(1) Next-generation Service Platform and Middleware Technologies

Wolfgang Kellerer, Chie Noda, Marco Sgroi, Anthony Tarlano, Srisakul Thakolsri, Matthias Wagner and Hendrik Berndt

DoCoMo Euro-Labs is researching advanced service platform capabilities such as personalization, context awareness, and service mobility to provide user-centric services, and a middleware platform including mobile middleware and cross-layer design to support a ubiquitous environment.

1. Introduction

The provision of user-centric services is becoming indispensable as network functions become ever more complicated and the variety of services increases. To provide users with personalized services that can be used anytime and anyplace, it must be possible to provide seamless services across heterogeneous access networks and to optimize Quality of Service (QoS) for users on the service platform. This calls for service-platform functions that can recognize the user's current environment and circumstances and provide appropriate services based on user preferences. It also calls for functions that can transfer and synchronize service information to enable the provision of seamless services across heterogeneous access networks and different mobile devices. There is also a need here for open interfaces that can abstract network-specific parameters to facilitate the provision of easy-to-use services and expand the set of applications offered by service providers.

Still another issue in achieving future mobile communication networks is the construction of a ubiquitous computing environment that will include the use of smart devices (high-performance, compact mobile devices such as future smart cards and sensors). Establishing compatible connections between these devices will be an important role of middleware.

This article describes our research results to date in the two fields of "advanced service platform capabilities" and "user-centric middleware platform" and present our plans for future research in these areas.

2. Advanced Service Platform Capabilities

This chapter describes a framework for providing context-aware services, a mechanism for dynamic service discovery and selection to provide personalized services based on user preferences, and technology for exchanging profiles and other service information [1] [2].

2.1 Provision of Context-aware Services

The research of context awareness is closely related to the field of ubiquitous computing [3]. In particular, the following issues come into play when an application determines and adapts its operation in real time on the basis of context information[*1] related to user environment (location, temperature, etc.) and user circumstances (working, driving, etc.).

- Recognizing and collecting context using sensors
- Distributing and selecting collected context information
- Adapting the application to selected context information

An important point to consider here is that the number and type of sensors that might be available for use cannot be estimated during the development of a context-aware application. There is therefore a need for a context-aware-application framework that allows the application to adapt in a scalable fashion to dynamic changes in an environment using sensors (such as increase in the number of sensors and addition of new types of sensors).

The framework proposed by DoCoMo Euro-Labs includes a conceptual model of context spaces, an information-sharing protocol called Application Message Interface (AMI), and dynamic application operation based on context acquisition as shown in **Figure 1** [4].

The context-spaces model describes an information system based on hierarchically distributed spaces. In this system, sensors represented as space peers make use of a context space to share and store context. A context space is a virtual meeting point for space peers belonging to a peer group. That is to say, multiple peers related to a common context form a peer group and exist in a common context space. A context space is addressable and capable of changing dynamically. Here, an AMI, which can access context continually, helps resolve the standard problems of scalability and usability in a dynamically

*1 Context information: Any kind of information that can be used to characterize the state of an entity. An entity may be a person, place, or object relating to the human-computer interaction.

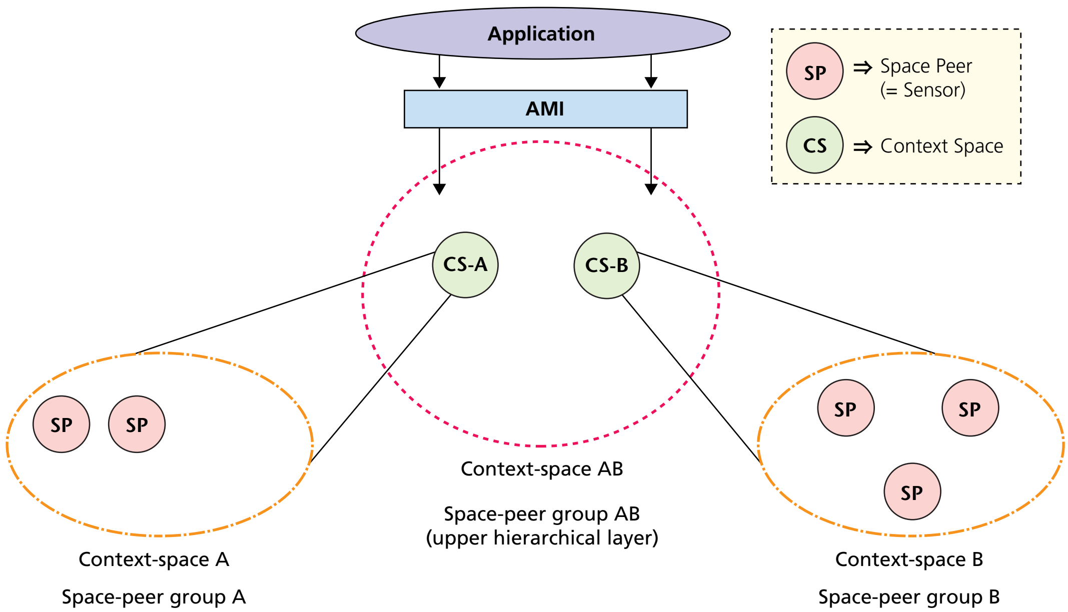

Figure 1 Framework for context-aware applications

changing environment.

To support communication between space peers, i.e., sensors, it must be kept in mind that sensors may not always be available for use in a mobile environment. Although a server must be available at all times in a server-client model, it is natural to assume that there is no mobile peer that can be accessed at all times without fail. To deal with this situation, we give each space peer both server and client roles by adopting a Peer-to-Peer (P2P) communication model (such as JXTA[*2]). It is important here to assume wireless interfaces between peers and limited bandwidth, which means that the control-signal overhead in existing P2P systems must be reduced and group communications for forming context spaces must be optimized.

In future research, we plan to include space peers capable of autonomous execution (proactiveness) in addition to performing the simple task of providing information in context space. This will make it possible to introduce logic programs in space peers that will in turn enable a variety of functions such as inferring context information before providing it to the application and "pushing" context information autonomously to the application.

2.2 Provision of Personalized Services

In addition to context acquisition and processing, we propose service and user modeling and cooperative service provisioning to achieve personalization based on individual user preferences.

1) Service Modeling

We have been researching knowledge-based modeling and the Semantic Web with a focus on service modeling and subsequent service provision. Generally speaking, the Semantic Web is targeting an extended Web that promotes a formal semantic structure for content on the Internet and rich service descriptions. It aims to describe the meaning of all types of information using machine-recognizable data that will allow that information is automatically processed by services. Semantic-based services of this kind will enable accurate representation of user preferences plus service provisioning adapted to specific circumstances and context through discovery, classification, and personalized selection.

Figure 2 shows an example of service classification in relation to a restaurant reservation service [5]. In this example, the restaurant category of "Asian" includes the more specific category of "Chinese" that itself is subdivided into Cantonese, Sichuan, and Shandong.

The DARPA Agent Markup Language-Service (DAML-S) ontology is considered to be the most mature formal Web-service framework [6]. It is an ontology-based approach to the description of Web services trying to provide a common ontology[*3] for Semantic Web services. As a Semantic Web description language, DAML-S corresponds to the upper layer of the

*2 JXTA™: A platform defining a set of protocols for the functions required by P2P applications. Initially proposed by Sun Microsystems, it is now managed by Project JXTA that makes source code, documentation, and other information publicly available (http://www.aisland.jxta.org/).

*3 Ontology: A theory specifying the relationships between a formal vocabulary and real-world concepts.

DARPA Agent Markup Language plus Ontology Interface Layer (DAML+OIL) [7], which is slated to be standardized as an Ontology Web Language (OWL) by the World Wide Web Consortium (W3C).

2) Personalization

In our work on personalization, we are researching mechanisms for providing users with convenient and easy-to-use services, particularly paying attention to the recent trends in the standardization of profiling languages for use in service delivery. These include Composite Capability/Preference Profiles (CC/PP) at W3C, User Agent Profiles (UAProf) at Open Mobile Alliance (OMA), and Generic User Profile (GUP) at 3rd Generation Partnership Project (3GPP). These profiling languages assume access to services that are independent of mobile device type and therefore adopt a framework based on eXtensible Markup Language (XML) and Resource Description Framework (RDF) [8]. Accordingly, these languages possess a degree of interoperability at the profile-metadata-description level, but in terms of advanced profiling requirements, they are insufficient. In contrast, "cooperative answering" as used in the Semantic Web [9] and database systems is thought to have great potential for the design of future profiling schemes and user-modeling languages.

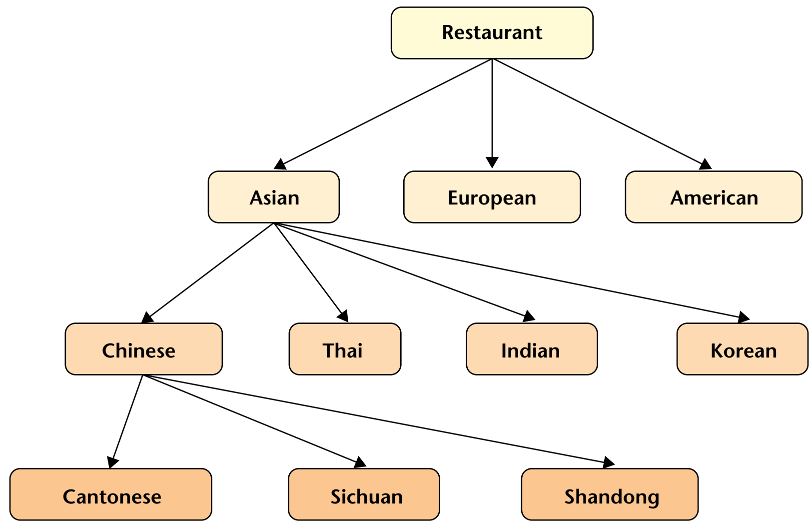

Figure 2 Example of classification in a restaurant reservation service

Figure 3 shows two key concepts behind the technique that we propose for extending user modeling. a) in the figure represents user preferences in the form of a preference order in relation to services and their attributes. In this example, the user prefers a Thai or Indian restaurant over a Sichuan restaurant but has no preference when it comes to Thai or Indian food. b) depicts the concept of usage pattern that can be used to assign default preferences to a certain user group. In this example, a user that likes "hot" or "spicy" food would, in general, prefer to eat at a Thai or Indian restaurant rather than a Sichuan restaurant.

In addition to basic preference orders and usage patterns, semantic-based ontology can also be useful in user profiling and service personalization as described earlier. An appropriate ontology can be used to represent dependencies and mutual relationships between users and user groups. It can also be used to introduce multiple viewpoints with respect to available services.

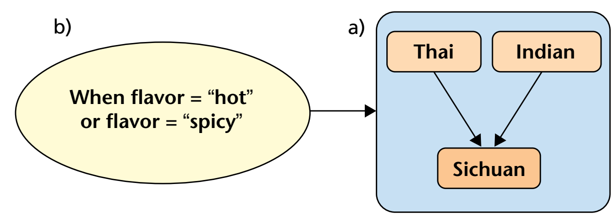

Figure 3 Preference order and usage pattern

3) Cooperative Service Provisioning

Recent efforts in promoting Universal Description, Discovery, and Integration (UDDI), Web Services Description Language (WSDL), and Simple Object Access Protocol (SOAP) are expected to raise the level of Web services. New Web services should significantly raise the potential of Web architecture by stipulating guidelines for service discovery, selection, and synthesis and by providing methods for service automation.

Providing user-centric, personalized services means satisfying individual user needs. Here, user needs can be expressed as complex tasks that can be further divided into simpler subtasks that correspond to individual services [5] [10]. As shown in **Figure 4**, "cooperative service provisioning" means providing the service that meets the user's request most closely even if that service does not completely agree with user preferences. User information (such as a profile that includes user preferences) plays an important role at each step in service provisioning (service discovery, selection, and execution). Reference [10] deals with personalized selection and execution in detail while

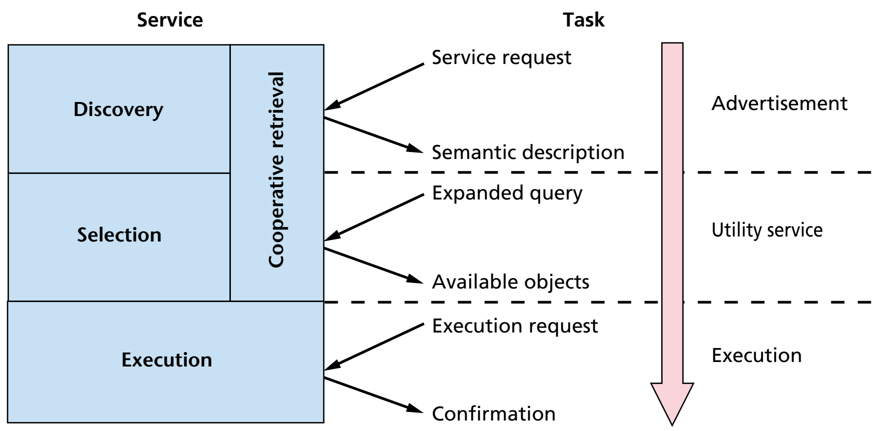

Figure 4 Cooperative service provisioning

reference [5] focuses on personalized service advertising and discovery. An extensive service ontology can make use of user preferences and usage patterns to provide the most appropriate service in response to an individual's request in a particular situation.

In addition, the application of "cooperative answering" will facilitate service discovery and selection that most nearly matches the user's request as described in 2) of this section. In the example of Fig.3 depicting user preferences, a Sichuan restaurant would be proposed if both a Thai restaurant and Indian restaurant were not available. In this way, the user can ease the search conditions associated with the restaurant that the search begins with (Fig.2). In short, user preferences are given priority over common information stored in the service ontology. This type of technique that combines ontology with dominating user preferences can be executed repeatedly in conjunction with a specific service ontology and individual user preferences to modify the initial service request and provide the user with a service of even higher quality. Integrating the above approach with the research of reference [10] makes it possible to achieve personalized, user-centric services.

2.3 Provision of Ubiquitous Services

The provision of ubiquitous services will require functions for transferring and synchronizing service information and currently executing services across heterogeneous environments. In this regard, it must be possible to:

- Provide services to users at anytime and anyplace;
- Provide services that support different mobile devices and heterogeneous access networks;
- Provide services without interruption even if moving or changing access networks; and
- Entrust system operation to a user-defined policy (to eliminate the need for direct exchanges between the user and the system).

1) Application-layer Mobility

The range of mobility that ubiquitous services must adapt to is not defined solely in terms of mobile device and user mobility—it also includes service, profile, and session mobility. Accordingly, to satisfy all of the above requirements, providing ubiquitous-service functions on the application layer is considered more appropriate than doing so on the network layer. The various types of mobility are defined below.

- Mobile device mobility: The mobile device can continue communicating and the session can proceed without interruption if the device's position changes.
- Personal mobility: The user can continue using the same user ID regardless of the mobile device and access point being used.
- Service mobility: The user can use the same service regardless of his position and the mobile device being used.
- Profile mobility: The user can access and synchronize his profiles stored on various mobile devices.
- Session mobility: A session currently in progress can move and be maintained between various mobile devices.

2) Technique

In this research, we use Session Initiation Protocol (SIP) of the Internet Engineering Task Force (IETF) used for session control in IP Multimedia Subsystem (IMS) of the 3GPP. Using SIP as a protocol for controlling the application layer makes it possible to establish, reconfigure, and terminate multimedia sessions.

SIP is flexible in terms of functional extensions [11], and provides mobile device mobility and personal mobility [12]. It must add on functions for real time transfer, profile mobility, auto-detection of mobile devices, and settings description.

3) Profile Mobility

Figure 5 shows an implementation of profile mobility that we have achieved using extended functions of SIP event messages (such as SUBSCRIBE, NOTIFY, and PUBLISH methods). This implementation assumes that a certain user will attempt to access his profile from different mobile devices. In the event that the user modifies his profile from a certain mobile device, his profiles stored on other mobile devices will be automatically updated to reflect those changes.

In our research, we aim to provide advanced context-aware services and personalized ubiquitous services, and to this end, we are paying particular attention to session mobility in IP-based wireless systems and to its performance evaluation.

3. User-centric Middleware Platform

This chapter describes a future middleware architecture for providing the service-platform capabilities described in Chapter 2 and a technique for providing user perceived QoS that takes into account several layers of communication system protocol stacks.

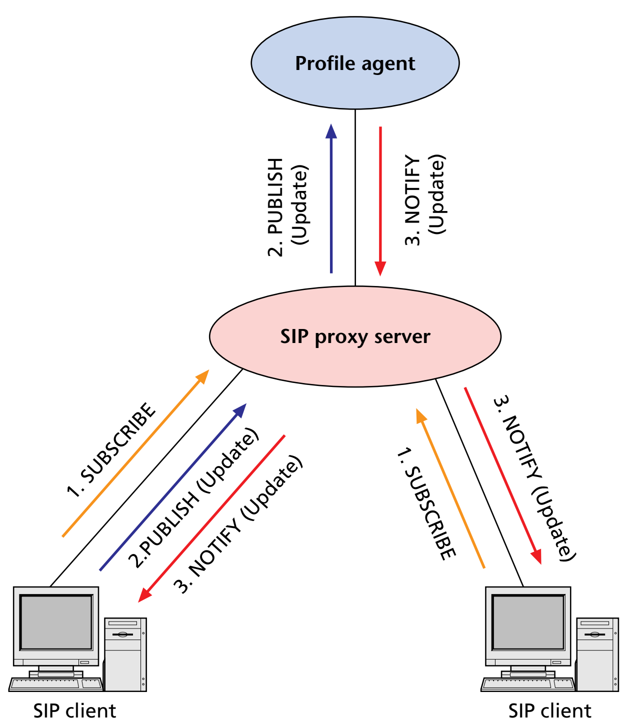

Figure 5 Implementation of profile mobility on a SIP-based system

3.1 Middleware Architecture

We propose a logical middleware architecture suitable for providing user-centric services in future mobile communications. This architecture must include the following capabilities as main constituent elements to achieve context awareness and ubiquitous services.

- "Personalization" to provide personalized services in accordance with user profiles (such as user preferences) and contexts, and "adaptation" to execute services dynamically in accordance with available resources such as mobile devices.
- "Community" to configure ad hoc virtual user groups for users having common goals, and "coordination" to make adjustments among members of a community.

Figure 6 shows the middleware architecture proposed by DoCoMo Euro-Labs. It assumes the reuse of existing middleware subsystems on hierarchical layers based on the conventional Open System Interconnection (OSI) reference model [13]. This service middleware features three layers: user support layer, service support layer, and network support layer.

The user support layer includes the four subsystems described above, namely, personalization, adaptation, community, and coordination, and is given autonomous execution not seen in past service middleware. Specifically, this layer can autonomously perform context analysis, personalization based on user preferences obtained from a user profile, and community formation and intra-community coordination, all of which reduces unnecessary user interactions with the system. This is a key point that makes it possible to provide user-centric services.

The service support layer corresponds to traditional middleware functions but introduces a dynamic service delivery pattern. This is a model that assumes the use of three subsystems—discovery and advertisement, contract notary, and authentication and authorization—to enable the discovery of services in accordance with a user request and to allow negotiation of service delivery conditions. As for service contracts, only a description framework is predefined, and after completing authentication, the contract content is dynamically updated and the service is delivered based on the updated contract.

The network support layer provides network-communication-control functions in support of heterogeneous environments. Here, mobility management includes routing and is carried out in a unified way over IP-based networks [14].

A prototype of this middleware platform is currently devel-

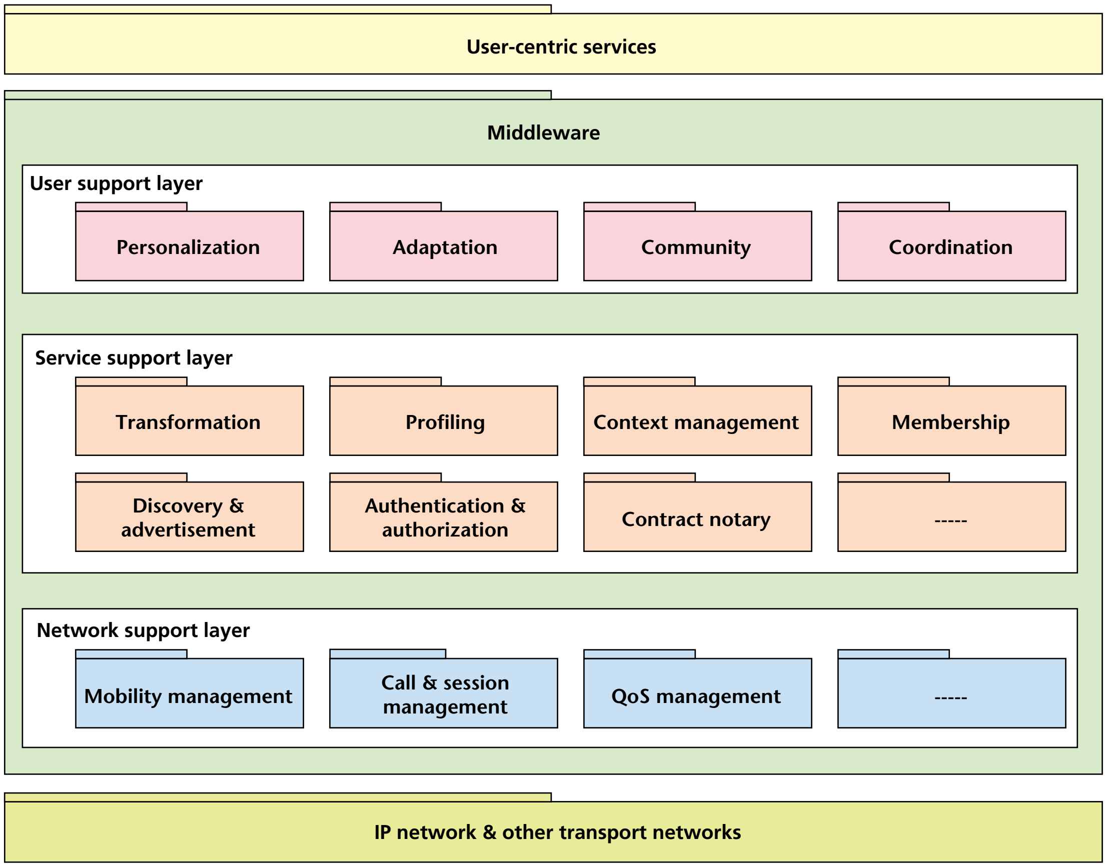

Figure 6 Middleware architecture

oped to evaluate the middleware architecture concepts introduced here. This prototype implements the dynamic service delivery pattern described above on top of P2P communications (such as JXTA), and we apply mobile-agent technology considering requirements for asynchronous communication in the event of network cutoff. This scheme makes it possible to load code dynamically, which provides an additional benefit in that the proposed middleware concepts can be applied to mobile devices for which resources are traditionally limited.

The architecture introduced here is currently being proposed to the Secure, Internet-able, Mobile Platforms Leading CItizens Towards simplicitY (SIMPLICITY) project in the 6th European Framework Program [15].

3.2 Cross-layer Design

Cross-Layer Design (CLD) aims to improve the user-perceived QoS by introducing a middleware platform that optimizes the use of network resources on different layers simultaneously.

As described in Section 3.1, a common approach for designing communication-network architecture is division of the network into multiple layers according to the Open Systems Interconnection (OSI) reference model. This is conducive to design modularization, and dividing up problems in this way can simplify the design process. At the same time, this technique may negatively affect capacity (the number of users that can be accommodated) as well as the QoS perceived by users. Considering, moreover, that the quality of radio channels in mobile communication systems changes dynamically, performance may drop if layers are designed separately. (In the case of Transmission Control Protocol (TCP), for example, errors on the radio channel can be misinterpreted as network congestion, forcing a reduction in transmission speed.)

The CLD approach developed by DoCoMo Euro-Labs is a new paradigm to design middleware architectures that takes into account the interactions and the information exchange between the traditional layers of the protocol stack including application layers.

As part of our research, we evaluated the effectiveness of this CLD technique in optimizing an application for multimedia communications and the radio link layer at the same time. The system used in the evaluation experiment consisted of a video streaming server and multiple mobile devices. The specific objective of CLD in this experimental system was to optimize the video quality perceived by users.

The optimization is achieved by considering three layers in the OSI reference model: the application layer, the data link layer, and the physical layer. **Figure 7** shows an example of CLD architecture that includes a protocol stack plus parameter abstraction, Cross-Layer Optimizer (CLO), and decision distribution elements. The radio link layer here corresponds to the data link layer and physical layer of the OSI model. The parameters of the application layer and radio link layer are converted into abstract parameters before being provided to the CLO, which searches the domains of these parameters to find values that optimize the objective function. The objective function varies according to the current goal such as maximizing the QoS of each user or maximizing the average QoS of all users. The decision distribution element here returns the values of selected parameters to each of these two layers. Reference [16] points out the advantages of using CLD for multimedia applications. In this case, the CLO parameters are channel bandwidth (time-slot allocation when using Time Division Multiple Access (TDMA)) and modulation system from the radio link layer and video transmission speed from the application layer.

In future research, we plan to study other applications of the CLD concept as well as a framework and middleware architecture for future mobile-communication networks that use CLD. Reference [17] introduces ideas for realizing CLD in open and programmable system environments.

3.3 Middleware for Smart Devices

Smart devices that incorporate sensors can act as wireless interfaces becoming important elements of a ubiquitous computing environment. For the future, with the aim of integrating ad hoc networks (such as sensor networks) and infrastructure-based networks (such as cellular networks), we plan to define a middleware architecture that can adapt to both environments and a middleware suitable to the physical limitations of smart devices.

4. Conclusion

This article has described key research techniques in relation to advanced service platform capabilities and user-centric middleware architecture. The validity of these research themes is supported by major telecommunication operators in Europe. It is also backed up by the results of Eurescom Project P1203 [18] that pointed out system concepts and application fields that would likely become important driving forces behind future mobile communication systems. Improving user satisfaction as services become more complex and diversified under a ubiquitous computing environment requires new concepts in service provision and operation.

Some of the research introduced here was accomplished in collaboration with other major research institutions and universities in Europe. It is our sincere desire that these research results be included in the 6th European Framework Program that major manufacturers, telecommunication operators, and academic institutions are joining to study unified architectural concepts for next-generation mobile-communication networks and individual technologies for realizing those concepts.

We are also participating actively in standardization organizations and industry forums such as the Object Management

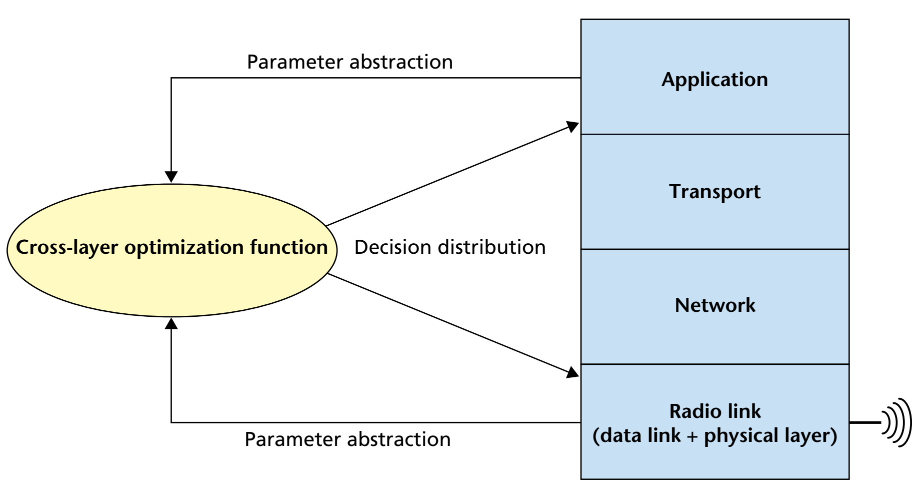

Figure 7 Example of CLD architecture

Group (OMG), W3C, IETF, WWRF, Wireless World Research Forum-Working Group2 (WG2) [19], and the Smart Card Project (SCP) of the European Telecommunications Standards Institute (ETSI) to disseminate the results of our research.

References

[1] M. Wagner, A. Tarlano, R. Hirschfeld and W. Kellerer: "From Personal Mobility to Mobile Personality," In Proc. Eurescom Summit 2003, Heidelberg, Germany, Oct 2003.

[2] W. Kellerer, R. Hirschfeld, A. Tarlano and M. Wagner: "Towards a Personalized Service Provisioning Environment," In Proc. IST Mobile Communications Summit 2003, Aveiro, Portugal, Jun. pp.15–18, 2003.

[3] K. Lyytinen and Y. Yoo: "Issues and Challenges in Ubiquitous Computing," Communications of the ACM, pp.63–65, Dec. 2002.

[4] A. Tarlano and W. Kellerer: "Context Spaces Architectural Framework," SAINT 2004 Workshop on Ubiquitous Services, Tokyo, Japan, Jan.26–30, 2004.

[5] W. T. Balke and M. Wagner: "Cooperative Discovery for User-centered Web Service Provisioning," In Proc. of the Int. Conf. on Web Services, Las Vegas, NV, 2003.

[6] A. Ankolenkar, M. Burstein and J. Hobbs, et al: "DAML-S: "Web Service Description for the Semantic Web," In Proc. of the Int. Semantic Web Conf (ISWC '02), Sardinia, Italy, LNCS 2342, Springer, 2002.

[7] D. Connolly et al: "DAML+OIL Reference Description," W3C Note, Dec. 2001.

[8] O. Lassila and R. R. Swick: "Resource Description Format Model and Syntax Specification," W3C Recommendation, 1999.

[9] T. Berner-Lee, J. Hendler and O. Lassila: "The Semantic Web." Scientific American, 284 (5): 34–43, 2001.

[10] W. T. Balke and M. Wagner: "Towards Personalized Selection of Web Services," In Proc. of the Int. World Wide Web Conf. (WWW), Budapest, Hungary, 2003.

[11] J. Glasmann, W. Kellerer and H. Müller: "Service Architectures in H.323 and SIP-A Comparison," IEEE Communication Surveys and Tutorials, fourth quarter 2003, Vol. 5, No.2.

[12] H. Schulzrinne and E. Wedlund: "Application-Layer Mobility Using SIP," Mobile Computing and Communications Review, Vol. 1, No. 2, Jul. 2000.

[13] C. Noda, A. Tarlano, H. Honjo, L. Strick et al: "Distributed Middleware for User Centric System," WWWF#9 Conference, Zurich, Jul. 2003.

[14] H. Yumiba, K. Imai and M. Yabusaki: "IP-Based IMT Network Platform," IEEE Personal Communications Magazine, Oct. 2001.

[15] IST-Project Simplicity, http://www.ist-simplicity.org/.

[16] W. Kellerer, Lai-U Choi and E. Steinbach: "Cross Layer Adaptation for Optimized B3G Service Provisioning," In Proc. of WPMC 2003, Yokosuka, Japan, Oct. 19–22, 2003

[17] C. Prehofer, W. Kellerer, R. Hirschfeld, H. Berndt and K. Kawamura: "Architecture Supporting Adaptation and Evolution in Fourth Generation Mobile Communication System," In JCN Journal of Communications and Networks, Vol. 4, No. 4, Dec. 2002.

[18] Eurescom Project P1203: "The operators vision on systems beyond 3G," 2003.

[19] WWRF, Wireless World Research Forum WG2, Service Infrastructure of the Wireless World, http://www.wireless-world-reserch.org/.

Abbreviations

3GPP: 3rd Generation Partnership Project
AMI: Application Message Interface
CC/PP: Composite Capability / Preference Profiles
CLD: Cross-Layer Design
CLO: Cross-Layer Optimizer
DAML: DAPRA Agent Markup Language
DAML+OIL: DAPRA Agent Markup Language plus Ontology Interface Layer
DAML-S: DAML Service ontology
ETSI: European Telecommunications Standards Institute
GUP: Generic User Profile
IETF: Internet Engineering Task Force
IMS: IP Multimedia Subsystem
IP: Internet Protocol
OMA: Open Mobile Alliance
OMG: Object Management Group
OSI: Open Systems Interconnection
OWL: Ontology Web Language
P2P: Peer-to-Peer
QoS: Quality of Service
RDF: Resource Description Framework
SCP: Smart Card Project
SIMPLICITY: Secure, Internet-able, Mobile Platforms LeadIng CItizens Towards simplicitY
SIP: Session Initiation Protocol
SOAP: Simple Object Access Protocol
TCP: Transmission Control Protocol
TDMA: Time Division Multiple Access
UAProf: User Agent Profiles
UDDI: Universal Description, Discovery, and Integration
W3C: World Wide Web Consortium
WSDL: Web Services Description Language
WWRF WG2: Wireless World Research Forum-Working Group 2
XML: eXtensible Markup Language