goals [4]. The fundamental goal is connectivity, i.e. to develop a technique for interconnecting the existing networks. Thus, networks are interconnected by Internet packet switches called gateways. The second level goals are survivability, support for multiple types of services and accommodation of a variety of networks. In order to achieve survivability, transient failures are masked by the lower layers, whereas some state information is kept by the upper ones, i.e. only at the endpoints of the network, which actually utilise the service.

B. *Mobility and Heterogeneity*

As the mobility of the users, the heterogeneity of the network and the number of interactive and multimedia applications increase, performance requirements on delay and packet loss more stringent than those of the Internet need to be met. As an attempt to overcome the limitations of the Internet, the peer-to-peer (P2P) technology is implemented, which enables the construction of overlay networks to conduct tasks without support from the infrastructure [5]. Nevertheless, the functions such as session continuity and seamless handover cannot currently be realised by P2P. Thus, P2P services have to converge into the managed services of NGN. Specifically, the handover process can introduce additional delay due to network discovery, configuration, authentication and authorization. In an attempt to reduce delay during these steps, the IEEE 802.21 group proposes a framework, which defines a media-independent handover function (MIHF) that helps mobile devices to experience seamless handover across heterogeneous networks [6]. MIHF provides abstract services to higher layers by means of an interface that exposes service primitives. An event service reports triggered events corresponding to link characteristics; a command service manages link behaviour related to handovers and mobility, and another service passes information on neighbouring networks and their capabilities.

Motivated by NGN complexity and heterogeneity, the idea of a cognitive network, which is capable of perceiving current network conditions, planning, learning and acting according to end-to-end goals, is suggested [7]. As an implementation of this idea, a three-layer framework is derived. The layers from top to bottom represent system behaviour, cognitive process and physical control, whereas the cognitive process operates in a manner similar to a collection of software agents.

III. THE NETWORK ECONOMY ARCHITECTURE

In order to address the issues states in the introduction, the network economy node architecture extracts the information for QoS, mobility and security support from the underlying network layers and converts it to the required functionality. The important aspects of service delivery are thus represented in an abstract manner, whereby the technical complexity is hidden from the user, service provider and network operator. The details of the proposed architecture are explained below.

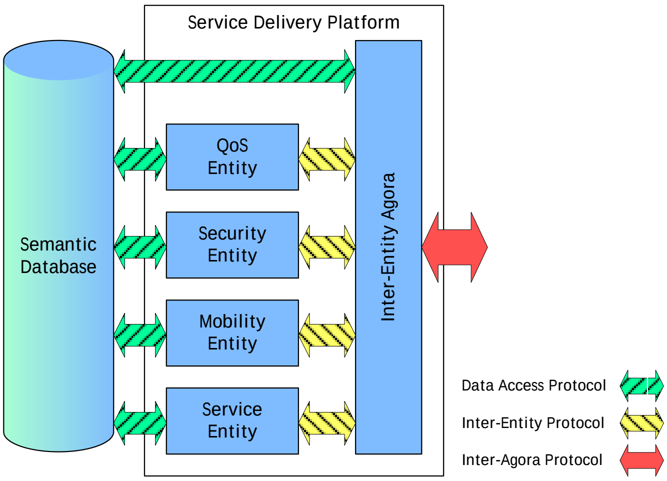

Figure 1: Node architecture with service delivery functionality.

A. *Service Delivery Platform*

In order to satisfy the requirements of a variety of services presented to the user of the near future, the layered node architecture of today has to be redesigned in such a way that it allows a node to organise itself according to the semantic and contextual information it gathers from the network. Although this information has been floating around in the network for quite a long time, the current architecture discourages its exchange between protocol layers [7]. Clearly, the raw network control data is not exactly what is meant by semantic and contextual information. However, it is necessary to access this raw data in order to compile, derive and interpret the results which will be aggregated to generate the semantic and contextual information. Thus, the new node architecture, should allow information to be shared between protocol layers, stored in a common semantic information database, and interpreted by a smart decision engine to be provided to the requesting service in a universal format.

Figure 1 shows the main logical components of the service delivery platform and their relations with the semantic information database mentioned above. According to the node architecture we propose, the service delivery platform consists of several functional blocks, also called *information interpretation entities*, each of which is responsible for one specific aspect of the network support that should be provided to the services at the application layer. Such an entity is capable of collecting information from the network at several layers with the help of its *intelligent engines*, which are explained in the following section. It shares this information with the other *local* entities by means of an inter-entity agora coordinating the communication between them. The entities also store information to their common semantic database to be used by the entities of other nodes in the network. For a *foreign* entity to access this information and submit queries to the semantic database of another node, an inter-agora connection needs to be established. Thus, two nodes can share semantic information only via their inter-entity agorae.

Each information interpretation entity fulfils a specific function of the service delivery platform in the node, such as QoS, mobility, security or other networking tasks, by making use of the relevant information within and outside the node. We use the term *information element* for the relevant local information that can be obtained from the network interface, the operating system and the applications running on the node. An example for a security information element is the security state of a certain IPSec tunnel, whereas the current velocity, location and Mobile IPv6 Care-of-Address of a device are examples of mobility information elements. The information elements are defined and grouped together according to their relevance for the information interpretation entities. Each entity is responsible for extracting its respective information elements in their native formats, interpreting them and converting them into the semantic database format, and storing them into the database. Information interpretation entities make use of the interpreted semantic versions of the local and remote information elements in their decisions later. The inner structure of the information interpretation entities is explained in the next section.

B. *Semantic Communication Space*

The new node architecture presented above decomposes a rather large task into smaller components. The major task is the creation of a semantic communication space, or a semantic web within the network, such that pieces of information can be aggregated and serve a holistic approach to efficient network planning. The parts of this task are semantic representation, query manipulation, classification and cognition. These parts are shared and realised by the intelligent engines composing the inner structure of an information interpretation entity as shown in Figure 2.

The major task stated above requires the interception and translation of information by the intelligent engines according to the needs of different information interpretation entities. For instance, the *learning engine* is responsible for the cognition and classification of information. It is the interface to the control data flowing through the network or stored in the node and has the ability to distinguish the usable information from other data with the help of a set of predefined rules. It extracts the useful information and delivers it to the semantic translation engine. It also uses the notifications of the central decision engine as a guide to generate new learning rules or apply changes to old ones.

Each learning engine has *learning agents* at those layers, in which the entity owning the engine has defined information elements to be extracted. These learning agents have different methods for extracting the information elements. They sniff network packets at various levels, read operating system variables, interact with device drivers, or use application programming interfaces running on the node. The learning engine is responsible for deriving the semantic information out of the raw data coming from its learning agents. It coordinates its learning agents such that the information elements are updated in a manner requested by the central decision engine; which can demand polling, event- or subscription-based update of information elements from the learning engine.

The *semantic translation engine* converts the interpreted information received from the learning engine into a semantic representation. It runs the data access protocol and stores this

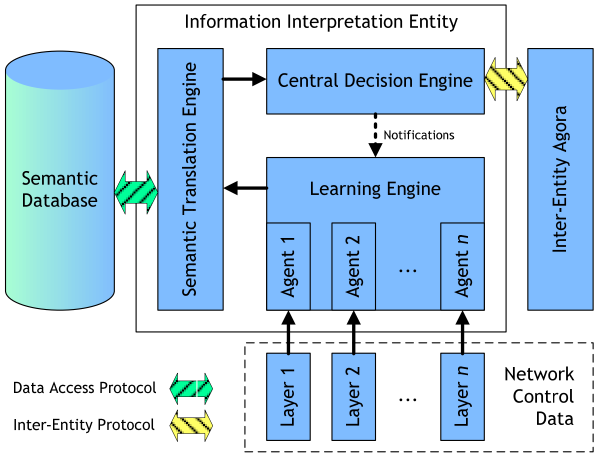

Figure 2: Structural details of an information interpretation entity.

information in the form of records into the semantic database. It also reads records from the database upon request from the central decision engine, which uses the semantic information and the inter-entity protocol to access the relevant information elements and make its decisions.

Finally, the *central decision engine* is responsible for the overall efficiency of the network with regard to the service- and application-level goals of the node. It also preserves the integrity of the information interpretation entity. It is this module that draws conclusions from the results of the database queries it runs and the status of the network. It also manipulates queries based on these conclusions and notifies the learning engine of the changes. The central decision engine is the core of the information interpretation entity, which forms a closed-loop control system through its feedback, thus supporting the other units and ensuring that each of them works in an adaptive and self-learning fashion and their actions do not contradict each other.

C. *Context-Awareness*

The node architecture of network economy is accompanied by context-aware, self-organising user devices in order to enhance the quality of the decisions made by the information interpretation entities. Context is generally defined by user preferences, current situation and device capabilities [8]. Thus, context-awareness is the ability of an intelligent device to provide its user with the most convenient presentation of the acquired service. This requires a certain amount of information to be gathered, either proactively before any service is selected or reactively upon the development of a situation, by the device such that necessary intelligent decisions can be made by the device in advance. As can be expected, context-awareness begins with the awareness of the device of its own properties, capabilities, limitations and restrictions. The intelligent device has to be able to make the necessary adaptation according to these such that the service can be presented to the user in the right format and under acceptable conditions.

At a more abstract level, context can be redefined as the set of circumstances and facts that surround a particular event. In the case of direct communication, this set of information is not necessarily exchanged between the peers explicitly but is known to those involved. It keeps the flow of conversation in the direction that is desired by both parties in a natural yet intelligent manner. Network economy extends this notion to computer communications which involves both humans and computers, where each node in the network is a producer and a consumer of context. Each node produces this information, as according to network economy it is self-aware. It also makes a subset of its context available to other nodes through the semantic database and the inter-agora protocol. Each node also uses this information as it builds a partial knowledge base about other nodes in direct or indirect communication. This is necessary in order to use the communication channels more efficient by understanding the context and to make more efficient local decisions using global inputs. The proposed node architecture allows the collection and utilisation of local and remote context information.

Location is a typical example for the information that can be used for context-awareness. The availability of the location data gives the portable device additional information to be used with the current service. Another important piece of information is the surroundings of the device, which basically consists of the personal area network (PAN) of its user. The user's personal intelligent equipment can be detected by an appropriate communication technology such as Bluetooth. The surroundings can also be extended to other publicly available devices in the vicinity of the user. Some of the less addressed but important components of context-awareness are the situation that the user is in and the intention of the user. If the portable device is intelligent enough to detect the former and anticipate the latter, or at least learn them from previous experience, a significant level of distinction can be achieved both in the search for and in the presentation of the service.

IV. SERVICE ACCESS WITH MOBILE QOS PROFILES

In this section we illustrate how the proposed architecture benefits the user, the service provider as well as the network operator with a service access and adaptation scenario in an all-IP network. We refer to the heterogeneous network in Figure 3, like the one described in the introduction, and show how our approach increases service accessibility in such a network. In our scenario, the user wants to access a multimedia service via a mobile terminal (MT) while on the move and initiates a session by using an appropriate session layer protocol. The service provider's node (SPN) is able to send the multimedia service in different formats, each with different bandwidth requirements. Bandwidth is a precious resource for the service provider, since it is leased from the network operator, and the goal is to resell this bandwidth in an efficient manner among as many users as possible. It is a waste of resources to send the high-bandwidth version of the multimedia service when the connections between MT and the access points (AP) cannot support this bandwidth. Thus, in order to choose the most efficient format, the service entity asks the QoS entity for the amount of available bandwidth.

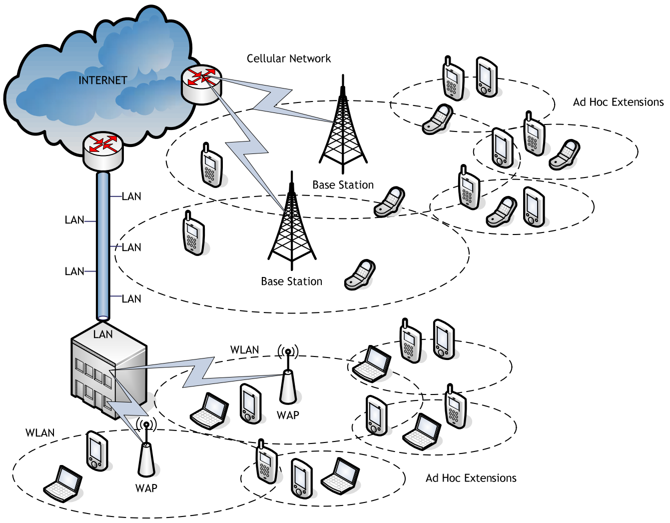

Figure 3: The heterogeneous all-IP network of the near future.

The QoS entity contacts its peer on the network operator node (NON) via the inter-agora protocol to inquire the capabilities of MT and of APs that are within the transmission range of the user. When the peer QoS entity receives this request, it turns to its semantic database where the mobility pattern of the user is stored. The mobility pattern is deduced from the location and velocity of MT as well as the locations of the visited APs. This information and the bandwidth that is available on these APs give yield a bandwidth range that the user can be supported with, which is returned to SPN as the user’s mobile QoS profile. Upon receiving the requested profile, the QoS entity of SPN makes an intelligent decision on the amount of bandwidth that has to be allocated to the user without exceeding the capabilities described in the mobile QoS profile and informs the service entity such that it can select the appropriate format. The mobility entity of SPN also registers itself at its peer on NON to be informed on the changes in the semantic database regarding the mobility state of MT so that SPN can adapt the multimedia service to the changing conditions of the connection. The QoS entity of SPN informs its peer at NON on the allocated bandwidth and also registers for changes in the QoS state so that the multimedia service can be modified according to the changes in the network load administered by the peer QoS entity.

With the proposed framework, a cross-layer distributed service access and adaptation scheme is realised as a response to the mobility of the user and the dynamics of network conditions without the need for defining separate protocols for each interface. This scenario ensures that the user enjoys the multimedia service on the move while the service provider and the network owner make the most efficient use of their respective resources.

V. CONCLUSIONS

This article presents a new concept, network economy, which defines an enhanced architecture in order to meet the essential requirements of NGN in a well-organised, integrated manner. The architecture allows network layers to interact with each other and exchange semantic and contextual information to support applications and services within a unified framework across the layers. It is a service-centric approach in the sense that as a result of the information shared among nodes, design and implementation decisions made independently by the applications do not violate but help support the overall efficiency of the network. This way, new applications and services can also be implemented much easier than today and the ones not anticipated yet can still be supported in the future.

There are a number of interesting research topics related to the development of the new network architecture we propose. The most important issue is the design of the individual components composing the information interpretation entities. These components confront different levels of data abstraction at different layers. They also have to be specially designed according to the requirements and functions of their respective layers. Their task includes the classification of the control data flowing in the network such that useful information can be extracted, annotated, stored and later queried with regard to certain criteria. However, it is not easy to process this data and make it available for the interpretation and use of different layers with QoS, security or mobility purposes. Conventional database management systems cannot provide the flexibility required to handle the vast amount of data generated by the network. Instead, an intelligent approach is needed to convert the raw control data to a semantic representation, make decisions based on the semantics and learn from experience. Nevertheless, network economy is a promising novel concept, which can pave the way for the efficiency of service-centric, context-aware communication networks of the near future.

REFERENCES

[1] Chakrabarti, S. and A. Mishra, “QoS Issues in Ad Hoc Wireless Networks”, *IEEE Communications Magazine*, Vol. 39, No. 2, pp. 142-148, February 2001.

[2] Zhang, Q., W. Zhu, G.J. Wang and Y.Q. Zhang, “Resource Allocation with Adaptive QoS for Multimedia Transmission over W-CDMA Channels”, *Proceedings of WCNC 2000*, pp. 179-184, Chicago, USA, 23-28 September 2000.

[3] Cavalcanti, D., D. Agrawal, C. Cordeiro, B. Xie and A. Kumar, “Issues in Integrating Cellular Networks, WLANs, and MANETs: A Futuristic Heterogeneous Wireless Network”, *IEEE Wireless Communications*, Vol. 12, No. 3, pp. 30-41, June 2005.

[4] Clark, D.D., “The Design Philosophy of the DARPA Internet Protocols”, *Proceedings of SIGCOMM’88*, pp. 106-114, Stanford, USA, 16-19 August 1988.

[5] Chen, W.-P., T. Hamada, J.J. Yao and H.-Y. Wei, “QoS Management and Peer-to-Peer Mobility in Fixed-Mobile Convergence”, *FUJITSU Scientific & Technical Journal*, Vol. 42, No. 4, pp. 535-546, October 2006.

[6] Dutta, A., S. Das, D. Famolari, Y. Ohba, K. Taniuchi, T. Kodama and H. Schulzrinne, “Seamless Handover across Heterogeneous Networks – An IEEE 802.21 Centric Approach”, *Proceedings of IEEE WPMC’05*, Aalborg, Denmark, 18-22 September 2005.

[7] Thomas, R.W., D.H. Friend, L.A. DaSilva and A.B. MacKenzie, “Cognitive Networks: Adaptation and Learning to Achieve End-to-End Performance Objectives”, *IEEE Communications Magazine*, Vol. 44, No. 12, pp. 51-57, December 2006.

[8] Yu, Z., D. Zhang and C.-Y. Chin, “An OSGi-Based Infrastructure for Context-Aware Multimedia Services”, *IEEE Communications Magazine*, Vol. 44, No. 10, pp. 136-142, October 2006.