CONTEXT AWARE SERVICES AS A STEP TO PERVASIVE COMPUTING

M. E. Anagnostou[1], A. Juhola[2], E. D. Sykas[1]

[1]ICCS, Institute of Communication and Computer Systems, Computer Networks Lab, Athens, Greece
[2]VTT, Technical Research Centre of Finland, Information Technology, Espoo, Finland

Introduction

The future of mobile computing after the death of WAP seems uncertain. WAP carried with itself an expectation that it would be an important driving force in areas ranging next generation mobile systems to microelectronics. The vacuum left behind by WAP and even a much greater space is now expected to be covered by the relative new buzzword of "pervasive computing" or its European version, namely "ambient intelligence". Pervasive computing is about "the creation of environments saturated with computing and communication capability, yet gracefully integrated with human users". It involves advances in a number of key areas, including sensors, interfaces, smart spaces, distributed and embedded computing technologies, service intelligence, and security.

Context aware service engineering can be seen as both the future trend in service engineering and an important aspect of pervasive computing. It has the additional advantage that it can be pursued today and thus provide one of the first steps towards pervasive computing. Context awareness in services is actually about closely and properly linking services, so that their user is relieved from submitting information that already exists in other parts of the global system. In this manner services are expected to act in a concerted mode, which finally add to user friendliness substantially.

Yet, context awareness cannot be achieved without an adequate methodology and a suitable infrastructure. The CONTEXT project deals with both. Regarding infrastructure, CONTEXT will rely on the capabilities of active networks. An active network contributes to the aims of this project in two ways: (a) by providing a network status awareness, which is part of the general context awareness, and (b) by allowing an improved network resource utilisation.

An essential aspect of CONTEXT work is the adoption of an end-to-end approach for service provisioning, from users to service providers and to the global Internet. A uniform way and underlying means for allowing the users to contract context-based services with service providers will be specified, complemented with the necessary intra- and inter-domain means for fulfilling and assuring the contracted service levels.

Background: Personal communications and VHE

This section serves the purpose of giving background information on a collection of relatively old concepts, which can be considered as relevant with the new concepts of ambient intelligence and context awareness.

The objective of the *RACE2104-PERCOM* (Personal Communication Space in Layered Networks, 1995-96) project was to develop and demonstrate personalised services features over an ATM network [Percom]. These features included authentication and registration procedures, session mobility and flexible service profiles (FSP). Essentially PERCOM tried to solve the problems of a user moving between terminals, and this approach was different from the one adopted by a mobile telephony system, which assumes that a user will always carry a portable terminal. Session mobility is the feature that allows connections within a session to be changed, while the session itself is maintained. FSP is a feature used to store, manage and access information related to the stakeholders and services. The users would be able to personalise their services by setting several service parameters according to their preferences. For example, a PERCOM system user would be able to handle incoming calls according to his/her preferences, to route them to the nearest suitable terminal in a visited environment, and to some extent to adapt the form of the message to the capabilities of the terminal.

While PERCOM tried to support a user's mobility between terminals, the ACTS325-MONTAGE project (MObile INTelligent AGEnts in Accounting, Charging and Personal Mobility Support, 1998-2000) [Montage] tried to support user mobility between administrative domains, by ensuring that the services, to which a user has subscribed in home domain, will also be available in a visited domain.

These ideas were compacted in various definitions of the so called Virtual Home Environment, as in [VHE99]: "Virtual Home Environment (VHE) is defined as a system concept for personalised service portability across network boundaries and between terminals." Two somewhat older VHE "definitions" are as follows: "VHE means that the user will have the same interface and service environment regardless of location" [UMTS97]. "VHE provides operator specific services that are accessible by the user even when this user is roaming outside the home network. The establishment of this concept realises that service provision and network operation may be separated, allowing services to the offered by networks other than those providing the home or visited call processing capabilities" [ITUT99].

A first experimental and partial implementation of VHE has been provided in the trial of the IST VESPER project. A final trial, which tested a VHE implementation was scheduled for autumn 2002. Nevertheless, the project was discontinued in summer 2002. Some key innovation areas expected from VESPER were:

- Service continuity.
- Service scalability (i.e. adaptability to network, terminal characteristics).
- Definition, standardisation and management of service and user profiles.

Note that older versions of the VHE concept are closely related with communication services, including the ones based on third generation mobile systems. The aforementioned VESPER project adopted a general definition, which involves the usage of fixed and mobile terminals of various types and capabilities. In essence, VHE is the necessary capability of the environment, which can relieve a person from dealing with various networks and terminals. In other words, in these projects it is understood that a person (or a machine) can possibly move from one environment to another, and that this changing context is somehow understood by the infrastructure so as to supply the user with an almost permanent set of services. However, the following factors have added to the complexity of this scenery:

(1) Services, which tend to move from merely supplying communication capabilities towards supplying content, become more and more dependent on each other. Services need to communicate with other services as people themselves do.

(2) Tomorrow's environment will be much more fine-grained than yesterday's. So far the environment context was only the administrative domain, which might as well be the same in a relatively large space, and the communication infrastructure of a specific space. Now the location services are emerging, so location has to be taken into account. Location brings in additional context, since many other factors can be taken into account in providing a service once the location is known, e.g. weather, temperature, local density of people, local shops, etc.

While personal communication space, VHE, and mobile systems play around the ideas of moving the user or moving the terminal or both, the introduction of flexibility in the terminal side is an interesting evolution. The requirements and design of re-configurable user terminals based on software defined radio is the main objective of the IST TRUST project [MDVN01]. A re-configurable terminal is able to reconfigure itself in terms of capability, functionality, and behaviour in order to accommodate the needs of the user. A fundamental property of such a terminal is the ability to adapt itself to different systems and standards (e.g. GSM and UMTS).

The new trend: Pervasive computing

"The word *pervasive* means having power to spread throughout. *Pervasive computing* is an environment where people interact with various companion, embedded, or invisible computers. It essentially means to enable networked devices to provide services to and use services from peers effectively" [GLPS01]. It refers to "the creation of environments saturated with computing and communication capability, yet gracefully integrated with human users" [Saty01]. Pervasive computing is essentially about interacting with a smart environment, which consists of humans, devices and open or closed spaces. Therefore, an "environment and context conscious" service can become more adaptive with minimal human guidance.

The *mobile computing* field involves results, which fall in the following areas [Saty01]:

- Mobile networking, which includes protocols and techniques for improving the performance of wireless networks.
- Mobile information access.
- Support for adaptive application, including adaptive resource management.
- System level energy saving techniques.
- Location sensitivity.

Pervasive computing goes beyond mobile computing by adding the following trends:

- Effective use of smart spaces: This idea has evolved from the intelligent building concept. A user's computer can take control over heating, cooling, lighting and other physical environment properties within a space. The inverse effect is also possible, as the user's device may adapt its behaviour to the environment or profit from environment resources, e.g. by exploiting (external to the terminal) local processing power.
- Invisibility: A pervasive computing environment is expected to (a) run smoothly with minimal human intervention and (b) meet conscious or even subconscious human expectations. Requirement (a) has been inherited from plain telephony systems.
- Localised scalability: The volume of a terminal interaction with the local environment decreases as the distance between the terminal and the aforementioned environment increases. This property is expected to solve the scalability problem, which will be created by the increased number of terminals and smart appliances.
- Masking uneven conditioning: Different environments will provide different capabilities, but even the poorer environment should try to compensate for its deficiencies. For example, currently available PDAs, which have only occasional internet access, try to download pre-selected web pages, which are often visited by the user.

Management issues for a pervasive computing environment include [MDAD01]:

- Dynamic IP configuration and autoconfiguration over entire networks or even dynamic topologies.
- User registration, authentication, profile maintenance.
- Mobility support, depending on a user's needs.

An important aspect of pervasive computing relates with capturing a user's needs and trying to exploit it in order to improve the final service quality. In the service scenarios that appear in [Saty01] pro-activity is a desirable property of the system, which brings forward a number of issues:

- The ability of the system to infer a user's intentions by analysing available information.
- The human intent representation model, its accuracy and detail.
- A cost vs. benefit analysis of the system's ability to offer pro-activity.

For example, if in a given environment a service cannot be offered at the appropriate quality level, the system might try to solve the dilemma of discontinuing the service or not, by predicting the user's decision, instead of asking annoying or unintelligible questions.

Furthermore, a technique, which may facilitate the implementation of certain functions that demand a processing capacity beyond the capacity available in a terminal, is known as *cyber foraging*. Usually a mobile terminal will exploit the resources of the nearby wired infrastructure. Service discovery, negotiation, charging, load balancing, security, and even battery life considerations are some of the problems, which must be solved before cyber foraging can be implemented.

The theoretical and practical problems that must be solved before a transition to a world of pervasive networking is possible are numerous and the general problem is always open to improvements and inventions. Reference [Nobl00] elaborates on a system for mobile, adaptive, applications and describes the

Odyssey platform for mobile data access. The whole issue (Vol. 7, No. 1) of the IEEE Personal Communications Magazine, in which [Nobl00] appears is devoted to pervasive connectivity enabling technologies, including a report on the IEEE 802.15 standardisation process. The scalability problem in smart spaces, which host wireless devices and sensors in large numbers, is explored in [BGS00]. The paper explains how to use distributed query execution techniques to leverage the computing capabilities of devices and reduce communication.

Context awareness

The transition to a world of ambient computing involves advances in a number of key areas, including sensors, interfaces, smart spaces, distributed and embedded computing technologies, service intelligence, and security. Also, the transition can only be gradual. Each transition phase will be shaped by technologies, which are available at the time. Each phase will introduce new elements in the puzzle of ambient computing and will increase the adaptation of the users to the new concepts and tools. It seems that currently the mobile communications technology, the location technologies, and the creation of new, intelligent, adaptive, proactive, context aware and linked services can lead the first wave, since all these technologies either are already mature or involve tractable problems.

The context of a person-service combination is circumscribed by the following non-exhaustive list of sample information items: Calendar, user location, weather, personal preferences, permission profiles, other services offered to the same person, other persons in the same area. Consider the following scenario:

"A tyre of a car, which moves on a motorway, is under gradual deflation. The next tyre repair facility is located 10 km away. At the present deflation rate the car is foreseen (by the intelligent system) to safely arrive at the repair facility. The onboard intelligent management system checks whether this facility will be open at the estimated arrival time, and arranges an appointment for the repair. Then, it gives a warning to the unlucky driver and asks him to proceed to the repair facility. Finally, it re-calculates the driving schedule and outputs the result to the passengers' calendars. The calendar of each passenger (a) corrects the arrival time at the hotel, (b) gives a notification to the hotel reception system and (c) postpones any appointments by a suitable amount of time. (For example, the hotel management system may use this particular piece of information to postpone the pre-heating of the rooms of the passengers.)"

At first sight this scenario is impressive, but a closer investigation reveals that a careful design of relatively simple service-to-service interfaces should suffice and that parts of the above systems already exist. For example, tyre pressure supervision systems, which give a visible warning to the driver filtered by an onboard computer, have already appeared in some high-end commercial cars. Location services, which can find for you the closest facility of a required type, are already offered by certain mobile operators. Car navigation systems, which can calculate trip time, are also available. The navigation module exports a single parameter value, namely arrival time, to the calendar.

The following problems must be resolved for context awareness:

- What is in a more or less complete (for the purpose it serves) context information and user status representation model? Should it be designed in a top-down fashion by including all possible factors that might influence a user's activity, or should it be designed in a bottom-up manner by referring to a user's interaction with the service provision system.
- How should a user be monitored? Who has the responsibility to maintain a user's status information? What privacy, security and other legal issues are involved?

Of special importance are the privacy/security and scalability problems. There is an inherent contradiction in trying to build context aware albeit secure services, because context awareness is about spreading information around, while security and privacy is about preventing information dispersion. Also, scalability is problematic due to the increase in possible service-to-service interfaces. In the *Jini* design [Jini] the problem is addressed by creating a lookup and discovery service, which operates a common service repository.

The implementation of a context aware system raises a number of important issues:

- Selection and implementation of a context representation model.
- Context storage capacity requirements and responsibility.
- Context processing and decision making algorithms.
- Information collection techniques for context awareness.
- Security and privacy of the context related data.

An answer to the problem of making a model for humans and machines for Web representation purposes can be found in [DeCa01]. This model is an evolution of the avatar concept. The avatar, i.e. the virtual entity, which represents a person or a machine, partially reflects selected current and permanent properties of the real one. By interacting with the virtual entity one can finally establish interactions in the real world. For example, a husband, who has privileged access to his wife's avatar and is currently abroad, may give an order a nearby florist to bring her flowers, while she has tea at a friend's house.

Active networks as an enabling technology for context aware services

The network support for mobile users enjoying adaptive and context sensitive services places unprecedented demands for the network platform used. It is clearly unthinkable to statically configure every (access) node to support the individual service demands of every single mobile user, so the ability to dynamically instantiate personalised service functionality only when and where needed is a practical prerequisite.
Currently there are not many networking paradigms daring to promise the fulfilment of this requirement: They are Active and Programmable Networking, including their derivatives.
So the choice of the networking platform for our purposes can be narrowed down to the selection of the suitable form of active/programmable networking, with no viable alternative in sight.

The branch of programmable networking taken to a serious consideration in CONTEXT makes a separation between IP forwarding and control planes in an active node. This is in line with the IETF FORCES [FORCES] approach. The main idea is that the forwarding plane can be dynamically configured according to the instructions received from the control plane. The CONTEXT active node's control plane will support dynamically loaded code ("active code") that will indirectly use the control plane's power to configure the forwarding plane (within limits) and to extract context information from the network. Other options available for the active code will include the possibility to "grab" packets from the forwarding plane into the control plane for further examination, possibly triggering subsequent actions and/or packet content treatment. The packet, or a packet modified based on it, can also be sent back to network.

The active code is the most important tool in implementing the context sensitive services, since the active code can be programmed to watch for context changes and to react on them. Say, to trigger forwarding plane configuring or other actions.

The injection of the active code can be made subordinate to operator control, thus avoiding security problems that might otherwise arise.

Noteworthy other developments related to active/programmable networking include the recently developed paradigm of Mobile Agents (MAs) Although these concepts were introduced by different research communities to address different problems, they have started overlapping in focus and applicability. Next - generation service networks can benefit from active networking paradigm with the dynamic deployment of network services that can be tailored to the user's requirements using context. Besides, ANs distribute code to proxies at the edge of the network, as well as, to the mobile device, thus reducing the number of necessary network transactions to provide a service, by local processing and local service customisation.

The specific platforms currently under consideration are a variant of [ABLE] (The Active Bell-Labs Engine for Network Management) and LANE (Lightning Active Node Engine, VTT). Both have chosen a simple, practical minded approach to the programmable networking and support the general features as described earlier in this chapter. In addition, they have similarities in implementation.

CONTEXT will deploy programmable network platforms in several member sites and implement the modifications and additions required for experimentation with context sensitive services to be defined based on the scenarios presented in the project's technical description.

The CONTEXT project for context aware services

The *IST-CONTEXT* project (Active Creation, Delivery and Management of Efficient Context Aware Services, to begin Sept. 2002) [Context] main objective is to design, develop and assess innovative models and middleware solutions for an efficient provisioning of context-aware services making use of active systems technology on top of dynamically configurable Internet IP layer. . This will allow the composability and dynamic adaptability of current and future context aware services for the benefit of the mobile user. There will be no need to develop services for "fixed" users specifically, since the "fixed" user can be considered to be a special case of the more general mobile one. From the technical point of view the CONTEXT solution will span into three domains, namely: the Service Layer (SL) domain, the Active Applications (AA) domain and the IPv6 domain. In the service layer domain, this project will deal with modelling of the information that expresses the context of services and establishing a framework for the creation of context aware services The AA based solution, with appropriate APIs to control the IP domain, will allow the actual delivery of policy based context aware services.

In the Service Layer domain, CONTEXT aims

- to identify mechanisms for the definition, exchange and acquisition of contextual information
- to identify mechanisms for creating context sensitive services (binding of contextual information into services, thus creating context-sensitive services)
- to identify mechanisms for provisioning 'contextualised' services (based on the available provisioning means of the existing service infrastructure of the domain)
- to propose a policy-based framework for service management including configuration of the services according to the personalised user profiles
- to specify and implement the interactions with underlying (Active) Network Management Systems to facilitate personalised service delivery.
- to allow the enforcement of monitoring policies for Service Level Agreement management

In the Active Application Layer domain the CONTEXT aims

- to enhance active network technology in order to provide efficient delivery of context based services especially in the mobile network environment. This will be accomplished by specifying and developing an API on top of existing AN nodes to allow the following functionality:
- Provision of context information collected from the data plane (packet) and the management plane for the needs of context sensitive services and management. The information made available can include network characteristics, network policies to be made known to users/other networks, network charging principles/tariffs, network's current status and network resources available for the requestor.
- Utilisation of the context information for context sensitive (user) services
- Utilisation of the context information for policy-based active/programmable network and node management. For example, both the user and operator could place conditional actions (policy rules) for matching user applications with their immediate networking options.

References

[ABLE] www.cs.bell-labs.com/who/ABLE/

[BGS00] P. Bonnet, J. Gehrke, P. Sashadri, "Querying the physical world," *IEEE Personal Communications*, Vol. 7, No. 5, pp. 10-15, Oct. 2000.

[Context] IST-CONTEXT project proposal, Annex 1, May 2002.

[DeCa01] P.Debaty, D. Caswell, "Uniform web presence architecture for people, places and things", *IEEE Personal Communications*, Vol. 8, No. 4, pp. 46-51, Aug. 2001.

[FORCES] http://www.ietf.org/html.charters/forces-charter.html

[GLPS01] S.K.S. Gupta, W.-C. Lee, A. Purakayastha, P. K. Srimani, "An overview of pervasive computing", *IEEE Personal Communications*, Vol. 8, No. 4, pp. 8-9, Aug. 2001.

[ITUT99] ITU-T Q.1711 "Network functional model for IMT-2000", March 1999.

[Jini] R. Gupta, S. Talwar, D. P. Agrawal, "Jini Home Networking: A Step toward Pervasive Computing", *IEEE Computer*, Vol. 35, No. 8, pp. 34-40, Aug. 2002.

[MDAD01] A. Misra, S. Das, A. McAuley, S.K.Das, "Autoconfiguration, registration, and mobility management for pervasive computing," *IEEE Personal Communications*, Vol. 8, No. 4, pp. 24-31, Aug. 2001.

[MDVN01] M. Mehta, N. Drew, G. Vardoulias, C. Niedermeier, "Re-configurable terminals: An overview of architectural solutions," *IEEE Communications Magazine*, Vol. 39, No. 8, pp. 82-89, Aug. 2001.

[Montage] D. Prevedourou, M. Anagnostou, et al., "Use of agent technology in service and retailer selection in a personal mobility context", *Computer Networks*, Vol. 31, pp. 2079-2098, Sept. 1999.

[Nobl00] B. Noble, "System support for mobile, adaptive applications," *IEEE Personal Communications*, Vol. 7, No. 1, pp. 44-49, Feb. 2000.

[Percom] H. Bakker, J. Heilesen, G. Schapeler, M. Anagnostou, C. Georgopoulos, "The Implementation of PSCS Features in a RACE ATM Demonstrator", Proc. of the Third International Conference on Intelligence in Broadband Services and Networks-IS&N'95, Heraklion, Crete, Oct. 1995.

[Saty01] M. Satyanarayanan, "Pervasive computing," *IEEE Personal Communications*, Vol. 8, No. 4, pp. 10-17, Aug. 2001.

[UMTS97] UMTS Forum "Report 1: A Regulatory Framework for UMTS", June 1997.

[VHE99] "Universal Mobile Telecommunications System (UMTS); Service aspects; Virtual Home Environment (VHE)", ETSI TR 122 970, version 3.0.1, Release 1999.

[Weis91] M. Weiser, "The computer for the 21st century," *Scientific American*, Sept. 1991.