An Analysis of Service Continuity in Mobile Services

Ivar Jørstad
Norwegian University of Science and Technology, Dept. of Telematics, O.S. Bragstads plass 2E, N-7491 Trondheim, Norway
ivar@ongx.org

Do Van Thanh
Telenor R&D, Snarøyveien 30 N-1331 Fornebu, Norway
thanh-van.do@telenor.com

Schahram Dustdar
Vienna University of Technology, Distributed Systems Group (DSG), Information Systems Institute, A-1040 Wien, Argentinierstrasse 8/184-1, Austria,
dustdar@infosys.tuwien.ac.at

Abstract

Nowadays, the users are making use of more and more devices like mobile phone, PDA, laptop, Stationary PCs, etc. and it is crucial to provide service continuity not only across heterogeneous networks but also across heterogeneous devices. The users should be able to use services with minimum interruption even when changing devices. This paper provides an analysis of the composition and distribution of generic mobile services. It then addresses the impacts these models have on service continuity across networks, devices and domains.

1. Introduction

Till now, the focus in mobile communications has been on providing service continuity when a mobile terminal is roaming between networks, i.e. avoiding abrupt access to services. The underlying mechanism for achieving service continuity is called handover or handoff. Handover was first implemented between networks of same type, e.g. GSM [1] and is gradually extended to networks of different types, e.g. handover between GSM and WLAN [2]. With the increasing number of devices users have at disposition it is quite relevant to provide service continuity across heterogeneous devices. The goal of this paper is to examine how service continuity can be provided, not only across networks, but also across devices. The paper adopts a formal and analytical approach.

First, an analysis of current services, i.e. their composition and distribution, is performed. Next, the analysis is related to movement across networks, domains and devices. The functions and capabilities that are necessary to achieve service continuity are thus derived.

2. Composition of a generic mobile service

Mobility is the ultimate requirement for mobile services; they should by definition be available at any time, any place using any device with communication capabilities (thus supporting many types of mobility [3]). However, mobility can be further divided into sub-types. One approach is to discuss what the mobility (movement) is relative to. In this case, the following mobility types are commonly used:

- Personal/user mobility (user moves between devices)
- Terminal mobility (terminal moves between network access points)
- Service mobility (service moves across devices)

In addition, the mobility for each of these categories can be said to be continuous or discrete, where seamless mobility is the extreme of the former and portability the extreme of the latter.

In particular, it is more and more important that access to mobile services can be moved from one device to another with minimum interruption. To be able to address these requirements and further study them, formal definitions of the composition and architecture of mobile services is needed.

Mobile services can be modelled according to their composition or their architecture. Whereas the compositional model is concerned with the division of a service into discrete components according to their nature and role, the distribution model is concerned with the distribution of these discrete components across system elements. Together, these models will allow an analysis of service continuity across devices, networks and domains.

2.1. Composition model

- Service Logic
- Service Data

Figure 1 displays a UML (Unified Modelling Language [4]) Class Diagram showing the composition of a mobile service.

Service logic is the program code that constitutes the dynamic behavior and provides the functions of a service. The logic of a mobile service can be subject to various distributions, as in any distributed system [5]. The most common models to describe the distribution of service logic are:

- Standalone
- Client-Server
- Peer-to-Peer
- Multiple distributed components

Service data are used in the execution of the service logic and reflecting the state of it. Such data are typical *transient* and stored in volatile memory (e.g. RAM). They are for example variable values, temporal parameters, register values, stack values, counting parameters, etc.

In order to provide service continuity and personalisation we propose to introduce two additional service components as follows:

- Service Content
- Service Profile

Service content refers to data that are the product of service usage. These data are *persistent* and stored in non-volatile memory (e.g. EEPROM, Flash or on a hard disk). For example it can be a document written in a word processor or the entries in a calendar. Service content can be produced or consumed by the user.

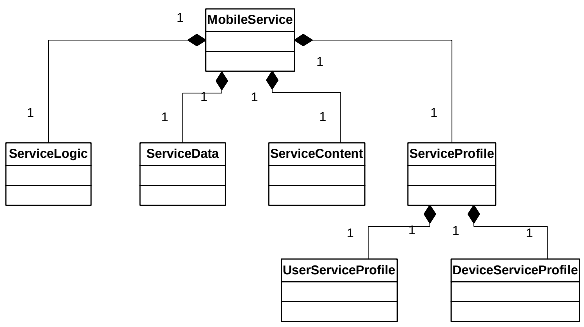

Figure 1. Composition model of a mobile service

Service profile contains the settings that are related to the user or/and the accessing device. The *User Service Profile* contains the user's settings that state how a specific service should behave for a specific user (e.g.

A generic mobile service is commonly modelled as consisting of two basic components:
what functions should be available), but it can also include personal information about a user that is used in accordance with a service (e.g. as input).

The *Device Service Profile* states the properties and qualities of a particular device (e.g. form factor), so that a service can adapt (e.g. the presentation layer) to this device.

The *User Service Profile* is stored or linked to the User profile, whereas the *Device Service Profile* is part of the Device profile.

The *User profile* can either be directly modified by the user or indirectly through the usage of the service. Alternatively, it can be implicitly altered by a separate service in the network according to the continuous usage pattern of a specific user.

2.2. Distribution model

According to the Distributed Computing paradigm, the distribution of a service/application should be hidden and the mechanisms to support distribution are incorporated in the Distributed Computing middleware such that the developer can concentrate on the core functions of the application [6][7]. However, for mobile services, distribution plays a crucial role that must be considered at service design time. Indeed, when the user is moving and is accessing services from different places, the location of a service and its components relative to the user's location will have great influence on its availability, quality, continuity and personalisation.

Important parts of a mobile service system are now defined. Networks consist of physical components and protocol stacks of distinctive types (e.g. a GPRS network or the Internet). Devices are the physical entities that end-users use to access services.

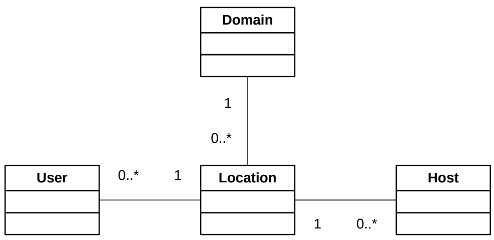

Figure 2. Relationship between parts of a mobile service system.

A Domain is a subset of a network (a host or set of hosts) that is controlled by an actor, and where access back and forth to the (larger) network is restricted. A domain can keep zero or several locations, and these locations can keep zero or several of both hosts and users (Figure 2). Locations are logical groupings of hosts (keeping service components) and users within a Domain. The procedure of combining the compositional model with the hosting system defined above is applied to obtain a distribution model for a mobile service.

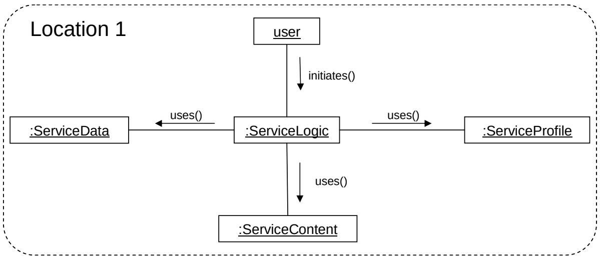

Figure 3. Distribution model #1 for a mobile service.

There are several distribution possibilities for mobile services, but only four cases are shown for illustration. Although these are not exhaustive, they expose the most critical properties that are needed for the further discussion of service continuity. The models are constructed by introducing Locations for and Associations between the components defined by the compositional model in Section 1.A.

1. All the objects including the user stay close to each other. This is the case of the user at home or the user travelling with a laptop containing standalone services. The access to services is continuous, reliable and simple but requires adequate processing, and storage capabilities in the laptop.

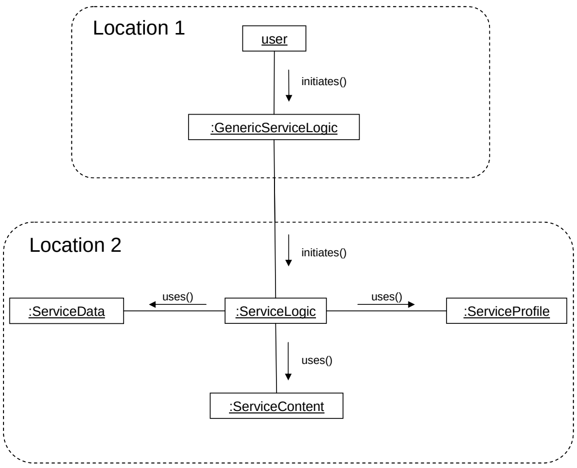

Figure 4. Distribution model #2 for a mobile service.

2. The user is located far away from the service and most of its components (shown in Figure 4). This is the case of the user travelling far way from services at home. The access to services is realised by a generic remote access application through a communication channel. For example, the user uses rlogin, telnet or a remote desktop client (VNC [8]) to access services and applications at home.
3. The user (see Figure 5), ServiceLogic1 and ServiceData1 are in the same location while the other objects ServiceLogic2, ServiceData2, ServiceContent1 and ServiceProfile1 are in a different location. Optionally, a ServiceProfile2 and a ServiceContent2 can be co-located with ServiceLogic1. This is the case where the user is using a service client to access the remote service server. This case is quite close the previous one but the client is *specific* to a service. The access to the service depends on the communication channel between the client (ServiceLogic1) and the server (ServiceLogic2).
4. The user and ServiceLogic1 and ServiceData1 are in the same location, whereas ServiceLogic2, ServiceData2, ServiceContent1 and ServiceProfile1 are in a different location. ServiceLogic3 together with ServiceData3 are in a third location (see Figure 6). The scenario is very similar to the previous, except that a third service logic part is accessed to render parts of the service, thus the service is also dependent on a second communication channel.

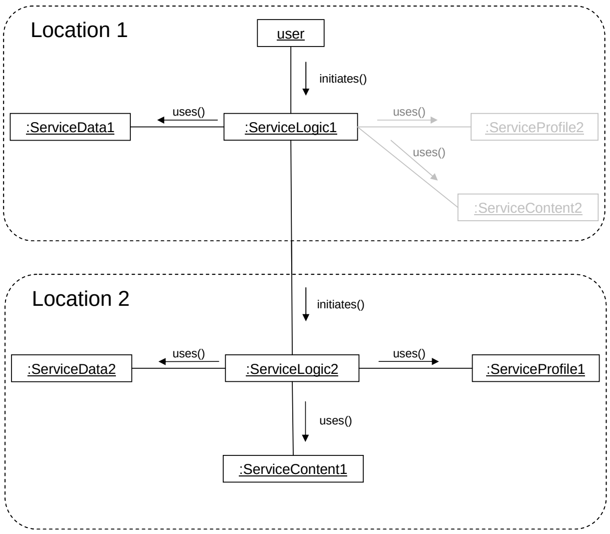

Figure 5. Distribution model #3 for a mobile service.

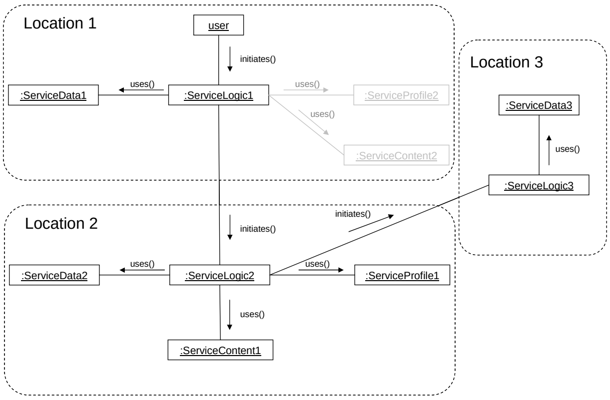

Figure 6. Distribution model #4, a more complex distribution of components.

The ServiceProfile in the composition model presented was initially divided into two sub-profiles, DeviceServiceProfile and UserServiceProfile, whereas in the four distribution models it was represented as one component.

Having presented a theoretical foundation of service composition and distribution, service continuity is now considered according to various relative movements of the elements in the service system. To avoid mixing concepts with other mobility terms, *service continuity* is introduced as an additional requirement of mobile services. It is defined as follows: "Service continuity implies that an end-user service is provided with minimum abruption in access, even when there is relative movement of user or components across devices, networks and domains."

3. Service continuity across networks

The roaming between networks of same type is addressed by handover mechanisms in GSM (Global System for Mobile communications) or UMTS (Universal Mobile Telecommunication System). For IP-based networks, Mobile IP [9] provides solutions for maintaining the communication channel at the network layer. There are also efforts aiming at providing handover between heterogeneous networks, e.g. GSM and WLAN or Bluetooth [10]. With better access to a common IP network as the result of these efforts, other challenges on higher layers can be investigated and hopefully tackled.

4. Service continuity across devices

In this section, a scenario is considered where a mobile service is moved from one device to another. The prerequisites for service continuity across devices vary with the distribution of the service in question, as well as the properties of the devices that are used. Due to the heterogeneity among devices, they must be categorised by type. If the following three properties are identical for two devices, they are regarded as the same *type* of device in this context.

1. Runtime Environment
2. Transport Layer Connectivity support in the runtime environment
3. Local Storage Support on the device

Of course, this is a simplification, because displays, processing power and other properties of the devices also vary a lot. However, to make this analysis easier, other properties are not considered. The application runtime environment ensures that the local service logic (GenericServiceLogic or ServiceLogic1) is executable on the new device, whereas the second property ensures the availability of communication primitives that can be exploited by the service logic to access service logic in another Location (e.g. ServiceLogic2). The third property ensures that local service content part of the service can be kept local on the new device.

For full service continuity to be supported across devices, the local service data should also be moved to the new device, thus preserving the *global state* of the service. However, the runtime environment is already taken into consideration, thus if the runtime environment supports such behavior (e.g. reviving objects from serialized data), state can be moved from one device to another.

5. Service continuity across domains

In the beginning of mobile telecommunication, mobile services were only provided by telecom operators. Lately, third party actors have been allowed to become service providers, thus extending the domains that can host services. This development will continue, and mobile services will be provided from any domain that is capable of hosting services. Today, mobile access to services in enterprise networks is provided [11], and soon, mobile access to services at home based LANs will be requested by consumers. Mobile services should soon embrace all domains shown in Figure 7.

Assume that some components of a service are spread across domains D = $\{x, y, z\}$, whereas the network access point (NAP) that the user's device is connected to is (together with at least one service component) in domain a not in D. A generalisation of the challenge is then, how to ensure access to all components of this service, thus fulfilling the ability to render the service.

In the definition of Domain in Section 2.2, an actor can for example be any of:

- Telecom operator
- Internet Service Provider (ISP)
- Independent enterprise

The access restrictions can be due to:

- Use of proxies
- Use of firewalls
- Use of Network Address Translation (NAT) in routers

For access to functionality in a telecom operator network, this is not possible without a special agreement (e.g Content Provider Access, CPA [12]) or through Parlay/OSA [13]. This is a hindrance for service continuity of services based in the telecom domain.

Enterprises often use a combination of proxies, firewalls and NAT to secure the network from the outside and consumers are often provided with a router/firewall that is used to connect to the Internet (e.g. with xDSL or cable) or they employ software only firewalls available in the operating system (e.g. when using dial-up connections).

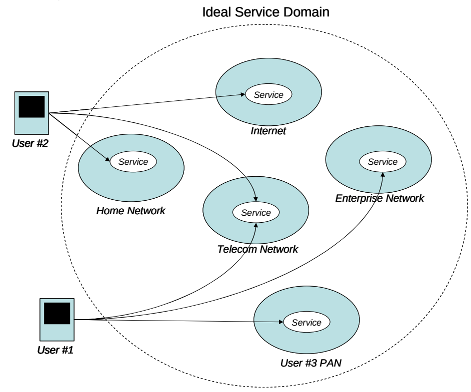

Figure 7. An expanded view of the mobile services domain.

All of these measures are needed to secure the property of the respective actors, but these measures are also prohibiting and complicating the expansion of the mobile service domain. Incoming connections to networks employing any of these mechanisms are not accepted without special configuration. Thus, solutions must be sought out that can allow services to access these domains while maintaining the existing high level of security. A general term to describe some of the restrictive elements mentioned is *middlebox*, and the challenge is often referred to as the *middlebox problem*. Solutions to the challenge are sought out in [14].

At first, service continuity across domains can be narrowed down to the existence of special service gateways between every domain. Initially, these can look like ordinary gateways, but in addition they must be capable of handling traffic back and forth between the domains based on information in service/application layer headers. This is also why the definitions of networks and domains are not the same, because while traffic back and forth between networks are governed by low-level packet header information (e.g. transport protocol information), traffic is now controlled at a higher layer. As seen, the importance and challenge of providing service continuity across domains is substantial, and must rely on support functions in all layers of the network stack and middleware.

6. Service continuity support functions

This section considers functions that are needed to support service continuity in the cases studied in the previous sections.

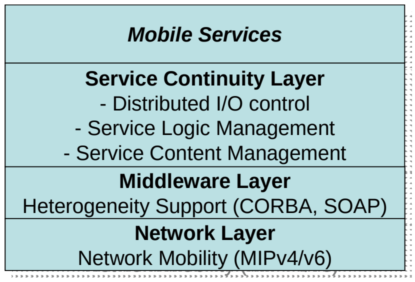

Figure 8. The introduction of a service continuity layer.

In Figure 8 we introduce a layer below mobile services which is called the Service Continuity Layer. The tasks of this layer are to allow services on top of it to provide service continuity. Some of the elements of this layer are depicted in Figure 9.

Elements of this layer will definitely be service logic management and service content management (part of the Service Composition Module).

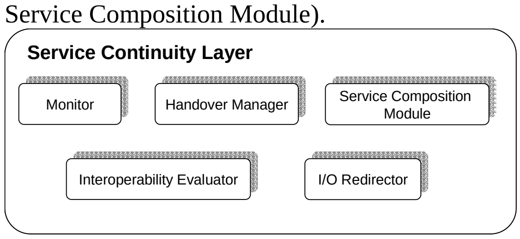

Figure 9. Elements of the service continuity layer.

It might be necessary to reorganise a service at runtime, thus moving and coordinating both service logic and content will be necessary. As an alternative to total reorganisation of the service, generic service I/O (input/output) redirection between devices to avoid replication of service logic and content throughout networks and devices could be provided by a novel function in this layer. An interoperability evaluator can determine if service components can be replaced at runtime, allowing service continuity using replacement components instead of moving components to new locations. Mobile agents [15] have previously been introduced to perform some of the tasks of the service continuity layer.

7. Future work

The distribution models described for mobile services are not exhaustive. The study of such models should be taken further in future work, and their properties investigated using a more formal methodology. This can help revealing weak spots that could restrict the development of mobile services in the future. Also, the briefly mentioned middleware for supporting service continuity must be further investigated in order to come up with a feasible architecture.

8. Conclusion

Distributed systems are in general complex to analyse particularly because of their asynchronous behaviour, but also due to heterogeneity. Realising mobile services with distributed systems is even more challenging, since mobility requires the service to be accessible also when there is some relative movement between components of the service. This paper presents a generic composition model for mobile services and thereafter four distribution models where the identified components are in varying degree spread across each of the distribution models. The resulting challenges to support service continuity across networks, devices and domains are thus discussed in relation to the composition model and the distribution models. Focus has been kept away from security and privacy issues arising as a result of enhanced distribution and availability of services. Rather, focus in this paper has mostly been put on functionality to support service continuity across domains, because the introduction of various elements in the networks between domains poses great challenges that must be solved by innovative technology. A Service Continuity Layer is introduced and some of its elements briefly described.

References

[1] Gunnar Heine, *GSM Networks: Protocols, Terminology and Implementation*, Artech House Mobile Communications Library, ISBN 0-8900-6471-7, January 1999

[2] Ye Min-hua et al., "The Mobile IP Handoff Between Hybrid Networks", The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2002

[3] Jun-Zhao Sun & Jaakko Sauvola, "On Fundamental Concept of Mobility for Mobile Communications", The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2002

[4] Object Management Group, UML 2.0 Superstructure Specification

[5] ITU-T X.901 | ISO/IEC 10746-{1,2,3,4}, Open Distributed Processing Reference Model Part 1,2,3 AND 4

[6] Kerry Raymond, "Reference Model of Open Distributed Processing (RM-ODP): Introduction", CRC for Distributed Systems Technology, Centre for Information Technology Research, University of Queensland Brisbane 4072 Australia.

[7] Object Management Group, "The Common Object Request Broker: Architecture and Specification, 2.0 ed.", July 1995

[8] Tristan Richardson & Kenneth R. Wood, "The RFB Protocol", ORL, Cambridge, Version 3.3, January 1998

[9] IETF, "RFC 2002: IP Mobility Support", 1996, available online at: http://www.ietf.org/rfc/rfc2002.txt

[10] D. Melpignano & T. Chiu, "Mobility Node White Paper", Philips Research, Monza, Italy, 2002

[11] Research In Motion, "White Paper: How BlackBerry extends the enterprise wirelessly", 2002, available online at: http://www.blackberry.com

[12] Telenor, "How does CPA SMS work?", available online at: http://cpa.telenor.no/cpa

[13] ETSI, "ETSI ES 202-915-1 v1.2.1 (2003-08), Open Service Access (OSA), Application Programming Interface (API), Part 1: Overview (Parlay 4)", ETSI, 2003

[14] IETF, "RFC 3303: Middlebox communication architecture and framework", IEFT, 2002, available online at: http://www.ietf.org/rfc/rfc3303.txt

[15] Stefano Camapdello & Kimmo Raatikainen, "Agents in Personal Mobility", Proceedings of the First International Workshop on Mobile Agents for Telecommunication Application (MATA'99). Ottawa Canada October 6-8 1999. World Scientific , pp. 359-374