Mobility support for distributed collaborative teamwork

Mathias Johanson  
Alkit Communications  
Sallarängsbacken 2  
SE-431 37 Mölndal, Sweden  
+46 31 675543  
mathias@alkit.se

Peter Törlind  
Polhem Laboratory  
Division of Computer Aided Design,  
Luleå University of Technology  
SE-971 87 Luleå SWEDEN  
+46 920 492412  
Peter.Torlind@cad.luth.se

ABSTRACT
The trend in computer and network architecture is towards smaller and lighter devices that are more or less constantly connected to the Internet through wireless network access technologies. For computer supported, cooperative work this offers increased opportunities for distributed collaborative teamwork by enabling mobile users to participate in synchronous and asynchronous information exchange processes. In this paper we investigate the emerging technologies for wireless network access and mobile computing, how mobility support can be built into the software tools used today for distributed collaborative teamwork, and the benefits it gives users. The main focus is on synchronous, real-time communication tools like multimodal teleconferencing, and the perspective is both technical and methodological. Until now text has been a valuable medium for mobile collaboration and our key finding is that user mobility support for video can substantially enhance the possibilities for informal, spontaneous communication between team members. Furthermore, we argue that mobility support for the applications in question is more appropriately implemented at the application level rather than at the network level.

Keywords  
Mobility, distributed collaboration, CSCW

INTRODUCTION
Distributed collaborative teamwork, empowered by state of the art information and communication technology (ICT), promises more efficient work processes, reduced travelling needs, and increased opportunities for personal interactions in many different fields of work. Specifically, collaborative work between geographically distributed teams of engineers and designers has the potential of cutting lead times in product and production development, thereby reducing the cost and increasing the quality of the final product.

Engineering design is fundamentally social, requiring a lot of interaction and communication between those involved [1]. Additionally, good design often relies upon the ability of a cross-functional team to create a shared understanding of the task, the process, and the respective roles of its members. Different locations and time zones complicate communication. Traditionally, in distributed collaborative teamwork, team members are often unaware of remote team members, have difficulties in locating people that are out of office, and have problems coordinating activities and initiating interpersonal communication. It is important to provide tools and methods so that geographically distributed design teams are also given the opportunity to engage in such social interactions that co-located teams are used to [2]. One important aspect is that people are mobile and do much of the design work away from their office. Belloti and Bly [3] respond that systems for collaborative work should be designed to support mobile collaborators.

In the distributed product development project Distributed Team Innovation (DTI) [4], a joint product development effort between Luleå University of Technology, Sweden, Stanford University, California, and Volvo Car Corporation, eight students from Luleå University of Technology and Stanford University were studied during a period of nine months while working together as a distributed design team. The goal of the project was to design ‘virtual pedals’, taking into account that the need for mechanical connections between pedals and actuators has disappeared with the introduction of ‘drive-by-wire’ technology in the automotive industry. The students used high quality videoconferencing and shared whiteboards for meetings and document servers, instant messaging, and e-mail for asynchronous communication.

In the DTI project a web-based collaboration tool called Contact Portal was developed to support informal communication and information sharing in distributed engineering design teams. The Contact Portal has been a valuable asset in the project primarily because it combines several information channels: awareness cameras, instant messaging, SMS, and asynchronous tools such as diaries, e-mail archives, and document servers in one place. The system supports multiple fixed locations [3] (the students at home, in the lab, or at the concept lab). However, since both the Luleå and the Stanford campuses have wireless network (IEEE 802.11b) infrastructures, a student could be available using a PDA or a laptop with a wireless

LEAVE BLANK THE LAST 2.5 cm (1”) OF THE LEFT COLUMN ON THE FIRST PAGE FOR THE COPYRIGHT NOTICE.
connection. To make the users of the system truly ubiquitously reachable, the Short Message Service (SMS) of the GSM mobile telephony network was used.

Findings from the DTI project revealed the need for sophisticated software tools to improve the possibilities for scheduled, formal meetings, but the key to success might very well be the ability to adequately support the communication that happens in between these meetings.

In this paper we use the findings from the DTI project to create a framework for mobility support for distributed collaborative teamwork. We explore different types of devices and software tools for mobile, distributed collaborative work and the information exchange processes they support. Furthermore, we describe the design and implementation of a system that integrates many different video communication tools into a comprehensive framework for distributed mobile collaboration, targeted at engineering and design teamwork.

BACKGROUND AND MOTIVATION
Informal communication is critical to a modern knowledge intensive organisation. These interactions are not pre-planned with a set agenda or a fixed location [5], but instead they take place spontaneously and almost everywhere. Everyday work is filled with these highly informal, accidental, spontaneous communication sessions that have an impact on work processes and task outcome that is sometimes even greater than that of formal communication [6].

When working in a co-located organization, there are many possibilities for informal communication, and issues can be discussed and resolved spontaneously rather than waiting for a suitable, and scheduled, time to make a formal decision [7]. Hence, these spontaneous interactions facilitate frequent exchange of useful information. The awareness of ongoing activity creates a shared knowledge and understanding of other’s activities which enables you to set the context for your own activities, goals and motivations [8,9]. Informal communication between remote team members can be supported via tools for lightweight interaction [10], such as phones, e-mail, instant messaging (IM), shared workspaces, and awareness technologies.

Many awareness technologies use an iconic representation of the availability of people such as Babble [11], ICQ and AOL. Other awareness tools are based on digital images, or live video [2,12,13] but these are mainly designed for desktop systems. Web based systems [4,14,15] are primarily designed for multiple fixed locations and not designed for mobility but can be viewed on a portable device.

Local mobility vs. remote mobility
Luff and Heath [16] classifies mobility into three different categories: micro mobility, local mobility and remote mobility.

Micro mobility is the way in which an artefact can be mobilised and manipulated in a circumscribed domain, for instance by moving a document or a handheld device so that another person can see the context.

Local mobility is used when moving within the company, e.g. visiting another office or going to the coffee machine. Systems that support local mobility are usually interconnected using wireless LAN technology, or by direct interaction between wearable or otherwise mobile devices. Examples of local mobility systems include the proximity based inter-personal awareness tools Hummingbird [17] and Proxy Lady [18]. The Roam Ware [19] system collects information of face-to-face meetings and synchronises information (notes and attendance list of the meeting) from several mobile devices.

Remote mobility is when geographically distributed users interact with each other over a distance using communication technology. Examples of remote mobility systems are mobile phones and the Hubbub [20] mobile instant messenger. In this paper the primary focus is on remote mobility and video communication.

Personal mobility and device mobility
Sometimes a distinction is made between mobility support mechanisms that provide personal mobility (user mobility) and schemes that provide device mobility [21,22].

Personal mobility refers to the ability of end users to access network services from any terminal system irrespective of physical location, and the ability of the network to identify end users as they move. Personal mobility thus relies on some unique identification and addressing of mobile users [23]. Device mobility is the ability of the network to support seamless roaming of mobile devices as their users move around, e.g. like a mobile phone. Consequently, for device mobility it is the devices rather than their users that need to be uniquely identified and addressed.

Device mobility typically represents a higher degree of mobility supporting continuous operation of communication sessions while the users traverse subnetwork boundaries.

COMPUTER AND NETWORK SUPPORT FOR MOBILITY
User mobility support must provide means for roaming team members to participate in both synchronous and asynchronous information exchange processes. Asynchronous collaborative work is typically supported by tools for remote access to shared project data (e.g., file sharing systems, workflow systems). These tools are generally not affected dramatically by the mobility aspect, although security and performance issues may need to be specifically addressed. Synchronous collaboration tools, however, are inherently location dependent and must explicitly provide means for locating a mobile user and arrange for an efficient routing of data packets between the participating hosts’ current locations. The mobility support in this situation can be implemented at different levels. At the network level, a mobile host's current location can be monitored and data packets destined for this host can then either be forwarded from the host's home location, or the
routing system can be updated to reflect the node's current position (as in mobile telephony systems). At the application level, the individual applications can assume the responsibility of maintaining location awareness and efficient distribution of data.

**Mobile IP**

The Mobile IP working group of IETF is developing a framework for extending the Internet to support mobile hosts [24,25]. Mobile IP provides transparent host mobility without requiring modifications to applications. A host is allowed to retain its IP address while roaming between different IP networks. A mobile node (MN) is required to acquire a "care-of-address" (COA) when visiting a foreign network. Network agents at the home and foreign networks are responsible for relaying packets destined for the MN between the home network location and the visiting address through tunnelling. Once a COA has been acquired by the MN from the foreign agent (FA) it must be registered with the home agent (HA) so that the tunnel can be set up. Packets originating from the MN need not be tunnelled (unless they are destined for another MH) but are transmitted in the ordinary fashion with the COA as source address. Although the transparency provided by Mobile IP is highly beneficial in terms of deployment and simplicity it also has negative effects for delay-sensitive applications like real-time multimedia conferencing. Since a packet bound for a MN must be directed to the MN's home location and then redirected through tunnelling to the current location a suboptimal routing path is achieved. This is sometimes referred to as "dogleg routing". Increased transmission delay due to dogleg routing and the tunnelling overhead can seriously impact the interactivity of interpersonal communication. The problems with dogleg routing are partly solved by the route optimization [26] extension of Mobile IP by sending binding updates to inform the sending host of the mobile host's current location. However, this solution has several drawbacks, as discussed by Wedlund et al. [27].

**Application-aware mobility support**

In contrast to Mobile IP an application-aware framework like the one proposed by Inouye et al. [28] can be used. In this model an agent on the mobile node known as a guard is used to detect when a predicate referred to as a quasi-invariant is invalidated. A quasi-invariant could for instance be the association of an IP address with a mobile host. This association is reasonably static, but may change as the host moves to a different location. The guard triggers an event to the applications that have been modified to support mobility, so that appropriate actions can be undertaken in order for the application to continue to work after a quasi-invariant is invalidated. For instance, a teleconferencing application could signal the change of IP address to the participants of the conference session or to a gateway used for distribution of multimedia packets of a session. This type of architecture for mobile networking puts a heavier burden on the application designer than the transparent Mobile IP solution.

The Session Initiation Protocol (SIP) [29] is an IETF protocol used for call setup signalling, providing application-level mechanisms to establish, modify and terminate synchronous multimedia sessions. SIP supports user mobility by transparent name mapping and redirection services. The user mobility functionality present in the SIP protocol makes it possible to initiate synchronous collaboration session where one or many of the participants are mobile. There is no support, however, for ensuring delivery of data packets to a mobile participant of a conference sessions that roams between different routing domains. Neither is there any support for determining the location of a mobile host at session setup time, and hence some mechanism external to SIP must be used for this.

**Multicast applications and mobility**

For multipoint conferencing, an efficient multicast routing mechanism is of paramount importance. It is interesting to note that multicast packet delivery is potentially easier to support in mobile environments than unicast packet delivery. Since a multicast address is not associated with a certain host, but rather represents a dynamic group of hosts, continuous operation of a synchronous conference session can be supported without requiring the mobile host to retain its unicast IP address. When the mobile host roams to a new location a new unicast IP address may be assigned with a dynamic host configuration protocol like DHCP whereupon the multicast group can be re-joined from the new location to continue the session.

Since multicast routing is not supported everywhere on the Internet, specialized gateways known as reflectors or multipoint conferencing units are commonly used. At session setup time these gateways must be configured with the host addresses of the participating users. In case a session includes mobile users the reflectors must be re-configured whenever a participating mobile user's host address is changed.

**Network Services**

Modern mobile phone systems provide data communication services, such as GPRS (General Packet Radio Services). These services allow you to be online all the time, but typically the bandwidth that can be delivered is too low for sophisticated multimedia conferencing applications. Thus, the use of mobile data communication services for collaboration is limited to low bandwidth applications such as asynchronous file sharing tools and text messaging. The emerging third generation mobile communications systems (e.g. UMTS) promise higher bandwidths for data communication. At this time, however, 3G wireless network services are not commercially available in Sweden.

In addition to the second and third generation mobile communication systems, wireless network services based on wireless LAN technology (IEEE 802.11a/b) are being developed and deployed. This type of wireless network service can deliver substantially higher bandwidths compared to 2G and 3G networks, but the coverage is
Due to the limited size of the display on a normal mobile phone it is best suited for synchronous voice communication and simpler asynchronous applications, such as reading e-mail and SMS. The limited interaction provided (generally no keyboard) makes the mobile phone primarily a tool for browsing and reading information. The trend, however, in mobile phone development is towards integration of more and more functions, suggesting an eventual convergence of technology between mobile phones and PDAs.

**DESIGN AND IMPLEMENTATION**

Experiences from the DTI project show that mobility support is needed both for asynchronous and synchronous applications. The team members need to arrange synchronous conferencing meetings - including mobile participants - and each team member also needs asynchronous access to project data more or less ubiquitously. The synchronous tools support the direct real-time collaboration between users, whereas the asynchronous tools support the individual work that is performed in between the synchronous sessions. In the borderland between the asynchronous and synchronous applications are tools that support the transition between asynchronous and synchronous work processes. These tools include meeting scheduling tools, awareness tools that can be used to check if remote users are available for ad hoc synchronous communication sessions and synchronous session initiation tools. Henceforth, we collectively refer to this class of tools as "awareness and transitional tools".

**Awareness and transition support**

Awareness systems are mainly used to support the transition between asynchronous and synchronous modes of collaboration. That is, they assist the users in finding a good time and place to interact and they support informal communication through opportunistic or spontaneous interactions.

Awareness applications supporting the initiation process of informal communication sessions can be divided into three different categories:

- **Location awareness**: Is a certain location (meeting room, collaboration studio, etc) populated by anyone with whom a synchronous communication session can be initiated? This type of awareness can be exemplified by the process of glancing into a co-workers room to see if he/she is available for a meeting.

- **Meeting awareness**: Which meetings relevant to a team of co-workers are available at the moment? (e.g. to overhear a local meeting).

- **User availability awareness**: Information about the current availability of persons with whom synchronous collaboration is desired.

**Location awareness**

With the aim to enhance the sense of working in a shared physical environment, continuously open video links [31] were integrated into the Contact Portal. The idea of the...
awareness cameras was to monitor several locations that the distributed team frequently worked in and provide information about actions of other members. By incorporating visual location awareness information in the web page, the teams only needed a quick glance to know if, or when, it was a suitable time to initiate interaction.

The awareness cameras are based on network cameras with a built-in web server and connected directly to a network as a standalone unit. Figure 2 shows an example of the mobile contact portal. In the DTI project the awareness cameras were often used for the initial contact, by checking the availability of other team members. Then the users switched to another media such as IM, or videoconferencing.

Figure 2 The Contact Portal with the awareness cameras.

Meeting awareness
Within a co-located team we become aware of others’ whereabouts and activities as we walk around in the work environment, overhear conversations in the hallway, or glance into the room of a colleague [32]. Informal workgroup meetings are frequently formed spontaneously with short notice and participants may join and leave these meetings continuously. In a distributed environment, this process can in part be supported by the "location awareness" tools described above. Moreover, by using WebSmile, an RTP to HTTP video gateway [15] we can combine high quality videoconferencing and a web-based interface that shows currently active conference sessions. The web-based interface makes it possible to easily access information about ongoing conference sessions, and to take part of the multimedia content being disseminated, without requiring the installation of specialized synchronous collaboration tools, an example is found in Figure 3.

Since the HTTP protocol is used for relaying the multimedia content of the session, the problems with firewall traversals are circumvented. This makes it easier for a mobile participant of the engineering team to participate in synchronous information exchange processes.

User availability awareness and session initiation
Research on business phone calls has shown that between 60% and 70% of the calls do not reach the intended recipient [2, 33, 34, 35]. It is therefore important to be able to check if a person is available and then easily initiate a communication session regardless of the user’s communication tools (desktop computer, handheld computer, mobile phone). Thus, the initiation of communication should be as easy as calling a person’s mobile phone: as long as you know the phone number the knowledge of the user’s location is unnecessary. However, the service should ideally also be able to tell you if the remote participant is available, to avoid wasting time calling people that are unavailable. In the Contact Portal the ICQ messaging tool provided availability awareness.

The teleconferencing application that has been used supports session initiation using the SIP protocol. Users that are unavailable can choose to hold incoming calls, returning a reason to the caller for not accepting the call. More sophisticated support for filtering incoming calls remains to be implemented.

Synchronous collaboration support
Synchronous interpersonal communication is a real-time process involving two or more participants. Multiple communication channels supporting different modalities can be used simultaneously. In the DTI project synchronous distributed teamwork session were conducted using audio/video teleconferencing in combination with application sharing and shared whiteboards. Mobile users participating using less capable devices, like PDAs or
mobile phones, typically only take part using a subset of the available modalities. Studies of asymmetrical communication show that asymmetries in the modalities supported affect measures of co-presence, awareness, communication effort and collaboration effectiveness [36]. However, the effect of the asymmetries depends largely on the roles of the co-workers and the nature of the collaborative task.

Mobile teleconferencing
The teleconferencing software used in the project is supported on a wide variety of hardware platforms, including PDAs, laptops and workstations (see Figure 1). For multipoint videoconferences, heterogeneity in terminal equipment and bandwidth availability can be overcome by using a layered video coding in combination with an adaptive layered multicast transmission architecture [37]. However, this mode of operation is highly experimental and will require additional development efforts to be fully operational.

The teleconferencing tool supports audio and video communication as well as text messaging, remote camera control and functionality for session initiation and management.

Implementation issues
As previously discussed there are several options available for implementing user mobility for distributed collaboration. The functional requirements that need to be considered when choosing the technological framework upon which to build the mobile collaboration support can be summarized as follows:

- Do mobile users need access to both asynchronous and synchronous services?
- What level of mobility is required? Is the system required to support continuous operation of applications while user's roam across routing boundaries? (cf. personal mobility vs. device mobility)
- What level of user involvement is appropriate in determining the user's current location?
- Is multicast packet delivery needed?

Experiences from the DTI project and elsewhere have revealed that there is not a great need to support continuous operation of synchronous collaboration services, like videoconferencing, while the mobile users are on the move. Typically a mobile user joins a conference session from his/her current location and stays connected to the session from that same location throughout the duration of the session. For this reason, network level mobility support as implemented by Mobile IP is not necessary. Instead an application level approach has been chosen. Since applications like videoconferencing typically require higher bandwidth than what is generally available in ubiquitous wireless networks, synchronous collaboration sessions are not going to be possible in the near future for true mobile users anyway. This restriction means that user mobility support becomes part of the session initiation process. For this purpose the proxy forwarding functionality of the SIP protocol has been used. Whenever a mobile user acquires a new IP address through DHCP from the visited network, the host contacts its home network's SIP server and registers the new address using the SIP REGISTER request. The SIP server then forwards SIP invitations to the mobile host, transparently to the caller. The mobile host's SIP user agent responds with 'accept' or 'decline' SIP messages through the SIP server, to the caller. An accepted call invitation is followed by RTP data exchange directly between the communication endpoints (i.e. the SIP server is bypassed). This avoids the adverse implications of dogleg routing that is typically present in network level user mobility protocols like Mobile IP.

For initiation of multipoint conferencing sessions the situation is a little bit more involved. In case multicast routing is known to be supported at each of the mobile user's possible locations, the "invite to" address of the SIP INVITE message is simply a multicast address and the session initiation proceeds in the same way as for point-to-point call setup, for each user being invited. In case a reflector is needed the "invite to" address is the address of the reflector. After each prospective participant have been invited through SIP signalling, the host initiating the session configures the reflector to relay RTP data between all hosts that accepted the invitation. This configuration is currently performed using a proprietary protocol, but could be implemented using SIP messages.

DISCUSSION
A clear trend in computer and network architecture is towards mobile wireless systems that are constantly connected to the Internet. Although the increase in CPU performance is rapid for stationary computers recent work indicate that advances in battery technology and low-power circuit design will not alone meet the demands of increasingly complex mobile applications [38]. We can therefore envision a future situation much in agreement with the present situation, characterised by considerable heterogeneity in terms of resource availability for different operating conditions. When designing a multimedia communication system for distributed and mobile teamwork it is hence of utmost importance to address scalability issues at different levels, including network bandwidth consumption, processing requirements and data presentation capabilities. The applications must be able to adapt to widely different conditions depending on the context of operation. This implies that the applications must provide a graceful scaling of the quality of the multimedia content being communicated and possibly also to support modality changes in ongoing collaboration sessions in response to dynamics in resource availability. For engineering and design sessions that rely heavily on sharing of multimedia objects, scalable encodings are needed. Improved video and 3D formats supporting multiple description encodings is a key component in the successful realisation of truly mobile collaboration systems.
in highly heterogeneous computing and networking environments.

CONCLUSIONS
In this paper we have investigated the need for mobility support for distributed collaborative teamwork. Furthermore, we have identified some of the emerging enabling technologies that will make future distributed collaborative work possible almost ubiquitously.

From related work and our own experiences we conclude that informal communication is a vital component in collaborative work and that this type of communication traditionally has been troublesome to support in CSCW systems. By introducing mobile awareness and mobile video conferencing we argue that the possibility of supporting informal communication is increased. More extensive behavioural studies are needed, however, to fully substantiate this claim.

We investigate the emerging computer and network technologies that will facilitate the development of truly ubiquitous collaboration systems. Specifically, we argue that the implementation of mobility support for collaboration systems is more appropriate at the application level, using protocols like SIP, rather than at the network level using protocols such as Mobile IP. Possibly a combination of network level and application level mobility support will be chosen in the future when designing mobile collaboration systems.

We exemplify the use of mobility support for distributed collaborative teamwork by reporting experiences from the design and implementation of the prototype system used in the DTI-project for collaboration among engineers and designers. As a proof of concept we also developed several tools to support remote mobility, the conference awareness tool, in which a remote user can browse all ongoing conferences, and mobile videoconferencing with SIP based initiation.

Apart from the need for more sophisticated wireless access networks and improved mobile devices, we conclude that one of the keys to realising a truly ubiquitous mobile collaboration system is to develop systems and protocols that are adaptable and scalable to heterogeneous conditions in terms of bandwidth availability and end-user equipment capabilities. This will require continuous research in scalable media encodings, adaptive network flow control algorithms and behavioural studies in asymmetric interfaces for collaborative work.

ACKNOWLEDGEMENTS
Support for this research is provided by Alkit Communications AB and the Polhem Laboratory at Luleå University of Technology. The Distributed Team Innovation project was funded by Volvo Car Corporation and Volvo Aero Corporation. Special thanks to the students in the DTI project and our colleagues at the Center for Design Research at Stanford University.

REFERENCES