The Bay Area Research Wireless Access Network (BARWAN)

Randy H. Katz, Eric A. Brewer, Elan Amir, Hari Balakrishnan, Armando Fox, Steve Gribble, Todd Hodes, Daniel Jiang, Giao Thanh Nguyen, Venkata Padmanabhan, Mark Stemm

Electrical Engineering and Computer Science Department
University of California, Berkeley, CA, 94720-1776
http://www.cs.Berkeley.edu/~randy/Daedalus/BARWAN/BARWAN_index.html
http://daedalus.cs.Berkeley.edu/

ABSTRACT

Wireless data services have thus far been more promising than successful. We believe that future mobile information systems must be built upon heterogeneous *wireless overlay networks*, extending traditional wired and internetworked processing "islands" to mobile hosts over coverage areas ranging from in-room, in-building, campus, metropolitan, and wide-areas. In this paper, we describe a new wireless data networking architecture that integrates diverse wireless technologies into a seamless internetwork. It is being implemented in the BARWAN testbed in the San Francisco Bay Area.

1. INTRODUCTION

Given the recent award of the PCS licenses, combined with the rollout of digital direct broadcast satellite services, the digitalization of cellular services, and the deployments of cellular packet services, alternative wireless data services are poised to proliferate. Unfortunately, network planners continue to think in terms of homogeneous wireless communications systems and technologies, providing either low bandwidth connectivity over the wide-area or high bandwidth connectivity over a small area.

We believe that future mobile information systems must be built upon heterogeneous *wireless overlay networks*, extending traditional wired and internetworked processing "islands" to hosts on the move over a wide area. For example, a high bandwidth in-room infrared (IR) network can be "overlaid" with a more moderate bandwidth radio frequency (RF) network to provide connectivity between rooms. Or a low-tier/low mobility in-building PCS system could be overlayed with a high-tier/high mobility PCS system to provide cost effective connectivity bridging the local and wide areas. Figure 1 illustrates this concept.

To verify this concept, we are creating a wireless overlay network testbed, spanning from in-building to direct broadcast satellite systems. We are developing pilot multimedia applications to drive the design and validation of the interfaces between applications and the network. Our goal is to demonstrate a scalable architecture that can support wireless access across multiple overlay networks while delivering high levels of end-to-end performance to applications.

The rest of this paper is organized as follows. In Section 2, we describe wireless data technologies and their strengths and weaknesses for supporting bandwidth and latency-sensitive applications. Our wireless overlay architecture is presented in Section 3. Mechanisms for network management and handoff are described in Section 4, and our approach for application support services is given in Section 5. Section 6 presents the current status of the BARWAN testbed and our summary and conclusions.

2. WIRELESS DATA TECHNOLOGY OVERVIEW

The success of wireless data networks has yet to be proven. It is estimated that there are no more than 100,000 wireless data users in the U.S. today. This is in comparison with over 20 million cellular telephone users and 30 million paging subscribers! To yield ubiquitous connectivity, what is needed is a wireless internetwork formed from multiple wireless overlays interconnected by wired segments. For many emerging applications, (near) real-time audio/video and collaborative support for subscribers on the move is an absolute requirement. Table 1 identifies performance characteristics of representative wireless technologies. It also indicates approximate video and audio performance.

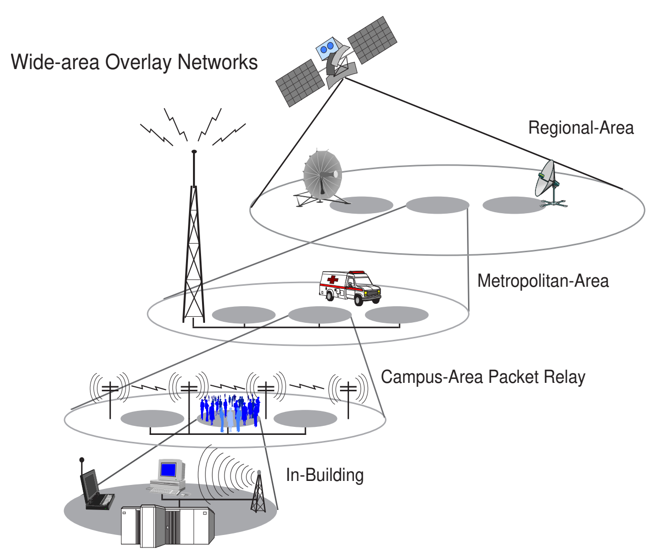

Figure 1 Wireless Overlay Networks

Type of Network (Coverage)	Bandwidth	Latency	Mobility	Typical Video Performance	Typical Audio Performance
Room/Building (Radio Frequency Infrared)	> 1 Mbps RF: 2-20 Mbps IR: 1-50 Mbps	<< 10 ms	Pedestrian	2-Way, Interactive Full Frame Rate (Compressed)	High Quality 16 bit samples 22 KHz rate
Campus-Area Packet Relay	≈ 64 kbps	≈ 100 ms	Pedestrian	Medium Quality Slow Scan	Medium Quality Reduced Rate
Wide-Area (Cellular, PCS)	19.2 kbps	> 100 ms Highly Variable	Pedestrian/ Vehicular	"Video Phone" or Freeze Frame	Asynchronous "Voice Mail"
Regional-Area (LEO/VSAT DBS)	Asymmetric Up/Down 100 bps to 4.8 kbps 400 kbps	>> 100 ms	Pedestrian/ Vehicular Stationary	Seconds/Frame Async Video Play-back	Asynchronous "Voice Mail"

Table 1: Characteristics of Alternative Overlay Technologies

Marketing literature focuses on bandwidth, but there is little discussion of other metrics are equally important, such as latency, expected packet loss rates, probability of packet retransmission, and bandwidth per cubic foot. For example, a 10 mbps Ethernet typically supports 50 users; a 1 mbps in-room IR network designed to support 5 users collocated in the same room yields just as effective performance. Certain applications, such as web browsing, are somewhat tolerant of latency. Others, like decoupled mail access, can operate quite well even in a network environment characterized by very high latencies. Consider the following wireless technologies:

- In-Room Infrared: Infrared technology is attractive for providing wireless connectivity for well defined physical spaces like offices and meeting rooms (typically 30 m x 30 m). Commercially available diffuse IR devices achieve 1 mbps. Observed latencies are comparable to wireline networks. Research devices have been demonstrated at 50 mbps [4]. This allows high quality compressed video and audio. Directed IR interfaces are expected to be inexpensive and ubiquitous in future PDAs and laptop computers.
- In-Building Radio Frequency: Numerous wireless local area networking products are commercially available using spread spectrum in the unlicensed bands. They yield bandwidth in the 1-2 mbps range, but whether this bandwidth is reused among adjacent cells varies among implementations. Quality compressed video and audio can be achieved if user densities are modest. A typical cell size is a several hundred feet—enough to cover perhaps half the floor of a modern university building. Picocellular architectures with much smaller cell sizes are more interesting in terms of aggregate bandwidths, but today's economics makes the approach expensive.
- Campus/Metropolitan Area Packet Relay Networks: Packet radio systems forward packets between mobile hosts, infrastructure radios, and wired access points. Consider the Metricom Ricochet Network deployed in the S. F. Bay Area. Radio link performance is in the range of 100 kbps, but this must be shared among all users within the cell, typically 0.2 to 1 mile in diameter. Latency is a function of the number of radio hops between the mobile host and the wireline connection. 2-3 hops, at approximately 20-40 ms per hop, are typical.
- Wide-Area Packet Switched Data Networks: Cellular Digital Packet Data (CDPD) is a wide-area data overlay to the analog cellular system. The system is designed for instantaneous transmission rates at 19.2 kbps, but this will be difficult to sustain. Pricing remains high, in the range of $0.10-0.18/KByte. Latencies will be highly variable, depending on the cellular voice traffic.
- Regional-Area Satellite Data Networks: There are numerous low earth orbiting (LEO) satellite proposals planned for the latter part of this decade, offering low data rate services. Deployed Direct Broadcast Satellite (DBS) and Very Small Aperture Terminal (VSAT) services provide shared high data rate downlinks. These incur high latencies due to geosynchronous orbits. For DBS, the uplink is asymmetric, over non-satellite links.

3. GATEWAY-CENTRIC NETWORK MANAGEMENT

3.1. Overlay IP

The emerging specifications for Mobile IP [4] and the existing TCP/UDP protocols provide the basic network and transport layers upon which overlay and wireline networks are integrated. However, Mobile IP has not been conceptualized as providing simultaneous connectivity over multiple independent wireless subnetworks. Overlay IP is our extension to Mobile IP that provides this capability

Figure 3 illustrates the concept. The Mobile Host (MH) has an IP address associated with each of its network interfaces.

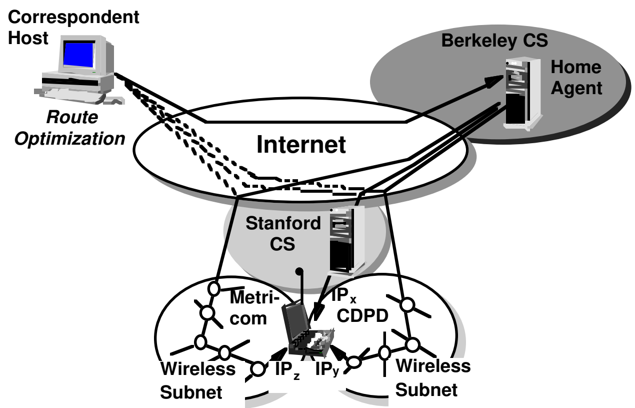

Figure 2 Overlay IP Architecture

Its wired Ethernet network interface defines the MH's home network, in this case, Berkeley CS. As in Mobile IP, packets are sent from a Correspondent Host (CH) to the home network of the MH. A Home Agent (HA) forwards the packet to a foreign agent (FA) in the subnetwork in which the MH is visiting, such as Stanford CS. Note that by using route optimization mechanisms as proposed by the Mobile IP working group, these route bindings can be performed at the CH directly.

Some subnetworks implement their own support for mobile roaming, such as in Metricom or CDPD. Nevertheless, the MH has a unique IP address in each of these, and packets can still be forwarded from the FA to the appropriate "care-of" address within the appropriate subnetwork.

The HA and the MH must monitor end-to-end network performance and collaborate to determine which among the possible subnetworks is the best one to carry a given packet.

To be effective, an architecture for wireless overlay networks must address the following:

- *Seamless Integration of Overlay Networks*: Overlay IP provides the basic routing substrate, but algorithms for choosing the best subnet, and handing connectivity over to it, are still needed. Overlay IP can be viewed as Mobile IP plus policy-driven mechanisms to choose which subnetwork over which to deliver a particular packet.
- *Support Services for Mobile Applications*: Handover across overlays changes an application's network bandwidth and latency. A new API is needed to allow applications to initiate handovers, to determine their current quality of connectivity, and to gracefully adapt their communications demands to this connectivity.

3.2. The Gateway-Centric Architecture

Conventional wireless networks place the mobile host at the center, with gateways to other networks around the boundaries. Such an architecture requires that the wireless network be homogeneous. Our *gateway-centric* approach places the subnet-to-subnet gateways at the center of the architecture. Diverse networks are integrated through software that mediates between the mobile host and the networks it could connect to, supporting the mobile host as it roams among the multiple wireless networks.

Most wireless networks, deployed by competing service providers, provide little controllability above the subnet. We call these *black pipes*, because packets transfer between the wireline gateway at one end and a wireless link to/from the mobile host at the other with no control over routing, priority, quality of service, etc. *Cooperative networks*, on the other hand, provide greater visibility of the network control functions to the gateway as well as application-level software, enabling management software to balance the network load among the cooperating networks or to choose a particular subnetwork for priority applications, etc.

3.3. The Architecture in More Depth

Figure 3 provides a conceptual architecture for wireless overlay internetworking and mobile applications support services. Working bottom up, the layers and the functionality they provide are described in the following subsections.

Wireless Overlay Subnetwork Layer

This corresponds to the physical, data link, and subnetwork layers, representing the wireless subnetworks to be integrated. These interface to the rest of the architecture through Overlay IP. The details of the underlying physical channels, media access, link-level, and routing protocols are not always visible above this layer. However, for in-building subnetworks, we fully expose the underlying mechanisms, especially those that control handoff.

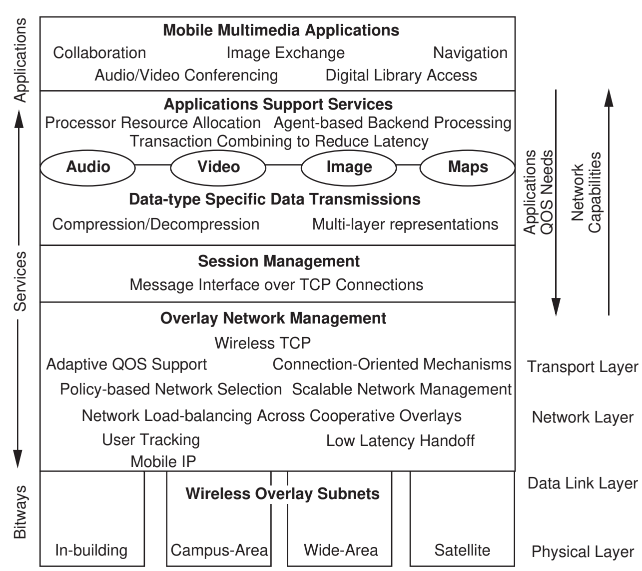

Figure 3 Layered Architecture

Overlay Network Management Layer

This layer corresponds to the network and transport layers, with extensions to manage the multiple overlays. It builds an internetwork on top of the independent wireless subnetworks. It handles the routing of packets across heterogeneous subnets, while choosing the most appropriate subnetwork to provide end-to-end connectivity given the application's specifications for quality of service.

Existing handoff algorithms have been designed for homogeneous wireless subnetworks. Algorithms to support handoff between heterogeneous subnets must base their handoff decisions on application-specified policies as well as periodic measurements of the quality of the underlying network connectivity between the MH and the HA. These can constrain feasible handoffs. Signal quality, bit error rates, and the resulting probability of packet loss and retransmission must be considered as part of the handoff decision process.

Maintaining multiple powered-on network interfaces consumes precious power. To extend battery life, only one network interface is normally active. Others are powered up periodically to determine the quality of their connectivity. A trade-off exists between the frequency of determining link quality, the power consumed by the operation, and how up-to-date the information is about the network state to drive handoff decision making.

Within the network management layer, existing protocols must be extended to improve reliable transport over wireless links. Since entire subnets are "black pipes," we must depend on end host adaptation to changes in the network characteristics, rather than strong support from the network, such as resource reservations.

Session Management Layer

This layer provides applications-independent session mechanisms, such as message-oriented interfaces built on top of a single TCP connection.

Data Type Specific Transmission Layer

This layer provides specialized functions for applications to manage the transport of type specific objects. Multimedia applications process audio, video, image, and other semantically rich data types. Section 5 describes the mechanisms we are implementing to support alternative representations and compression/decompression schemes for commonly encountered data types for multimedia applications.

Applications Support Services Layer

This layer provides resource management and distributed processing services that allow applications to migrate computation between the mobile host's environment and the wireline infrastructure.

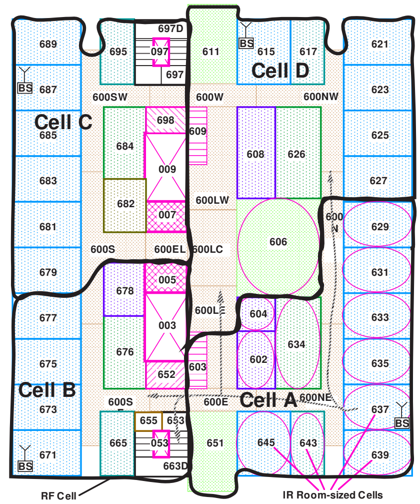

Figure 4 Exploiting Geographical Constraints to Assist Hand-off Processing

Mobile Multimedia Applications Layer

This layer consists of mobile applications exploiting the underlying services and network management functions.

4. OVERLAY NETWORK MANAGEMENT

4.1. Low Latency Mobile Handoff

Network management exploits the location of users and their physical environment to achieve low latency handoffs. Particularly inside buildings, it is possible to use the layout to localize the collection and analysis of tracking information, and to drive the handoff algorithms. For example, consider a mobile host receiving a continuous media packet stream while moving through a building. Its path is constrained by the building's physical layout. Figure 4 shows a floor of a typical building, covered by four RF cells. It is not possible to traverse directly from Cell A to C.

Existing Mobile IP implementations require high latency operations to communicate location updates back to the MH's home network. We have implemented a handoff strategy based on IP multicast to reduce this cost [5]. By tracking the speed and trajectory of the MH, and with knowledge of the geographic layout of base stations, we form multicast groups containing the mobile host's current base station as well as a small number of base stations that the mobile host can reach in the near future. In Figure 4, the base station in cell A could be formed into a multicast group with those in cells B and D. Packets for a mobile host in cell A are simultaneously delivered to cells B and D. As the MH comes within range of one of these, the MH uses the relative strength of base station beaconing signals to choose when to register with a new primary base station. The multicast groups are then reconstituted as the MH moves. The new base station already has the tail of the packet stream destined for the mobile device.

Support for "Vertical" Cooperative Handoffs

Conventional wireless handoffs are *horizontal*, i.e., within a homogeneous wireless subnet. The mobile host or associated base station detects a degraded signal as it reaches the fringe area of its cell. The mobile host listens for beacon signals from base stations in adjacent cells, choosing to register with the cell with the strongest signal.

Vertical handoffs allow mobile hosts to roam between overlays. The mobile host, or higher level network management, determines when to switch the connections to an alternative overlay network, driven by signal quality, network load, or the costs of using one overlay versus an alternative.

For in-building networks, we selectively violate strict network layering to expose information from below the network layer to higher level overlay management software. The allows wireless subnets to cooperate to reduce vertical handoff latency as well as to better share connectivity and network loading between alternative overlays. Our cooperative overlays are formed from independent in-building IR and RF physical networks. They are treated as logically independent networks that can only be integrated through overlay management. The multicast method described above can span base stations for different wireless subnets.

Consider a group of mobile users in a meeting room. The in-room IR network provides the preferred connectivity. But if its bandwidth becomes oversubscribed, resulting in poor service to the mobile hosts in the room, some hosts can be forced to handover to the RF network. Since the overlays are cooperating, new connections can be set up in the RF overlay in advance of actual handoff using the same multicast strategy described above. This ensures that the handoff is executed with low latency.

Connection-Oriented Services for Mobile Hosts

The standard TCP protocol interprets lost packets as congestion, employing mechanisms that unnecessarily decrease the rate of transmission. Our approach caches the unacknowledged TCP packets in the base station while performing aggressive local retransmissions over the wireless link and managing duplicate acknowledgements from the mobile host [2]. It requires no changes to TCP's semantics nor to any implementations of TCP in the wireline portion of the network, except at base stations under our control.

Figure 5 shows the performance of this "snoop" approach to reliable transmission over wireless. We injected a single packet error every 256 Kbits transmitted. While errors in wireless channels tend to be bursty, the controlled errors we introduced represent a relatively low error rate for real wireless links. The figure shows the speed up and the predictable latencies that can be achieved when local methods are used to fixed local transmission problems, rather than relying on TCP's end-to-end mechanisms.

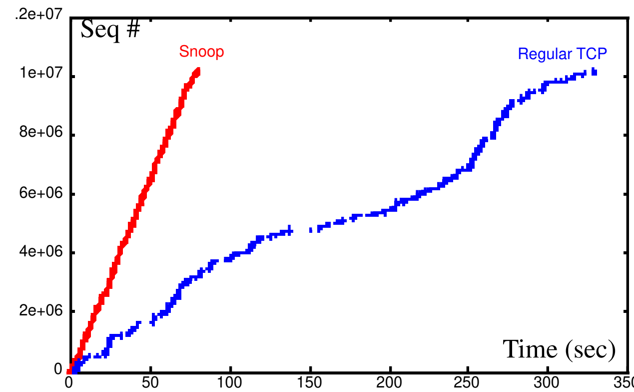

Figure 5 Snoop TCP versus Conventional TCP

5. APPLICATIONS SUPPORT SERVICES

5.1. Network Services for Applications

Mobile applications demand access to the same remote computing power available from desktops. Where network performance is poor, we use network servers to execute agents on behalf of mobiles, and to perform application- and data-specific compression before transmission over the wireless link. Agents mitigate poor latency by reducing the number of link crossings, while compression increases the link's effective bandwidth. Both of these strategies can require substantial cycles from the network servers.

5.2. Uniform Architecture for Applications Support

Our development toolkit supports type aware mobile applications. The toolkit dynamically adjusts the application's bandwidth and latency expectations to the capabilities of the network. Each mobile device has a *proxy*, responsible for managing the wireless connection, choosing an appropriate level of compression and encryption, and performing computation on behalf of the mobile both interactively and in the background. The representative has the same access rights and security privileges as the mobile end user.

The architecture relies on strongly typed transmissions. A dynamically extendable type system enables type-specific compression levels and abstraction mechanisms for progressive object transmission, thus husbanding link bandwidth. For example, depending on link quality, we send a raw, compressed, or lossy/highly compressed bitmap to a mobile client. As an extreme example, we can perform edge detection and OCR on the bitmap of a topographic map, only transmitting the text and edges. This *distillation* process produces a summary version that can be used to evaluate the value of the full bitmap. These techniques exploit the backbone compute servers to trade software latency to compress for increased effective bandwidth, thus reducing overall latency.

Figure 6 illustrates the image proxy at work. A multicolored image of Stempy is reduced to a 4-level gray scale thumb-

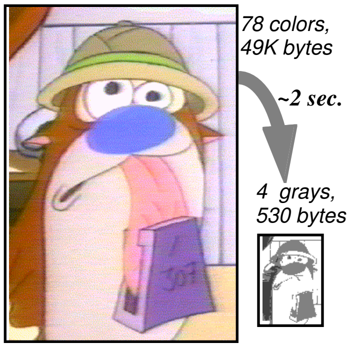

Figure 6 Real-Time Translation By Image Proxy

nail in about two seconds of processing on a SparcStation20 workstation.The original image would require approximately 25 seconds to transmit over a 19.2 kbps modem. The translated image requires only 0.33 second!

Figure 7 illustrates another real-time proxy for video transcoding [1]. It maps between alternative formats in real-time. For example, motion JPEG arrives over a high speed wired network at 1 mbps and is transcoded into an H.261 format suitable for transmission over a wireless local area network at 256 Kbps. For low motion video streams, we have achieved full 30 fps transcoding rates on a SparcStation 20; for higher motion streams, the transcoder rate drops to a still acceptable 21 fps.

These prototypes are first steps in a comprehensive approach for data type specific proxy services. We are developing a range of bandwidth/computation tradeoffs for each data type, distinguishing between independent *versions* and *layers*, which form a sequence of increasing detail useful for progressive transmission of objects.

6. IMPLEMENTATION STATUS, SUMMARY, AND CONCLUSIONS

Figure 8 provides a network stack-oriented view of the BARWAN architecture. It is being designed, deployed, and evaluated with wireless technology and network services from AT&T, DEC, GTE Mobilenet, Hughes Aircraft Corporation, IBM, Metricom, and Pacific Telesis. Berkeley's

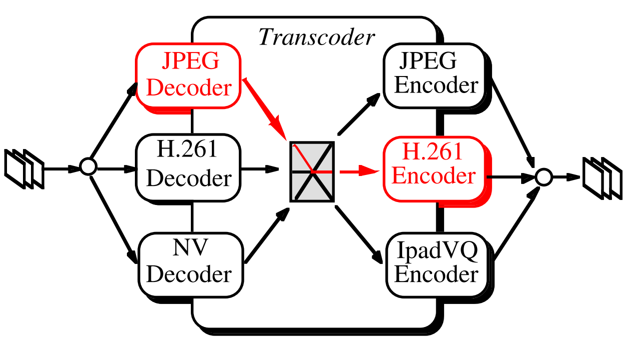

Figure 7 The Video Gateway Transcoder Proxy

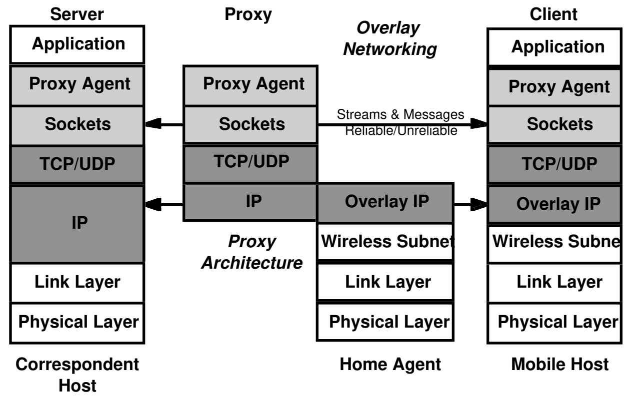

Figure 8 BARWAN Network Protocol Stacks

Soda Hall is being equipped with a building-wide wireless LAN based on AT&T WaveLAN technology. Metricom has deployed their wireless packet relay network in the East Bay as well as on campus, making it possible to roam around the Bay Area while still providing reachable from your home network. We also have research access to CDPD services from GTE Mobilenet and the DirecPC DBS system from Hughes.

We have already demonstrated mechanisms for low latency handoff [5], reliable wireless transport [2], real time video transcoding for wireless transmission [1], and a prototype image proxy for restricted bandwidth web browsing. Rudimentary user-initiated vertical handoff between multiple wireless local area networks and the Metricom packet relay network is now operational. We expect to demonstrate network-initiated handoff over the local and wide-areas by the Summer of 1996.

We have described wireless overlays as a way to combine the advantages of wide-area coverage while achieving the best possible bandwidth and latency for mobile devices. We are developing support mechanisms that enable applications to adapt to the changes in the quality of their network connectivity. In our model, agents moderate these changes in network connectivity on behalf of applications.

7. REFERENCES

[1] E. Amir, S. McCanne, H. Zhang, "An Application Level Video Gateway," ACM Multimedia '95, S.F., CA, (October 1995).

[2] H. Balakrishnan, S. Seshan, E. Amir, R. H. Katz, "Improving TCP/IP Performance over Wireless Networks," ACM Mobile Computing and Networking Conference, Oakland, CA, (November 1995).

[3] J. M. Kahn, *et al.*, "Non-Directed Infrared Links for High-Capacity Wireless LANs," *IEEE Personal Communications*, V 1, N 2, (Second Quarter 1994), pp. 12-25.

[4] C. Perkins, "Providing Continuous Network Access to Mobile Hosts using TCP/IP," *Computer Networks and ISDN Systems*, V 26, (November 1993), pp. 357-369.

[5] S. Seshan, "Low Latency Handoff in Wireless Networks," Ph.D. Dissertation, University of California, Berkeley, CS Division Report, (December 1995).