Network-Centric Architecture for Context-Aware Information Delivery Services in 3G Mobile Networks

Eiji Kamioka¹, Shigeki Yamada¹ and Takako Sanda²
¹National Institute of Informatics
²The Graduate University for Advanced Studies
E-mail: ¹{kamioka, shigeki}@nii.ac.jp, ²sanda@grad.nii.ac.jp

Abstract
This paper proposes a context-aware information delivery system (CAIDS) using a UMTS Release 5 architecture and a wireless LAN. We present the concept and the system architecture of CAIDS, discuss the disposition of functional modules using the network-centric approach and clarify the CAIDS can provide the users with context-aware information delivery services including context-aware personal communication services. In addition, we discuss the preliminary evaluations of CAIDS performance in terms of the response time. The result indicates that the performance of application servers is a dominant factor of the response time in CAIDS.

Key Words: UMTS, mobile network, WLAN, ubiquitous computing, context awareness
1. Introduction

Next-generation mobile-computing networks are expected to evolve into ubiquitous computing networks in which an extraordinarily large number of computers will be embedded in various objects connected to the network, which will work cooperatively to support human activities. Ubiquitous computing\[1\] encompasses two major concepts: ubiquity and invisibility. Ubiquity means that numerous computers are located everywhere while invisibility implies computers will disappear below the threshold of our awareness. Context-aware computing is considered one way to achieve this invisibility, where user context such as the user's location, time, activity status, activity history, and preferences are automatically captured and analyzed by computers on the network to provide users with relevant content and services.

So far, most of the research efforts on context-aware computing have been devoted to sensor devices, computer-human interactions, and context-aware application software, but little research has focused on networking issues.

We believe that context-aware information services must seamlessly be available to users anywhere and at any time, i.e., from indoor to outdoor environments, from LANs to wide area networks (WANs), and from stationary to mobile networking environments. A universal network infrastructure for these must be constructed, for context-aware services to really penetrate into the fabric of our daily lives, by making use of the WANs' capabilities of covering both local and wide areas, enabling seamless network connections.

In this paper, we propose a context aware information delivery system (CAIDS) architecture using a UMTS Release 5 architecture and a wireless LAN. The CAIDS obtains user's context, analyzes and interprets this, retrieves the content from the content servers that the user needs based on the interpretation, converts the medium for the content into one adapted to the capabilities of the user device if necessary, and provides the user device with the converted content.
2. Requirements of Context-Aware Services

2.1 Context-Aware Service Scenarios

Here, we discuss examples of context-aware service scenarios to clarify the requirements for context-aware information delivery services and personal communication services.

The first scenario involves a tour guide in Paris and the user's interaction with one of the context-aware information delivery services. Suppose that a tourist walking in Paris tells his/her wearable computer that he/she wants to visit the Eiffel Tower. The context-aware information-delivery service system, connected through the network to his/her wearable computer, then analyzes his/her voice through the microphone in the computer, interprets this request, captures his/her geographical location and direction through a GPS device attached to his/her computer, and autonomously displays the information he/she needs on his/her head-mount display. The information includes the walking route to the nearest subway station, displayed on a map of Paris, with the correct subway fare to the Eiffel Tower.

The second scenario concerns context-aware personal communication services. Suppose that businessman Bob is sleeping on an airplane on the way to Singapore on business. His boss, Alice, phones him at his office. The context-aware personal-communication-service system detects the call and checks the network for Bob's location (on an airplane to Singapore), his activity status (sleeping), and his available communication devices (a handset near his seat on the airplane). The personal-communication-service system analyzes this context and recognizes that Bob is not able to answer Alice's call because he is asleep. Accordingly, it determines that Alice's voice should be recorded in a voice-mail server and this recording should be transcribed to an e-mail message so that Bob will be able to read it on his PC (that is connected to his handset) after he wakes up.

The two scenarios utilize various kinds of user contexts such as user location, direction, activity status, and available communication devices, and they include different types of context-aware services such as information-delivery and personal-communication services. Therefore, we first need to clarify the functional requirements and then consider an underlying infrastructure that is suitable for context-aware services.
Functional Requirements of Context-Aware Services

**Context-Awareness**
- Captures and updates user’s context
- The captured context is stored in a context database
- Deals with frequent changes and reuse of the context.

**Information Binding**
- Analyzes and interprets the context
- Binds the context to the content

**Information Provision**
- Converts the media content into one adapted to the receiver’s communication environment, if necessary.
- Delivers the (converted) content to the receiver at the right time in the right way

2.2 Functional Requirements of Context-Aware Information Delivery Services

The context-aware information-delivery service first captures user context, including that for the user's device, analyzes and interprets this, retrieves the content from the content servers that the user needs based on the interpretation, converts the medium for the content into one adapted to the capabilities of the user device if necessary, and provides the user device with the converted content.

From the two service scenarios detailed in the previous section and the above explanation, we identified that the following three kinds of functions would be necessary to implement context-aware services.

1. **Context-Awareness**

   Context-awareness requires context to be captured by various sensors or by the computing devices themselves from users and their environments. The captured context is stored and updated in a database that deals with frequent changes and reuse of this context. Context-awareness may be implemented by computers collaborating on the context-sender and context-receiver sides. Generally, raw context is recognized in the context-sender's computer while more refined, and thus abstract, context is recognized in the context-receiver's computer. Hereafter, we will focus on abstract context-awareness because it is mainly used to implement context-aware services over the network.

2. **Information Binding**

   The captured context must be analyzed and interpreted to identify or discover what content and services should be provided to users. We refer to context analysis and interpretation as the information binding context and content. Multiple bindings are possible between M kinds of contexts and N kinds of content, where M and N are integers greater than 1.

3. **Information Provision**

   The information provision function is divided into two kinds of sub functions, i.e., content conversion and its provision.

   The content that is delivered to the user should be converted into a form that is adapted to the capabilities of the receiver's device and the functions to display it.
3. System Architecture

3.1 Functional Modules

We introduce four functional modules of CAIDS into the network to handle the context information sent from the user and the content delivered to the user. The details are as follows:

(1) UIN (Universal Information delivery Navigator)

The UIN is a decision making module that accepts user context sent from the user device, executes user authentication, and analyzes and interprets the context. After this procedure in the context-aware information delivery service, the UIN decides the content the user needs based on its interpretation, refers to a directory server to obtain the location of the content and sends the location information to the user.

(2) DA (Directory Agent)

The DA manages the locations of content servers that are located in the local area networks connected to the Internet. It processes service discovery queries from the UIN and responds with a service reply that includes the locations of content matched against the request from the UIN.

(3) UMD (User Management Database)

The UMD is a database that manages information about a user who has subscribed to context-aware services, such as authentication information, preferences and context information that he/she has sent in the past. Context information in the UMD is updated when the user's context changes.

(4) MPS (Media Processing Server)

The MPS converts media content downloaded from a content server into an appropriate form, if necessary. This conversion is done based on the UIN directives.
3.2 Network Architecture

3.2.1 Access Network

For context-aware services to penetrate into our information society, they must seamlessly be available anywhere at any time, i.e., from indoor to outdoor environments, from LANs to WANs, and from stationary to mobile networking environments. To achieve this, wide-area coverage and mobility management capabilities must be supported. We therefore chose an underlying infrastructure that integrates the UMTS (Universal Mobile Telecommunications System) architecture[2] with a wireless LAN (WLAN). The UMTS is a third generation wireless system, which is being standardized by the 3rd Generation Partnership Project (3GPP)[3], that has been designed to provide high data rates of up to 2 Mbps and enhanced services to subscribers. The latest architecture, the UMTS Release 5, will offer both traditional telephony as well as packet switched services over a single converged packet-based network.

This figure illustrates a high-level view of the network infrastructure integrating the UMTS Release 5 with a WLAN. The UMTS mandates that the PS-CN be completely insulated from radio functionality so that the UMTS can work with different types of radio networks such as the WLAN. Therefore, as we can see in this figure, the underlying infrastructure also includes the WLAN, which is connected to the UMTS through the Internet Service Provider's (ISP) network and it complements the UMTS by supporting shorter-range but higher data rate wireless transmission. The integration of UMTS and the WLAN allows a broad range of data rates, a wide coverage from local to wide areas, and indoor to outdoor environments with complete seamless service/terminal mobility that can be supported with a single dual-mode MT.

Context-aware services in a wireless access environment require implicit inputs from a huge number of sensors and wearable devices in addition to conventional communication devices. These sensors and wearable devices may be connected to MTs through ad hoc personal area networks (PANs) with short-range radio technologies such as Bluetooth wireless technology to access the UMTS network and the Internet.
3.2.2 Disposition of Functional Modules

In terms of the three kinds of functions in the section 2.2, the abstract context-awareness and information binding are directly related to the context handling. Hence, the allocation of the CAIDS functional modules in the network plays a key part in the network architecture of context-aware services.

Here, we propose a network-centric approach (hereafter, we call this approach NCA) as a disposition of the CAIDS functional modules. The NCA causes the context handling functions to be located and managed in application servers inside the network and assumes the network operator that owns the network is also responsible for managing the context database and context-related servers. Therefore, the context-aware network infrastructure can be deployed all over the places in a uniform way by the network operator. As the end-to-end interoperability between users is also guaranteed by the network operator, they can enjoy the same classes of context-aware applications without user's context-aware facilities no matter where they are.

We have proposed EAPEC[4] (Environment-Adaptive Personal Communications architecture) as a network architecture of the NCA. The EAPEC captures the communication context (e.g., available devices, media and services) of a communication receiver, accepts a message from a communication sender, analyzes the relation between the communication context and the message type, selects the most appropriate communication device and medium/service type for the receiver, converts the sender's message type (the medium/service) into a form acceptable to the receiver if necessary, and finally forwards the message from the sender (or the converted message) to the receiver.

The EAPEC was also designed as a context-aware computing network architecture and we focused on only receiver's communications context. The EAPEC network architecture is based on 3GPP all-IP SIP network architecture[3]. The function modules of EAPEC are located in the IMS (IP Multimedia Subsystem) of UMTS network and connected to the SIP server as the extended HSS (Home Subscriber Server).

We adopt the IMS as the network subsystem where the CAIDS functional modules are placed and replaced the HSS with the CAIDS as a service component module in the UMTS network architecture because the CAIDS can be a context-aware information delivery system including personal communications provided by the EAPEC. As a matter of fact, the EAPEC is a person-to-person communication service with the context for the receiver's device, while the CAIDS is regarded as a person-to-machine or a machine-to-machine communication service with user context. These two services are regarded as almost the same in terms of service system functions and input/output data flows. Therefore, we can operate the CAIDS the same as EAPEC with user's context.
This figure illustrates the CAIDS network architecture including the access network. The CAIDS has enhanced the UMTS Release 5 architecture by replacing the HSS in the IMS of the UMTS with the UIN, the UMD, and the MPS, and adding the DA to the IMS. Context-aware services must handle various types of user contexts, but the existing HSS in the UMTS architecture only incorporates the user’s location context and user profile databases. Therefore, the NCA replaces the HSS with the more powerful UIN that incorporates the UMD that stores all types of user contexts, and has the capability of analyzing and interpreting the user context. The DA is a service information repository, which processes service discovery queries from the UIN and responds with a service reply that includes the URLs of all content, matched against the request from the UIN. Thus, information binding from context to content is done with the collaboration of the UIN, the UMD, and the DA. The MPS converts a medium type of content that the user downloads from the content server into a form that is acceptable to the user device if it is needed.
3.2.3 Communication Mechanism

We can see a typical information flow in the CAIDS in this figure that has two phases, i.e., from capturing context to providing the content-location information and from providing content-location information to downloading content from the content servers. Note that we omit the communication flow through the MPS for the sake of simplicity.

When any user context update is detected in the first phase, numbered from #1 through #16 in this figure and also shown in the figure of CAIDS architecture (Slide No.8), by the user's sensors connected to the MT, the update context information is transferred by the SIP[5] REGISTER message, either from the MT through the WLAN and ISP network to the IMS, or from the MT through the UTRAN and PS-CN to the IMS. The REGISTER message in the IMS is transferred through various SIP application servers (the P-CSCF, I-CSCF and S-CSCF) to the UIN located in the MT's home IMS. No matter where the MT moves, the UMTS architecture guarantees that the REGISTER message will reach the UIN in the MT's home network through the CSCF routing mechanism[6]. The UIN stores the updated context into the UMD, interprets the context, generates the required attributes, and sends them as a query to the DA through the Service Location Protocol (SLP)[7]. The SLP is an IETF standard, decentralized, lightweight, and extensible service discovery protocol. If the DA discovers any content whose attributes satisfy the conditions provided by the attributes in the query, the relevant content location information such as service URLs is returned. This location information is sent back to the user from the UCN to the MT of the user through the SIP 200 OK message.

In the second phase, the user checks the content location information and clicks his/her favorite location information. This triggers the transfer of the HTTP-GET message through the Internet to the associated content server. The content server returns the specified content. As the SIP uses UDP in its transport layer, it helps in reducing the protocol-processing overheads per SIP message. As the CAIDS is a natural extension of the UMTS, this means that only a minimum investment is required in building a context-aware network once the UMTS network is constructed. This also allows the UMD to be confined in the IMS and managed by the network operator. This may be desirable in terms of secure context data management.
4. Network Performance Evaluation

4.1 Performance Model

To evaluate the performance of the CAIDS network architecture, we constructed performance models to measure the response time of the networks. Here, the response time was defined by the interval from when the user generated a user context to when he/she received the content location information. Therefore, this response time directly corresponded to the round trip delays in the previous slide. The respective mean and 95th percentile response times of the architecture were obtained by queuing theory and simulation. We made several assumptions in the calculation and simulation of response times as we explain below.

The response time for the CAIDS architectures could be broken down into three delay elements: a processing delay\[8\], IP-network delay\[9\], and wireless communication delay.

Processing delay is caused by the application servers such as various CSCF servers, UIN (including process in UMD), and DA in processing SIP/UDP/IP, HTTP/TCP/IP, and SLP/UDP/IP packets. We assumed that all application servers had the same exponential processing time per IP packet. We also assumed that the ISP network and PS-CN in the figure of CAIDS architecture (Slide No.8) had the same IP-network delay time for the sake of simplicity. Each IP packet generated by the MT was based on a Poisson arrival.

The second delay element, IP-network delay, is produced in IP packet networks such as the IMS, LAN and the Internet, which include all of the network-layer, data-link-layer and physical-layer delays, caused by routers and other network devices. We assumed that the IP-network delay in each type of network obeyed a normal distribution.

The third delay element, wireless communication delay is generated at wireless access sections either from the MT, wireless link, and UTRAN to the PS-CN, or from the MT, wireless link, and WLAN to the ISP network. This wireless communication delay was assumed to be the same, between the UTRAN route and the WLAN route, for the sake of simplicity. Consequently, we will only focus on UTRAN access from the MT after this. We also assumed that this wireless communication delay had normal distribution characteristics. Mobility such as roaming and handover were not considered here, i.e., the MT was stationary during communication sessions from sending the context to receiving content location information.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Traffic Distribution</th>
<th>Values for Calculation and Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wireless-Released Delay either in MT, UTRAN and PS-CN or in MT, WLAN and ISP network</td>
<td>Normal Distribution</td>
<td>10 ms, as an average 3 ms, as an standard deviation</td>
</tr>
<tr>
<td>IP Network Delay in the IMS</td>
<td>Normal Distribution</td>
<td>10 and 20 ms, as averages 3 and 4 ms, as standard deviations</td>
</tr>
<tr>
<td>Processing Time in Application Servers</td>
<td>Exponential Distribution</td>
<td>10 and 50 ms per IP packet</td>
</tr>
<tr>
<td>Utilization of Application Server</td>
<td>Constant</td>
<td>From 0.1 to 0.9 with the interval of 0.1</td>
</tr>
<tr>
<td>IP Packet Arrival Rate</td>
<td>Poisson Distribution</td>
<td>Arrival rates are determined so as to satisfy the specific utilizations of application servers</td>
</tr>
<tr>
<td>Processing Time in MT</td>
<td>Constant</td>
<td>10 ms</td>
</tr>
</tbody>
</table>
4.2 Performance Results and Analysis

The upper figure represents the response times of the CAIDS for low-performance application servers, with the processing time of 50 ms per IP packet and the IP network delays of 10 ms and 20 ms. In this case, both the mean and 95th percentile response times with the IP network delay of 20 ms are slightly larger than those with the IP network delay of 10 ms in the wide range of server utilization. This figure also shows that the response times drastically increase as the server utilization becomes high.

The lower figure represents the response times for high-performance application servers with the processing time of 10 ms per IP packet, and the IP network delays of 10 ms and 20 ms. Compared with the result for low-performance application servers, the tendency of the response times to the server utilization is similar, however, the absolute values of the response times are much smaller than the previous ones.

To summarize, although the IP network delay does not very influence the response time in the CAIDS, the performance of application servers strongly impacts the response time. However, we expect the performance of application servers will increase year by year, so that the response time will meet our requirement. For the implementation cost of the CAIDS, as the application servers are provided by the network operators, these results will be useful for them to design the server installation in their networks.
5. Conclusions

We proposed a context-aware information delivery system (CAIDS). We presented the concept and the system architecture of CAIDS, discussed the disposition of functional modules using the network-centric approach and clarified the CAIDS could provide the users with context-aware information delivery services including context-aware personal communication services. In addition, we discussed the preliminary evaluations of CAIDS performance in terms of the response time and realized that the performance of application servers was a dominant factor of the response time in CAIDS.

The disadvantage of the network-centric approach is that building the network infrastructure is completely dependent on the network operator so that it may take a long time and a large investment to spread context-aware services. We will investigate the problem and also implement the CAIDS prototype system in a test bed network.

References