System-Initiated Context Modification to Improve Network Resource Usage

***Utz Roedig[1], Cormac J. Sreenan[1],**
[1]Mobile & Internet Systems Laboratory,
Department of Computer Science,
University College Cork,
Ireland
E-mail: {u.roedig|c.sreenan}@cs.ucc.ie

Abstract: The conventional goal of a context-aware computer system is to improve the user's interaction with the system. But it is useful also to employ a user's context information to optimise the operation of the system itself. In particular the context can be used to improve the use of network resources within the system. Furthermore, in achieving this goal, it might be useful to allow the system to influence the user's context in a controlled and restricted way. In this paper we report on our investigation regarding the impact of system-initiated context modifications on the usage of network resources in a context-aware system. A model is presented that allows this impact to be succinctly described and quantified. The model is then evaluated by analysing its behaviour for a specific class of applications - context-aware guides for a museums and art galleries.

1. Motivation and Introduction

Many computer systems today use context information to improve the services they provide. These computer systems are said to be *context-aware*. In the literature many definitions for the terms context and context-aware can be found, so their usage is often rather ambiguous. Within this paper, the definition for the term context as given by Dey [1] is used: *Context* is any information (e.g. user location) that can be used to characterize the situation of a user. This information is considered relevant for the interaction between the user and the system and may be obtained by various means, including from physical sensors – a system subclass known as sentient systems. The definition used for the term context-aware is derived from the definition given by Schilit [2]: a system is considered *Context-Aware* if it uses context information to adapt its behaviour.

Available context can be used within a context-aware system in several ways. Most context-aware systems employ context to improve the interaction between the user and the service provided by the system, for example by adapting a graphical user interface based on a user's current location [3]. Additionally – and this is the focus of our research - it is also possible to use the available context to improve the system itself. More specifically, context can be used to improve the usage of network resources *within* a system.

Beyond the possibility that computer systems might use available context for resource management, it is also possible that the computer system may achieve additional gains by being allowed to influence the user's context in a controlled manner. This approach has received little attention in the literature, but offers potentially rewarding benefits, for example by being allowed to alter the sequence or timing of user requests to download or access information. For obvious reasons, this influence might only be allowed in a limited and controlled way.

In this paper, the impact of system-initiated context modifications on the usage of network resources in a context-aware system is addressed. More specifically, it is investigated under which conditions the following hypothesis can be accepted:

If users in a context-aware system allow context modifications by the system, network resource usage in the system can be improved.

For our analysis a model is developed which succinctly captures the influence of system-initiated context modifications on the usage of network resources in a context-aware system (Section 3 and Section 4). Thereafter, simulation is used to analyse its behaviour for a specific class of applications - context-aware guides for a museums and art galleries (Section 5). Finally it is shown under which conditions the stated hypothesis can be accepted for this class of applications and the achievable improvements in this system are quantified (Section 6).

2. Related Work

Some research investigates the area of context-aware communication (see [4] for an overview). Context-aware communication is defined as the class of applications that apply knowledge of people's context to reduce communication barriers. Context is used to decide when, where and how communication between users should take place. The main goal is to optimise the offered service, not the communication system itself. However, the context-based decision which network resource (e.g. channel selection [5]) should be used for communication could also be done to optimise the system.

Another relevant research field is the adaptation of a single communication path based on available context information. For example in [6] it is suggested that the applications have to be aware of and adapt to variations in the user's context and available resources. Part of the research on context-based adaptation of a communication path is the context-based handover. For example in [7] and [8] it is investigated how handover decisions can be improved by using context information. In the aforementioned research, context is used to improve the system and the interaction between the user and the system.

3. Context Usage

3.1 Unidirectional Context-System Relation

The context C is used in a way described by the function F to adapt the system S to the user's context. This relation can be described as follows: $S = F(C)$.

The user does not benefit directly from that type of context usage, but it possibly allows us to save network resources and therefore perhaps to provide improved service. Hence the user profits indirectly from this type of context usage in the system. The context can be used in two different ways to achieve an improved usage of network resources.

Approach I: Many procedures exist that can be used to optimise the dimensioning and usage of network resources (e.g. network design, traffic engineering, explicit resource reservation [9]). Regardless of the procedure that is chosen to plan and assign the resources, a prediction of the possible expected network access and usage is necessary. The better this prediction is then the better the usage of network resources can be planned.

Context can be used to enrich the information used to predict network access and usage.

Approach II: Network efficiency also depends on how accurate the necessary Quality of Service (QoS) is specified. Often it is difficult for the user or application to determine and specify the QoS needs correctly. If the QoS is set unnecessarily high, network resources are wasted.

Context can be used to refine the necessary Quality of Service (QoS).

3.2 Bidirectional Context-System Relation

In this case the context is used to adapt the system as specified by the function F. In addition the system influences the user's context as specified by the function G. This relation can described as follows: $S = F(G(S))$.

The resulting relation in this case represents a control loop. Obviously some problems arise if it is assumed that the system is allowed to change a user's context. The context represents the (physical) situation of a user. To enable system-initiated context changes, the user has to accept the influence of the system. However, if this modification of user context is limited, controlled and negotiated between user and system it might be acceptable for the user. In the real world there exist examples where such systems are used and accepted. A car navigation system uses the user's context, location and intended destination, to adapt its user interface. In addition, the system modifies the user's context, the location and route. However, in this example the context is not used to improve the usage of network resources, as considered in this paper.

In what follows it is described how context usage with this bidirectional relation helps to improve the Approach I described in Section 3.1.

Approach I: The user's context can be used to enrich the information used to predict network access and usage. If it is considered in addition that a system can influence the context of a user, an additional optimisation parameter exists. In addition to predicting when, where and how network resources are used, the system can actively influence when, where and how resources are used. Of course, this influence will be limited.

Context can be used to enrich the information used to predict network access and usage. Context can be modified to influence the prediction results.

4. Network Resource Usage in Location-Aware Systems

It was shown that systems with bidirectional relation between context and system exist and provide a potential for further network resource optimisation. The question is, how large a performance gain might be and what costs have to be paid to achieve the gain. If these questions are answered, it can be decided under which conditions the hypothesis stated in Section 1 can be accepted. For this analysis a model is considered in the remaining paper with focus on "location" as specific context. While the model is by necessity a relatively simple abstract representation of the real world, it serves well as a basis to develop the bidirectional approach, and is general enough to be easily applied to a wide range of real scenarios.

4.1 Model

It is assumed that the context-aware system is a closed system (see Fig. 1). New users enter and leave the system with a controlled and known behaviour. Therefore the amount of users in the system and the rate at that they enter and leave the system is known.

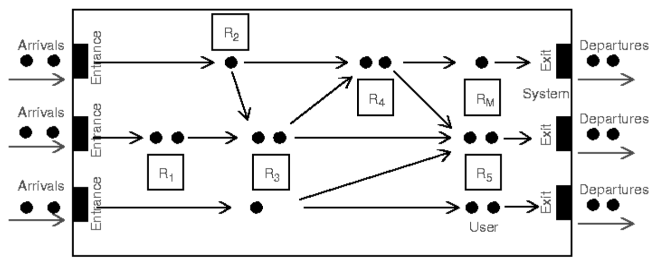

Figure 1 – Context-Aware System

The type of context considered in the model is the location *L(x,y,z,t)* of the users. At each possible location in the system, each user is either using (or trying to use) exactly one network resource *R* or no resource. Within the system *M* network resources are available. It is assumed that all resources are from the same type with equal capacity (e.g. wireless network basestations to access a backbone network with the same available bandwidth). Users that enter the system will follow a certain path $P=(L_1,L_2,\ldots,L_p)$ through the system thus using some of the *M* available network resources in a special order and at specific times before leaving the system.

If a user allows a limited and negotiated influence on his location context, his path *P* through the system can be slightly modified. Therefore different possible sequences for the usage of resources in the system exist for this user. This influence option can be used to improve the network resource usage in the system. To be able to quantify the improvement and the costs of them, *Gain* and *Cost* indexes as described below are defined.

4.2 Gain Index

This index describes the quantity of the gain of a system allowing context modification compared to the same system without influence on users context.

The parameter *N* gives the average number of users in the system waiting or accessing one network resource. If *N* is lower in a system allowing context modification than in a standard system ($N_{cm} < N_{st}$), less resources are necessary. Of course any improvements will be associated with costs that are described later. The gain achieved by allowing context modification can be expressed by:

$$G_N = (N_{st}-N_{cm})/N_{st} \qquad (1)$$

In a system *N* depends on the paths *P*, defined by a sequence of locations *L*, of each user. Thus G_N depends directly on the context “location”.

4.3 Cost Index

The cost for implementing location modification manifests in two parameters: *distance* and *time*. If the locations *L* of a user and therefore his optimal path *P* are modified by the system, the new path might result in a longer travel distance *D* through the system. In addition the new path suggested by the system might result in a longer travel time *T* through the system.

Both parameters can be used as a quantitative measurable indicator how inconvenient the system will get by implementing context modifications. The user will request a compensation for this inconvenience to allow the context modifications. Obviously, it will depend on the type of the investigated system how a user will weight the different parameters. The costs caused by allowing context modifications can be expressed by:

$$C_D = (D_{cm}-D_{st})/D_{st} \qquad (2)$$

$$C_T = (T_{cm}-T_{st})/T_{st} \qquad (3)$$

5. Application Example - Museums Guide

As example based on a specific class of applications is now considered - a context aware guide for museums and galleries. Visitors receive a personal digital assistant (PDA) device at the entrance of the museum. When the user approaches an exhibit, this is detected by the system either implicitly by its proximity to an access point or using some explicit positioning system such as [10]. The location of the user is used as context information in the system. Each sample in the museum provides a transceiver connected to the backbone network of the museum. A database server connected to this backbone provides information about the samples that are displayed. Thus, a user standing in front of a sample is able to receive information about the sample via the transceiver attached to the exhibit on his PDA. The information accessed might be multimedia content including video, audio and/or text. The user walks through the museum on a self-chosen or pre-defined path. After the user visited all exhibits, the user leaves the museum and returns the PDA at the exit. Such a museums guide is for example described in [11].

If in such a guide system too many users access the central database through the same transmitter two different problems might occur. Either the necessary QoS cannot be fulfilled for the users at this location, or some users have to wait to access the system. In this example it is assumed that users have to wait.

If the system allows an influence on the users location context, the system might propose the order in which the samples should be visited such that the occurrence of the aforementioned problems is reduced. This, of course, will result for the user in a not distance or time optimal path through the museum.

It is assumed that the exhibition consists of M samples and thus M network resources. The network resources are represented by the transmitters. Each resource can be accessed by k users simultaneously. Further it is assumed that the samples are arranged along a corridor with an equal distance d between each sample. It is assumed that at each entrance of the system users arrive with a rate of λ and an exponential distribution. The user inter-arrival time is exponentially distributed which is reasonable for most practical cases. Each user will access a resource (watch the exhibition sample) for an exponentially distributed time period $1/\mu$. It is also assumed that the time necessary to move from one sample to the next sample is significant less than the time used to watch a sample.

In the standard system (no context modification) users visit one sample after the other ($P=(L_1,L_2,\ldots,L_M)$). In the system with context modification the system decides which sample a user visits after finishing to watch the current sample. The system sends the user to the sample with the least users currently served (unless the user has already visited this sample). Therefore each user might have a different path through the system ($P=(L_1,L_M,L_5,\ldots,L_2)$). The described system, the standard system without context modification, is shown in Fig. 2.

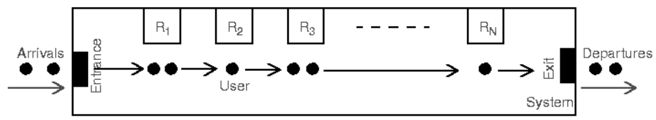

Figure 2 – Context-Aware Museums Guide

To investigate the system allowing context modification, a simulation experiment is used. The goal of the experiment is to determine the gain and cost indexes described in Section 4.

5.1 Experiment Set-up

For the simulation experiment, the museums guide application example described is implemented within the simulator. The standard system and the system with context modification can be used for simulation. The following parameters were used to specify the experiment conditions:

- Number of samples in the museum: $M = 10$
- Concurrent users per resource: $K = 5$
- Distance between each sample: d = 5m
- Resource access rate: $\mu = 0.001\ 1/s$

The simulator was used to determine the performance of the standard system (ST) and the system with context modification (CM). The number of users per resource, travel time and travel distance were measured in dependence of the user arrival rate λ for both systems.

5.2 Experiment Results

The determined values of both systems can now be used to calculate the gain and cost indexes described in Section 4.

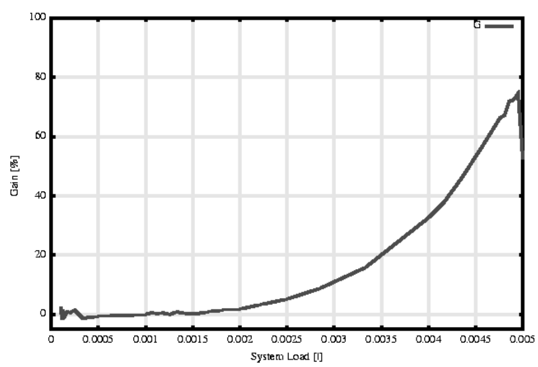

Figure 3 – Gain $G(\lambda)$

Fig. 3 shows the gain $G(\lambda)$ that can be achieved by implementing the location-context modification. For a low arrival rate λ of users, the modification of the location context does not provide an advantage. With increasing arrival rate, the gain then increases continuously. At the point $\lambda = 0.005\ 1/s$, both curves tend to infinity and the gain is zero again. As stated at the beginning of the paper, a price for this gain has to be paid.

The price for the gain is an increasing inconvenience for the users in the system that can be described by the cost indexes for travel distance and travel time.

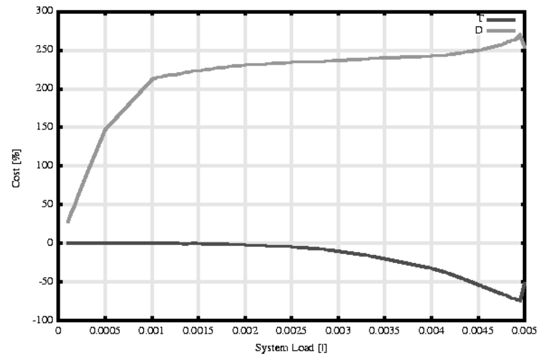

Figure 4 – Cost $C_D(\lambda)$

Fig. 4 shows the two cost indexes according equation (2) and (3). The cost index of the travel distance $C_D(\lambda)$ is increasing fast and is finally nearly stable around 230%. The increasing at the beginning of the curve is caused by the implementation of the location-context modification. For low arrival rates, it is very likely that many resources have the same minimal amount of users queuing. In this case the context modification algorithm selects the closest of these resources. The more loaded the system gets, the more unlikely resources with the same number of users queuing are. To allow the gain shown in Fig. 3 the cost described by $C_D(\lambda)$ has to be paid. For a system load of $\lambda = 0.0025$ 1/s for example, $C_D = 234.6m$ is obtained. That means that for this load the users in the museum have to travel in average 2.6 times further (234.6m instead of 90m).

As it is assumed that the time necessary to walk from one exhibit to the next one is zero, the travel time through the system is proportional to the queue length. The travel time through the system increases for the users as they spend more time in a queue of a resource before accessing it. In our selected example scenario, the cost index of the travel time is negative and is therefore de facto a gain (see Fig. 4).

6. Conclusion

The analysis of the selected example application class represented by the museums guide, shows that by allowing system-initiated location-context modification a gain in the usage of resources can be achieved. This existing gain can be used to

- serve higher arrival rates at the same service quality.
- reduce the necessary resource capacity.

It is also shown that a price in form of an increased travel distance for the users has to be paid. It depends on the perception of the users in the system, the visitors of the museum, if they will accept the introduced inconvenience in form of increased walking distances. The perception of the visitors might be influenced by the price they pay or by how strong the location-context modification can be noticed. In the simulated scenario, a visitor spends in average two and a half hours in the museum. It can be assumed that in this case a visitor will not notice if he walks 235 meters or 90 meters during his visit.

6.1 Hypothesis Verification

In Section 1 the following hypothesis is stated:

If users in a context-aware system allow context modifications by the system, network resource usage in the system can be improved.

A goal of this research was to investigate under which conditions this hypothesis can be accepted. The experiments are based on an application example of a museums guide, representing a generic class of guided applications. In our experiments, only “location” as specific type of context was considered. Thus we cannot conclude that our obtained results can be achieved using other types of context. Taking the results of our investigation and the aforementioned statements into account it can be concluded that the stated hypothesis can be accepted if the considered system is an exhibition/tour guide system and the considered context is “location”. However, this combination captures a large number of possible applications, including what is perhaps the most high-profile context-aware deployment to date [12]. In addition it is important to recall that the possible system gain is associated with costs and depending on the particular system, the costs might be higher than the achievable gain.

7. Acknowledgements

The support of FEXCO Ireland (www.fexco.com) is gratefully acknowledged.

8. References

[1] A. Dey. Providing Architectural Support for Building Context-Aware Applications, Ph.D. thesis, December 2000, Dr. Gregory D. Abowd (advisor), College of Computing, Georgia Institute of Technology.

[2] B. Schilit, N. I. Adams and M. M. Theimer, "Disseminating Active Map Information to Mobile Hosts," IEEE Network, vol. 8, no. 5, pp. 22-32, September 1994.

[3] A. Harter, A. Hopper, P. Steggles, A. Ward, P. Webster. The Anatomy of a Context-Aware Application. Proc. of 5th ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), 1999.

[4] B. Schilit, D. M. Hilbert, and J. Trevor, "Context-aware communication," IEEE Wireless Communications, vol. 9, no. 5, pp. 46-54, October 2002.

[5] E. R. Pedersen. Calls.Calm: Enabling Caller and Callee to Collaborate. Proceedings of CHI 2001 Conference, Seattle, ACM Press, 2001.

[6] D. Chalmers, et al.: "A Survey of Quality of Service in Mobile Computing Environments" IEEE Online communication Surveys, Vol.2, No.2, Second Quarter 1999.

[7] P. Vidales and F. Stajano. Context-Aware Automotive Telematics. First European Workshop on Location Based Services (LBS-2002), September 16th. 2002 London, UK.

[8] T. Strang, C. Linnhoff-Popien, M. Roeckl: Highlevel Service Handover through a Contextual Framework. MoMuC 2003, 8th International Workshop on Mobile Multimedia Communications, Munich/Germany, October 2003.

[9] S. Jha, M. Hassan. Engineering Internet QoS. Artech House. 2002.

[10] P. Bahl, V. N. Padmanabhan. RADAR: An In-Building RF-Based User Location and Tracking System. Proc. of IEEE INFOCOM, 2000.

[11] Margaret Fleck, Marcos Frid, Tim Kindberg, Eamonn O'Brien-Strain, Rakhi Rajani, and Mirjana Spasojevic. "From Informing to Remembering: Ubiquitous Systems in Interactive Museums." IEEE Pervasive Computing Magazine, Vol. 1, No. 2, pp. 13-21, April/June, 2002.

[12] K. Cheverst, N. Davies, K. Mitchell, A. Friday. Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. Proc. of 6th ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), 2000.