System Support for Location-dependent Applications over Indoor Wireless Networks
9807-04

Proposal for 1998-1999 Funding

Hari Balakrishnan and John V. Guttag

I. Motivation

Until now, most research in wireless networking has focused on link-layer, channel access, mobile routing and transport protocols. Little attention has been paid to developing novel location-dependent applications for wireless mobile systems, especially on a large scale involving hundreds of devices. Much of what has been done attempts to be location transparent and focuses solely on the transport of information (e.g., [1]). We intend to focus much of our attention on applications with a particular emphasis on applications that exploit knowledge about the geographical or physical locations of users and devices. We believe that the deployment of a ubiquitous robust and scalable general purpose system infrastructure for location-dependent applications will foster the development of novel and compelling applications and accelerate the deployment and use of wireless technologies.

Our goal is to design and deploy such an infrastructure and develop a suite of applications tailored to the office and campus environment to validate our architecture. We propose to explore the space of in-room (infrared) and in-building (radio frequency) wireless technologies and a variety of devices (laptops, PDAs, office equipment, rooms, etc.) as our implementation platform. As an important side-effect, we hope that our work will lead to the deployment of production wireless networks based on IR and RF that everyone in our labs can use.

II. Research Approach

There are three fundamental challenges to meeting this vision of a wirelessly networked office campus with mobile users and devices: heterogeneity, changing conditions and scale.

* Heterogeneity: There is a tremendous degree of heterogeneity in end devices and their communication capabilities. Unifying numerous disparate components into a single network infrastructure to perform useful tasks is therefore difficult. Unfortunately, it is insufficient to simply ensure that all devices have IP addresses and an IP ``dialtone'' to communicate with --- this only ensures that they can be reached at the network level, and

says nothing about how they can interact together to function at the application level. What is needed are application-level protocols that integrate these diverse end devices. It is important to ensure that these protocols are lightweight and minimalist, so that they can run even on computationally impoverished devices, and extensible so that new services and applications can be developed easily.

Our approach to dealing with variation is making the components of the system self describing. That is, with minimal prior configuration, components and devices must be able to learn about each other and their characteristics and inter-operate. We will explore different approaches and protocols that enable this.

* Changing conditions: The wireless and mobile environment is often characterized by rapidly changing conditions. This is caused by the nature of wireless media and by the new environments that devices and computers encounter as users move around. Handling this requires protocols and application support software that adapt to external conditions.

* Scale: An average office building has hundreds of people and thousands of entities that need to be accessible via a building-wide network. Thus, it is important to ensure that our system scales to such large numbers of communicating entities. This includes graceful degradation under heavy load, suitable access to the network and entities in the system as the system grows, and the elimination (or reduction in the number) of hot spots in the system.

There will be two intertwined themes that will play major roles in our research and system design: learning and activity. After discussing these aspects, we describe some important problems in security issues that we intend to address.

* Learning: This is the process by which an entity learns about and becomes a part of an existing portion of the system. This can happen both on device startup and (more commonly) when a device or user wants to communicate with a part of the system that it hasn't interacted with in the past or that has changed in relevant ways since the last interaction. One way of thinking about learning is to view it as a self-configuration mechanism that allows devices to deduce what communication facilities are available and what other entities can be reached. In most distributed systems today, configuration schemes are invoked only on initialization or upon component failure. In contrast, in our system, entities constantly probe the external environment to learn about new devices and services. In addition, they also cache properties of these environments for use both by themselves and by other entities. This property of continual shared learning amongst entities makes failure recovery and mobility, which are treated as special cases in most current systems, part of normal operation.

Learning is facilitated by making entities self-describing, so that other entities can learn about their capabilities. There are multiple ways of doing this; we intend to design various solutions and to evaluate the trade-offs between them.

One approach is to design a protocol between devices that allows them to query each other, tailoring their subsequent queries to the responses they obtain for previous ones. This is akin to their playing ``twenty questions'' to deduce the capabilities of each other.

Another approach, which we detail in this proposal, is to associate an environment with each entity in the system. Every entity in the system --- entities include physical devices such as desktops, laptops, PDAs, printers, copiers, wireless cameras, etc., the services they offer, and specific locations like seminar rooms, classrooms, cafeterias, etc. --- exports an environment to the rest of the systemin a standard representation. This environment defines the way in which components in the system learn about each other and their capabilities, as well as communicate with them. We note that it is only the format of the exported environment that is pre-defined; the specific environmental components are fully customizable and depend on the characteristics of the entity.

An integral part of learning is the dissemination of their current environments by entities in the system. This must be done without consuming excessive resources or introducing instability. Drawing from work done in the context of multicast applications on the Mbone [5], we intend to develop scalable protocols that use periodic, soft-state messages that are self-refreshing to disseminate this information. We will use techniques such as multicast damping and suppression, coupled with message aggregation to enhance the scaling behavior and stability of these dissemination protocols.

A key research issue to be solved is the representation format of environments in the system. On the one hand, each entity's environment contains rich semantic information that needs to be disseminated to and understood by other entities. On the other hand, we cannot force a rigid semantic structure on exported environments. Therefore, our approach is to use a flexible representation based on tuples of attribute-value pairs. A small number of attributes are well-defined and understood by all entities, which provides the necessary system-level communication ``dialtone.''

* Activity: Once an entity has learned about and become acclimatized to its immediate environment, it starts interacting with it --- we call this "activity." In a dynamic environment with a large number of entities, many of which are mobile or changing as the environment around them also changes, current approaches to communication that involve the knowledge of end-point IP addresses (and/or DNS names) and the establishment of connections are cumbersome and require extensive system support (e.g., DNS registrations for all entities, TCP for reliability, schemes like Mobile IP for mobility support, etc.). It is on top of such primitives that application-level communication occurs today, typically involving the establishment of a connection before any useful information can be exchanged.

In contrast, the approach we intend to explore directly uses the environment representation as the addressing mode for entities to communicate and exchange data. (This approach is related to work done by Van Jacobson in the context of adaptive Web caching [3] and has recently been discussed in an ISAT study group on Simple Systems as a promising approach [2]. We intend to collaborate with members of that study group in the future.) Here, if you want a piece of data whose name you know (in our context, this would be based on the environment with which we are interested in communicating), you send out a packet with that name. This packet diffuses through the network in the direction of the data provider (you don't know where that is, since you haven't established any connection with it) and eventually you get a response either from the provider or from another entity that has the data cached. We intend to thoroughly explore the effectiveness of using such ``data diffusion'' as a communication primitive in our system, obviating the need for heavyweight protocols like Mobile IP or TCP which are ill-suited to our location-dependent applications.

As an example of how this approach based on environment representations and data diffusion eases communication, consider the following example. Imagine a security surveillance system in a 10-floor building with 10 cameras placed on each floor. Suppose you want to send a message to a subset of these cameras asking (reliably) for the last three images they have captured, with the subset being the cameras at the west entrance of every odd floor. Furthermore, you want them to reply only if they have seen any people in their view. A little thought shows that implementing this using today's TCP/IP networking primitives is cumbersome and leads to inefficient solutions. For instance, performing a DNS lookup for every camera and establishing individual TCP connections with them incurs high overhead and latency. Alternatively, sending the request to all cameras by making them subscribe to an ``all-cameras'' multicast group and having them respond does not use wireless bandwidth efficiently. Finally, pre-defining multicast groups limits the flexible addressing of subsets of system entities (especially if cameras may be mobile) and could lead to a combinatorially enormous number of active multicast groups in the system.

In contrast, our system would send out a packet with a template in its tuple-based environment specification --- any receiver that matches this template (including wild-card entries) sends a reply to the originator (itself a template and not an IP address, for it could be mobile). The request environment template diffuses through the system and eventually reaches the relevant cameras, which then reply if they meet the request criteria. Doing this in a bandwidth-efficient manner is a problem we will have to solve so that entities in the system are able to communicate asynchronously without prior rendezvous.

This template-based routing would be facilitated by treating certain entities (e.g., cameras or rooms) as transponders that continuously export their environments to their immediate neighborhoods. Entities in the system cache information about the directions in which entities matching various templates are to be found. Then, when a packet arrives intended for recipients matching a certain environment template, a cache lookup allows forwarding to occur. In the absence of cached information, we perform a local broadcast and hope that this information is cached elsewhere. The design of an efficient protocol mechanisms that achieves scalable template-based application-level routing is a key research problem. We expect to build on recent work done in network support for Web caching to accomplish this [4].

We note that while we do not explicitly use multicast routing to achieve this, our solutions will extensively leverage previous work on multicast routing protocols for many-to-many communication. We intend to explore two implementation strategies: (1) over a conventional IP network using tunneling, and (2) using an Active Network infrastructure (ANTS toolkit) [6]. Ideally, we will be able to compare and evaluate both implementations, which can coexist in the same system because of the common environment representations they will use.

We emphasize that both learning and activity simultaneously occur during normal operation. In particular, the former is a constant feature, helping entities assimilate information about new and changing environments, enabling adaptation to changing external conditions.

Security Issues: There are several interesting research issues related to authentication and security that we will address in this work. We outline some of them below.

The dissemination of entity environments must be done in a secure manner for devices and services that want to advertise their environments only to authenticated entities. The protocols that enable self-describing devices have to be authenticated to prevent impersonation.

Location-dependent security mechanisms, which allow access to certain data only if the user is in a certain geographic location, are important for some applications. We propose to investigate this using credentials that users (or agents acting on their behalf) obtain when they move into the authorized locations. Mechanisms to revoke these credentials when they leave are also needed. An example application that uses this is a secure briefing in a meeting room, where all participants in the room automatically view the materials presented online, but no one outside the room has access to them.

Anonymity and privacy are useful features in some applications. There are services in the system that users would want to use without anyone knowing their identity. For example, employees in an organization may not want their employer knowing that they extensively use certain facilities or to continually track their geographical whereabouts in the building. While a number of these issues are social in nature, giving them some thought will help us gain a better understanding of what system support is meaningful to provide. Our goal will be to provide flexible mechanisms for describing and enforcing policies, not to design polices.

III. Demo Applications

A large number of useful applications are enabled by designing a system that meets the above goals. Sample applications that will test our design include:

* Intelligent meeting rooms: When people get together in a conference or meeting room with their laptops, PDAs, and other networked devices, we intend to automate the process of setting up the network and enabling connectivity to the outside world. Today, this process involves frustrating human intervention and requires knowledge of arcane configuration parameters to set things up. We would like to demonstrate a user walking with a PDA into a meeting or conference room and immediately obtaining the seating map and identities of all the participants in the room (with automatic exchange of electronic business cards). Similarly, walk into a cafeteria, and on your PDA appears the menu and specials of the day (and maybe even recommendations from those who have already dined).

* Intelligent device control: Control A/V equipment, VCR's, cameras, etc. via a PDA, laptop or desktop computer using this system. Demonstrate authenticated and secure access to system entities in this application.

IV. Relevance to the NTT Theme

For the foreseeable future IR and RF technologies will dominate communication to and from inexpensive mobile devices. Unfortunately, the laws of physics make it hard for such communication to run at petabit rates. Nevertheless, we believe that the work proposed here is relevant to the petabit theme.

The current (and almost certainly, future) Internet landscape will not be dominated by a single networking technology. Rather, heterogeneity will pervade its infrastructure. This implies that petabit-per-second backbones will have to coexist and often feed traffic into lower-bandwidth loss-prone wireless links before they reach end users. In addition, there are likely to be numerous servers disseminating data from behind wireless links. We therefore believe that the proposed research will contribute the understanding the issues that arise when technologies with radically different characteristics coexist in the same system. Our work will provide a graceful gateway between high-speed backbones and lower-speed devices connected over impoverished networks. We look forward to collaborating with researchers at NTT in understanding these aspects of bandwidth heterogeneity in the future.

We also intend to identify the correct security and abstractions and primitives for our system, collaborating with and leveraging the expertise of researchers at NTT and at MIT in the area of network security. We believe that these issues are relevant in many network scenarios, wireless and wireline, and that our findings will be relevant in both contexts. In particular, we would like to collaborate with Dr. Ono's group at NTT Multimedia Communication Labs on these topics.

V. References

[1] Balakrishnan, H. Challenges to Reliable Data Transport over Heterogeneous Wireless Networks. PhD Thesis, University of California at Berkeley, 1998.

[2] Estrin, D., et al., Simple Systems. ISAT study group, 1998.

[3] Jacobson, V. How to Kill the Internet. Talk at the SIGCOMM 95 Middleware Workshop, available from http://www-nrg.ee.lbl.gov/nrg-talks.html, August 1995.

[4] Legedza, U. and Guttag, J. Using Network Level Support to Improve Cache Routing. In Proc. 3rd Int'l WWW Caching Workshop (June 1998).

[5] MBONE Information Web. http://www.mbone.com, 1998.

[6] Wetherall, D., Guttag, J. and Tennenhouse, D. ANTS: A Toolkit for Building and Dynamically Deploying Network Protocols. In Proc. IEEE OPENARCH (April 1998).