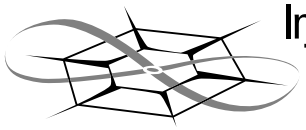
The University of Kansas



Information and Telecommunication Technology Center

Technical Report

NSF Wireless Networking Workshop: Final Report

Thomas F. La Porta, The Pennsylvania State University
Mario Gerla, The University of California, Los Angeles

ITTC-FY2004-TR-32950-01

January 2004

Project Sponsor:
The National Science Foundation
under grant ANI-0344042

Copyright © 2004:
The University of Kansas
2335 Irving Hill Road, Lawrence, KS 66045-7612
All rights reserved.

NSF Wireless Network Workshop
Preface

Wireless networking today is the challenging problem of networked systems.

The issues of tenuous connectivity, limited processing capacity, batteries, and on-the-move operation increases the complexity of achieving robust communications. At the same time, there is great demand for this technology. Personal schedules change on a day-to-day and hour-to-hour basis and people increasingly need to communicate on the move. Businesses recognize this need to utilize ubiquitous communications and they also seek to provide solutions. However, new wireless networking services generally require new radio spectrum allocations, which are limited.

While the demand for mobile, wireless services is increasing, scientists and engineers are developing new techniques to better utilize the radio frequency spectrum through advanced circuit technology, digital processing, and network coordination. These technology advances provide the foundation for agile radios that coordinate their actions and adapt to the surrounding communications environment. These advanced radio technologies combined with advanced networking technologies promise a rich, mobile, robust communications infrastructure.

The National Science Foundation has recognized these advances and developments and has organized a series of workshops addressing the issues of innovative radio technologies, mobile networking, and policy.

This report on the NSF Wireless Networking Workshop represents the concepts, ideas, and issues in wireless networking and elicits the important research steps to be taken. The workshop participants have volunteered their time and effort towards this endeavor, and for that we offer our thanks. Tom La Porta of the Pennsylvania State University and Mario Gerla of the University of California, Los Angeles collected comments and edited the report. James Sterbenz provided substantial comments and advice. Without these contributions, this report would not have been possible.

The future is bright for innovative wireless networking services. There are numerous opportunities to move beyond traditional point-to-point, broadcast, and cellular services to a rich and dynamic radio-based communications infrastructure. This workshop collects the thoughts and visions of how the wireless networking community may reach that future.

Gary J. Minden and Joseph B. Evans
The University of Kansas
Workshop Organizers

Table of Contents

NSF Wireless Network Workshop
Final Report

1 Introduction

This report presents the findings and recommendations of the NSF Wireless Workshop held in Chicago, IL, on July 29-30, 2003. This workshop followed a previous NSF Workshop on The Future of Spectrum: Technologies and Policy and focused on wireless networks. The findings of The Future of Spectrum workshop (see the Appendix A for a summary) pointed out many advances in radio technology that enable and require innovative new wireless networking techniques. Therefore, the Wireless Networking Workshop was organized and held to quantify the research required to maximize the benefit of the new wireless communication technologies.

The combined output of the research on wireless communications and wireless networking will have a major impact on the scientific community, homeland security and defense, and commercial wireless applications. The key conclusions from the Future of Spectrum workshop are (a) new wireless broadband technologies are emerging to provide more flexible and higher bit rates on wireless links; (b) new implementation technologies are maturing to enable the building of more powerful systems; and (c) new agile radios, capable of operating across a wide range of frequencies, waveforms, and bit rates, will be prevalent. Combined with network intelligence and control, these could become cognitive systems, capable of adapting to environmental conditions and application requirements, as well as learning from past experience, creating a much more powerful wireless system. There are several emerging applications that will benefit from these advanced radios and from the new network architectures that they will enable. In the next section, we introduce two representative examples.

1.1 Network scenarios

In this section, we describe two sample applications that illustrate the research challenges and the potential power of wireless networks. These are not meant to be a comprehensive list, but are discussed to provide the reader with a meaningful example of wireless networking applications.

Two emerging wireless network scenarios that will soon become part of our daily routines are vehicle communications in an urban environment, and campus nomadic networking. These environments are ripe for benefiting from the technologies discussed in this report. Today, users in cars connect to the cellular system, mostly for telephony services. The emerging technologies will stimulate an explosion of a new gamut of applications. Within the car, short range wireless communications (e.g., PAN technology) will be used for monitoring and controlling the vehicle's mechanical components as well as for connecting the driver's headset to the cellular phone. Another set of innovative applications stems from communications with other cars on the road. The potential applications include road safety messages, coordinated navigation, network video games for passengers, and other peer-to-peer interactions. These network needs can be

efficiently supported by an "opportunistic" multi-hop wireless network among cars, which spans the urban road grid and which extends to intercity highways. This ad hoc network can alleviate the overload of the fixed wireless infrastructures (3G and hotspot networks). It can also offer an emergency backup in case of massive fixed infrastructure failure (e.g., terrorist attack, act of war, natural or industrial disaster). The coupling of car multi-hop network, on-board PAN, and cellular wireless infrastructure represents a good example of hybrid wireless network aimed at cost savings, performance improvements, and enhanced resilience to failures.

In the above application, the vehicle is a communications hub where the extensive resources of the fixed radio infrastructure and the highly mobile ad hoc radio capabilities meet to provide the necessary services. New networking and radio technologies are needed when operations occur in the "extreme" conditions, namely, extreme mobility (radio and networking), strict delay attributes for safety applications (networking and radio), flexible resource management and reliability (adaptive networks), and extreme throughputs (radios). Extremely flexible radio implementations are needed to realize this goal. Moreover, cross layer adaptation is necessary to explore the tradeoffs between transmission rate, reliability, and error control in these environments and to allow the network to gradually adapt as the channel and the application behaviors are better appraised through measurements.

Another interesting scenario is the Campus, where the term "Campus" here takes the more general meaning of a place where people congregate for various cultural and social (possibly group) activities, thus including Amusement Park, Industrial Campus, Shopping Mall, etc. On a typical Campus today, wireless LAN access points in shops, hallways, street crossings, etc., enable nomadic access to the Internet from various portable devices (e.g., laptops, notebooks, PDAs). However, not all areas of a Campus or Mall are covered by department/shop wireless LANs. Thus, other wireless media (e.g., GPRS, 1xRTT, 3G) may become useful to fill the gaps. There is a clear opportunity for multiple interfaces or agile radios that can automatically connect to the best available service. The Campus will also be an ideal environment where group networking will emerge. For example, on a University Campus students will form small workgroups to exchange files and to share presentations, results, etc. In an Amusement Park groups of young visitors will interconnect to play network games, etc. Their parents will network to exchange photo shots and video clips. To satisfy this type of close range networking applications, Personal Area Networks such as Bluetooth and IEEE 802.15 may be brought into the picture. Finally, "opportunistic" ad hoc networking will become a cost-effective alternative to extend the coverage of access points. Again, as already observed in the vehicular network example, the above "extensions" of the basic infrastructure network model require exactly the technologies recommended in this report, namely: multimode radios, cross layer interaction (to select the best radio interface) and some form of hybrid networking.

These are just simple examples of networked, mobile applications drawn from our everyday lives. These applications, albeit simple, will be immediately enhanced by the new wireless technologies here discussed. There is a wealth of more sophisticated and demanding applications, for example, in the areas of pervasive computing, sensor

networks, battlefield, civilian preparedness, and disaster recovery, that will be enabled and spun off by the new radio and network technologies. More examples will be offered in the following sections.

1.2 Summary Recommendations

During this workshop comments, suggestions and recommendations were made by keynote speakers as well as attendees. This wealth of material has been carefully recorded and is reported in the various sections of the report. In the present section we summarize the key findings and recommendations:

NSF should support building a wireless networking community that includes communications and networking researchers. This support may consist, for example, in organizing periodic workshops and mixed PI/non-PI meetings.

A process should be implemented for continued interaction between regulatory and technical communities so that the full ramifications of policy and technology may be well understood.

It is of high importance for NSF to fund research on wireless networking along the following guidelines:

- Balance fundamental research among theoretical, systems, and experimental projects;
- Addresses wireless networking with both current and new radio technologies;
- Interdisciplinary work combining the physical and networking layers; and
- Diverse technical approaches.

There is a critical need for provisions and procedures for procuring community tools for wireless networking. This includes “open” facilities for prototyping programmable radio systems that may be used by many researchers as toolkits for experimentation, and national test facilities so that wireless networking solutions may be tested in realistic environments. The workshop participants envision that these national facilities will provide key support for academic research and will facilitate interactions between academia and industry. The key idea is to provide one or more centralized institutions that provide “corporate” memory and technical support for realistic wireless experimentation. A national laboratory could follow a model like Fermi or Argonne Laboratories in Physics. The workshop participants felt strongly that this national testbed facility should not be funded out of current NSF program funds, but should be part of a consistent national effort to bring more research funding to bear on the important problems in information technology.

There should be a strong education component to this research to train engineers and researchers.

It was recommended that NSF examine the funding levels of wireless networking research compared with other disciplines to ensure that this community, which is noted

for being very selective in its review processes, receives significant support for its research programs.

The rest of this report describes the technical areas of research that the consensus members of the workshop felt were of high importance and impact over the next five years. In summary, the areas are:

1. Wireless network architecture: In addition to traditional cellular and point-to-point wireless systems, over the past several years peer-to-peer, ad hoc wireless networks have emerged in which wireless devices communicate directly with each other, often using other wireless nodes as intermediate relays or routers. New architectures that fall between this dichotomy of fully centralized and fully distributed systems are also beginning to emerge. This research will address issues related to these network architectures, new network architectures, and heterogeneous networks in which multiple wireless access networks coexist. We discuss issues such as the impact of new air interfaces and agile radios on access network architectures, core networks, access protocols, hybrid networks, cross-layer interactions, and protocol layering.

2. Management of networks of radios: This research addresses challenges in managing and controlling wireless networks. Key themes include supporting auto-configuration and self-organization under policy and security constraints; dealing with the uncertainty present in a wireless environment by both making systems resilient to uncertainty and reducing uncertainty; making survivable systems in the face of wireless links that have characteristics such as asymmetry, weak, intermittent, and episodic connectivity by reliance on eventual connectivity mechanisms and open-loop control; architectures and algorithms that expect and exploit mobility; developing a framework for adaptive and agile networking; and pricing, including trade-offs of cost, price, and resource allocation. Furthermore, research must properly understand resource tradeoffs between processing, memory, data rate, energy consumption, and latency in communication.

3. Wireless Systems: This research addresses challenges in wireless networks from a systems perspective. Key themes are interactions of protocol layers and different access networks including cross-layer optimizations and feedback/control mechanisms, vertical handoffs in which users change between wireless interfaces, and the trade-offs between maintaining solutions that are independent of layer 1 and 2 technology vs. information sharing; application-driven networks including protocol design, network architecture and adaptation; and network evolution including interworking, overlay network architectures, and efficient gateway designs.

4. Pervasive Wireless Networks: This research will focus on environment sensors from the point of view of monitoring systems and smart spaces. The research includes both theoretical and systems components. The theoretical components address issues related to capacity, such as limits considering node density, mobility, and application profiles; optimal levels of hierarchy; and the impact of radio designs on capacity limits. The systems components address architecture, including deployment and distribution of functions; management, such as cross-layer optimizations; and energy-related topics such

protocol designs that account for power-state and the interactions of protocols with electronics.

5. Security for wireless systems: This research addresses the wireless security challenges posed by fluidity, scale, and trade-offs with performance. Wireless networks will support a multitude of users, interconnected by multiple networks, using different applications, many with conflicting security requirements. The complexity and dynamic nature of the networks is increasing, for example with ad hoc peer-to-peer networks, while the demands for stronger security are also increasing. The security solutions must accommodate policy, ease of use, ease of deployment, and allow for high performance communication; these are frequently at odds with strong security. Research includes defense in depth (layers of security defense), adaptive security solutions, and security for transient relationships.

6. Evaluation of wireless systems: This topic addresses the need for realistic and affordable means for carrying out representative, repeatable, and verifiable experiments to validate research on wireless networks. This includes open tools and simulation models, and the ability to use a national test facility to access realistic environments, as well as mapping experimental results to models that can be used in simulation.

These areas are justified and discussed in detail in Sections II–VII. In Section VIII we summarize the report and acknowledge the efforts of those that provided assistance with this workshop. Appendix A provides a summary of the findings of the May 2003 NSF Workshop on the Future of Spectrum. Appendix B includes a list of attendees and contributors to this report.

2 Wireless Architectures

Advances in wireless transmission technology are expected to enable cost effective ways to construct communication networks. We believe it is important for NSF to solicit proposals that explore new paradigms for network architectures that exploit such technology advances. New paradigms should consider features inherent in wireless communications that are generally not present in wireline networks.

A key distinguishing feature of wireless networks, as compared with wireline networks, is the potential ability of the transmission resources to be **reconfigured on a fast time scale**, for example with agile radios. In a wireline network, a decision to invest in a communication link may be made on a time scale of months or years, and once the investment is made, the communication link persists on comparable time scales. In contrast, in wireless networks it is possible to quickly reconfigure transmission resources, so that decisions to allocate a communication link may be made on a time scale of milliseconds or less.

A second important feature of wireless networks is that **signal interference** between different communication links in the same network, as well from external entities, may severely limit the rate at which communication can take place. In the worst case a wireless channel could be rendered useless by either external noise or excessive data

traffic. This introduces a dimension of uncertainty that is important to consider in network architecture.

Third, users and communication nodes may be **mobile**. This introduces the need to cope with a constantly changing network topology and environment. Mobility also introduces the reliance on battery-powered devices that places important constraints on energy consumption. Furthermore, both wireless communication channels and constantly changing environment bring about a new set of open issues in building a scalable and resilient communication infrastructure, as DOS attacks over wireless channels are more difficult to defend and mobility imposes new challenges in network scalability.

It is presently unclear what role the wireless transmission technology will take in shaping the structure of communication networks of the future. In the near term, it is apparent that this role will be in support of access networks (e.g. to the Internet), as well as in support of fixed point-to-point transmission links (e.g. in a wide area network). It is suggested that NSF encourage networking research that considers the use of wireless transmission technology in these contexts, as well as in other more futuristic contexts that exploit wireless transmission to a greater degree. For example, it may be become feasible to deploy wide area networks where almost all communications is wireless, exploiting directional antenna and free space optical communication. . It is important to understand whether such a wireless-based infrastructure would require a fundamentally new architecture that is different from that of the today's Internet, and if so how this new architecture should look.

In general, it is important to explore a diversity of possible uses for wireless transmission technology. For example, this includes cellular and Wi-Fi access networks, ad-hoc networks that use multiple wireless hops between source and destination, hybrid access networks that combine a cellular structure with multi-hop communication, sensor networks, wireless networks to support communication between large numbers of motor vehicles, and wireless wide area networks. We now discuss some architectural issues that are likely to be important in some of these scenarios.

2.1 Access Networks

The majority of wireless networks today are used as access networks, i.e., wireless is used as a link to access resources on the Internet or other wired networks. These access networks can be classified further into two types, namely, wide-area wireless networks (e.g. cellular) and local-area wireless networks (e.g. Wi-Fi). In this section, we describe the current architecture of these networks and identify possible evolutionary trends in these architectures as a result of emerging research ideas and/or technologies. The question of the appropriate architecture for the wireless access networks that trade-off performance versus flexibility remains an open research problem.

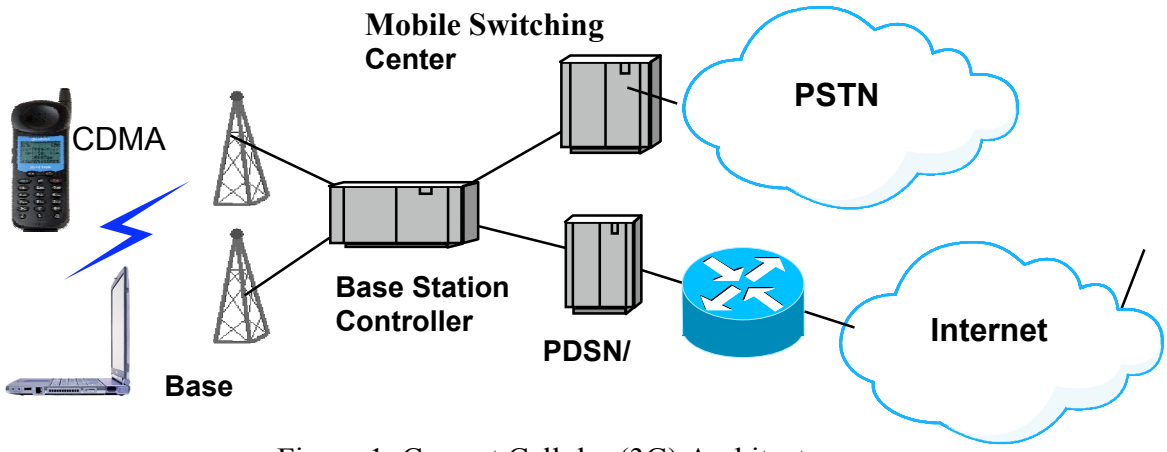


Figure 1: Current Cellular (3G) Architecture

The third generation (3G) cellular architecture is shown in Figure 1. Mobile devices connect to base stations using the CDMA-based air interface. Base stations are connected to a base station controller (BSC). The BSC performs several network-based functions such as soft-handoff and reverse outer-loop power control. At the BSC, the voice and data traffic take separate paths – voice traffic is routed to the Mobile Switching Center and into the Public Switched Telephone Network (PSTN) and the data traffic is routed through a packet data service node and a home agent into the Internet. The hierarchical nature of the cellular architecture results from the high level of functionality in the network as well as the need to scale to millions of wireless users. The separation of voice and data traffic is a result of the need to evolve from an existing legacy voice network.

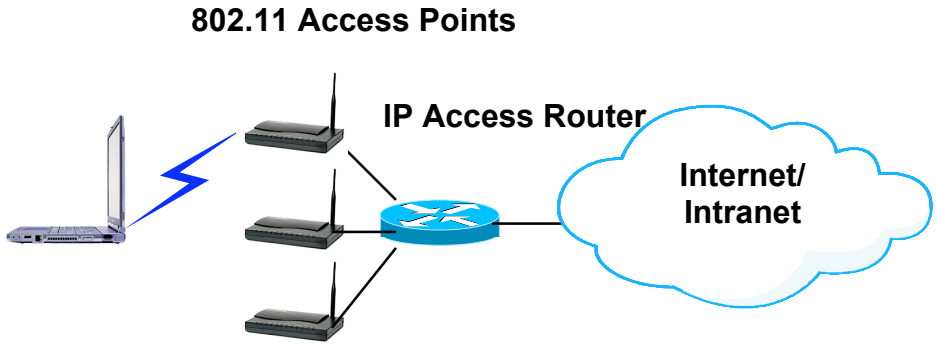


Figure 2: Current 802.11 (Wi-Fi) Architecture

The architecture of current 802.11-based Wi-Fi networks is shown in Figure 2. Obviously, the architecture is much simpler as compared to the cellular networks – the wireless specific functionality is relegated to the access points that simply connect to the switches and routers of the Internet.

The open question is how the architecture of access networks evolves given the following trends:

- Voice and data services are converging
- Wireless and wireline networks are converging

- Different access networks (802.11/UMTS/CDMA2000) are also converging to provide integrated services to devices that are converging (e.g. PDA and Phone)

For example:

- Is it possible to implement the functionality of cellular networks in a distributed manner at the base stations, thereby preserving the flexibility of an architecture that is typical of current Wi-Fi networks while maintaining the rich functionality? In other words, can we relegate all wireless-specific functionality into the base stations or is this dependent on the specific air-interface technology employed?
- On the other hand, will Wi-Fi networks evolve the way cellular networks have evolved, i.e., through the addition of centralized network controllers in the access network that limits network flexibility but enables richer set of services?
- If there is a given trade-off between performance and architecture flexibility, when is it better to choose the former and when the latter?

These are just a few of the open research issues that will determine how architectures of wireless access networks will evolve in the future.

2.2 Hybrid Networks

While highly centralized access networks and completely decentralized ad hoc networks form two ends of a spectrum of wireless architectures, several interesting architectural possibilities lie in between these two extremes. These intermediate architectures are sometimes referred to as hybrid networks as they have elements of both centralized and decentralized architectures.

Hybrid networks can be divided into two types: ad hoc networks that enhance infrastructure networks and infrastructure networks that enhance ad hoc networks. In the former case, examples include using ad hoc relays to migrate traffic from a highly-loaded cell to a neighboring lightly loaded cell, or migrating traffic from the edge of a cell (with poor signal quality) towards the center of a cell (with better signal quality). In the latter case, infrastructure can be used to reduce the complexity of ad hoc routing protocols, increase reliability of message delivery between the ad hoc nodes, and reduce the number of hops in routing packets. Furthermore, hybrid networks can also be differentiated by the wireless technologies used – the wireless interface corresponding to the ad hoc and infrastructure modes can be of the same type (e.g. GSM) or of different types (e.g. unlicensed 802.11 and licensed CDMA), resulting in different performance and protocol trade-offs.

Research in hybrid network architectures is just emerging and early research results are promising. However, a number of open issues need to be addressed before the full implications of hybrid networks can be realized, including

- Are there other, different, and useful ways of combining ad hoc and infrastructure networks beyond the ones described above?

- How to augment ad hoc routing protocols to take advantage of the presence of an infrastructure?
- What is the impact of hybrid networks on improving availability/reliability (researchers have so far focused on improving performance)?
- How best to combine the physical (coding, power, directional/MIMO antenna), media access, and networking layers to arrive at a optimal cross-layer design that trades-off performance versus flexibility (see the following section for more details)?
- What are the security and survivability characteristics of hybrid networks?

2.3 Protocol Layers

The protocol stack provides a useful abstraction for a network architecture. The traditional seven-layer network protocol stack has the physical layer isolated from the higher layers. One of the open question in protocol layering in wireless networks is whether this isolation of the physical layer is appropriate, or do we need new abstractions to capture the inherent characteristics of the wireless media and expose it to the higher layers. This can be achieved through the design of new layers or through appropriate mechanisms for enabling cross-layer optimization and interaction (knobs and dials).

The increased awareness that inter-layer interactions can help achieve significant gains in performance of wireless networks has led to many proposals for improving performance of various layers of the protocol stack in a wireless environment. This interaction consists of a control loop in which lower layers convey their characteristics to allow the adaptation of upper layers (dials), and upper layers influence the behavior of lower layers (knobs). In some instances, the inter-layer interaction has been too tight in that the protocol implementations are not portable across differing systems. Specifically, there have been several protocols designed for antennas with improved capabilities, such as beam steering. From a systems perspective, it is undesirable to have to deploy a different protocol for each type of antenna implementation. At the same time, it is necessary to adapt protocol behavior to the antenna behavior. This makes it important to develop suitable abstractions for capturing the antenna behavior, and making the abstract representation available to higher layer protocols via suitable interfaces. In effect, the antenna becomes a separate layer of the protocol stack with appropriate interfaces to communicate with other layers of the stack.

Open questions include:

- What information should be conveyed across layers (in some cases balancing performance against policy and security)?
- What is the appropriate abstraction for this information, balancing flexibility against tractability?
- What degree of influence and control should be exerted to lower layers, and how to deal with conflicting goals from different upper layer applications?
- Should a new layer, such as an antenna layer, be considered?

3 Management, Monitoring, and Control

The management, monitoring, and control of networks in general is difficult to achieve, but essential to their use and operation. In the case of wireless networks, in particular, management, monitoring, and control is extremely challenging due to the dynamic characteristics of the channels and resulting topology (which is further exacerbated by mobile nodes). In this section we will describe some of the critical research issues that need attention in this area: auto-configuration/self-organization, uncertainty, survivability, and pricing.

3.1 Auto-Configuration and Self-Organization

In wired networks, there is generally a network engineering effort to determine the placement and interconnection of network elements and links that results in the deployment of the network to meet the needs of the users. A primary motivation for wireless networks is to avoid the inflexibility of fixed infrastructure, and thus such manual network configuration efforts are very undesirable. It should also be noted that in the case of ad hoc networks, there doesn't even exist a central authority to which configuration might be assigned. We therefore need mechanisms for the auto-configuration of nodes and their self-organization into a usable network, driven by appropriate policy and security concerns.

We can list a set of mechanisms and dependencies from individual node configuration through network formation and operation:

- Auto-configuration of nodes based on their environment and usage needs, including identifier or address assignment.
- Neighbor discovery to establish the set of directly reachable nodes, typically using beacons.
- Link formation by the exchange of information (such as ID, type, capabilities) to determine which links to establish and keep alive.
- Self-organization and federation into a network by forming hierarchical clustered federations. Federations provide the basic network layer infrastructure over which routing and QoS can occur.
- Topology optimization and maintenance based on a number of criteria (including policy) and adjusted to the dynamic behavior (join/leave of nodes and merge/split of federations).
- Resource discovery, such as content servers, caches, and media stream transformation (transcoding and mixing). Resource discovery algorithms need to locate a set of usable resources and select the best one, based on application functionality, application and network QoS optimizations, and policy considerations.
- Internet dependencies – the Internet provides a wealth of useful resources and services, and may provide some of the links in which a distributed federation forms. However, wireless nodes should be able to operate autonomously even when links to the Internet fail, or are unavailable due to remoteness or security (e.g. Faraday cage in

a secure building). A balancing act is needed between Internet dependency and the ability to perform autonomously.

While some preliminary research has been done in these areas, significant work remains, particularly for policy-constrained, secure, and survivable auto-configuration and self-organization. For example, a number of techniques are well known to organize a network into a clustered structure, but significant research is needed on clustering based on policy so that only desired nodes of a coalition are part of a network federation, as well as continuous optimization and re-organization of the federation to account for dynamic link characteristics and group membership.

3.2 Uncertainty

A primary difference between wired and wireless systems is uncertainty. This manifests itself at the physical layer as rapidly varying link qualities or availabilities, owing to both channel conditions and variable points of network attachment for mobile users. This variability wreaks havoc with standard Internet protocols owing to implicit underlying structural assumptions about link reliability and due to the strict layering abstractions.

For example, a delayed response from an endpoint in a wired network typically suggests congestion along some part of a route. TCP/IP requires sources to immediately and strongly limit transmission when congestion is detected (backoff). However, in wireless systems, the lack of an acknowledgment is also likely to occur when corrupted packets are dropped; in this case the desired response is immediate re-transmission rather than backoff.

Furthermore, what begins as uncertainty associated with the physical layer of the system percolates up through every layer of the protocol stack since each layer depends on the mechanisms used by the layer below. For example, there exists uncertainty in user mobility, in the wireless network topology and load, in traffic characteristics and in the resource availability, to name a few. Thus, coping with and/or reducing uncertainty is a key challenge and perhaps an organizing principle for understanding the role of wireless and its seamless and successful integration into the Internet.

3.2.1 Coping with Uncertainty

One approach is to simply react to conditions as warranted. To this end, research is needed to articulate appropriate responses to various wireless/mobility-induced network events under differing scenarios as well as to understand how relevant information can be gathered and disseminated to appropriate places. For instance, if mobility of user network points of attachments is given, then timely knowledge of user network itineraries would certainly help in efficient resource allocation. It is an open research question whether it is possible or advisable to construct and maintain such itineraries, settling upon the proper levels of abstraction, and the necessary degree of information dissemination.

Likewise, for uncertainty owing to the wireless channel, developing models which capture and relay the relevant features from a network perspective, such as the fading

channel or the interference environment, requires similar consideration of appropriate information collection, abstraction, and dissemination at the physical layer and above.

3.2.2 Reducing Uncertainty

As opposed to simply coping with uncertainty, one might also ask which aspects of wireless system design are most deleterious to efficient network operation and seek to reduce them in some way. For instance, suppose the most challenging aspect of wireless is physical channel variability, and the strategies necessary to overcome it using network protocols are provably prohibitive. If true, such a result suggests that effort would be best expended in controlling variability in the wireless channel using, for instance, recent advances in multiple antenna systems to better guarantee channel quality, or even more recent advances in exploiting mobility and channel variability for improved throughput.

Likewise, one could imagine that with proliferation of various wireless devices, mutual interference is a serious problem. One could then imagine developing network services that help wireless transmitters and receivers better coordinate their use of shared wireless resources, thereby controlling the interference environment to the extent possible. These considerations suggest that an open area of research is the determination of methods for modeling and coping with uncertainty

3.3 Survivability

Survivability is the capability of a system to fulfill its mission in a timely manner, even in the presence of attacks or failures. While there has been considerable research on fault-tolerance, which assume random failures, the ability to survive coordinated attacks on the network and its infrastructure is a significantly harder problem. In the case of wireless networks, the problem of survivability is even more difficult, since an open channel allows an adversary (enemy, competitor, or cracker) to eavesdrop and attack the network without needing physical access to a network node.

We can divide the problem of survivability into survivable access to information, and survivable end-to-end communication (sessions between users can be created and remain active when needed). To achieve this we consider four broad areas of research that would benefit wireless networks: survivable connectivity, survivable communication, agile/adaptive networks, and airborne nodes.

3.3.1 Survivable Connectivity

The first major goal in survivability is to establish and maintain a connected network, whenever practical. This allows conventional routing and end-to-end protocols to function with reasonable performance. There is a fundamental tradeoff between connectivity and transmission power. Increased transmission power enhances connectivity, but does so at the expense of increased energy demands (particularly important for self-powered mobile nodes), reduced stealth (adversaries are more likely to detect, intercept, and exploit communication), and congestion (competition for limited spectrum).

While some work has recently been done on dynamic transmission power control and the use of direction antenna, a number of issues are still poorly understood and in need of further research, including:

- What level of connectivity (e.g., biconnectivity, triconnectivity) yields the best tradeoff between robust connectivity and stealth?
- Adaptive adjustment of the transmit powers and topology to evade jammers and interceptors.
- Combining network-layer approaches with physical layer approaches (including cross-layer interactions)
- Energy management that allows optimum power consumption for the node and/or network

3.3.2 Survivable Communication

There will be times, particularly in extremely challenging environments, where it will not be possible to maintain network connectivity. In these cases, we are interested in survivable communication (information access and end-to-end sessions) even though there is no stable network connectivity. Since we must assume that the conditions in a wireless channel are time-varying (particularly as nodes move), the result is a channel that may be asymmetric, may be weakly connected, and may suffer episodic connectivity during which there are periods of disconnection. It is crucial for network survivability that the protocols and algorithms expect these conditions as part of their normal operation, and communicate in spite of them.

Asymmetric Channel Connectivity: Conventional network and transport protocols have traditionally assumed bidirectional connectivity for proper operation. At the network layer, this means that routing protocols do not have to account for unidirectional disjoint paths; at the transport layer this means that a reliable back channel is assumed. There are a number of wireless scenarios where this is not the case. For example, one of a pair of communicating nodes may have less transmission power, or unidirectional communication may be desired in a case were the receiver wishes to remain radio silent (which includes no transmission for ACKs).

Research is needed in network routing and signaling protocols and in end-to-end protocols for asymmetric and unidirectional paths. In the case of intermediate links along a path, it is essential that the routing protocol support disjoint forward and reverse paths. In the case of an asymmetric end user, the routing protocol must support disjoint unidirectional paths and network layer signaling must not require a back channel.

Similarly, in such situations it is necessary to maintain end-to-end sessions even when the link shuts down in one direction. Closed loop control mechanisms (such as TCP error, flow, and congestion control) generally assume a reliable return channel for acknowledgements to properly function. While some work has been done on enhancing transport protocols for asymmetric channels, additional investigation should consider how to apply open loop mechanisms when necessary for highly asymmetric and

unidirectional paths, and how best to exploit hybrid open- and closed-loop control mechanisms.

Weak and episodic connectivity. Routing protocols currently require that a route (complete path) exist from source to destination before communication is even attempted. This *eventual stability* model of ad hoc routing assumes that routing converges eventually after partitioning. Under this model, a complete path to destination must exist at a given time; otherwise, communication is not attempted at all. Note that this is true whether or not datagrams are to be forwarded along the path, or a connection is to be established.

There are a number communication scenarios where wireless channel conditions are extremely challenging (e.g. noise and fades) and mobility patterns are such that routing algorithms rarely or never converge to stable paths. In these cases, research is needed toward using the *eventual connectivity* model from distributed computing, which relaxes the traditional assumptions so that communication can proceed along partial segments of paths between communicating nodes. Research is needed to understand how to apply these techniques to survivable networking in which extreme conditions are *expected* rather than treated as faults.

Mobility. Wireless networks frequently are deployed to support mobile access. Traditionally, mobility in networks has been handled as a necessary evil, with routing protocols adapting as best as possible to mobile and nomadic nodes. Just as survivable networks should expect challenging channel conditions as a normal mode of operation, they should be designed to *expect* and *exploit* mobility.

High mobility often poses challenges to conventional ad hoc routing protocols especially after they reach their reactive limit. In this case it is necessary to use knowledge of the location and trajectories of nodes to predict future location without requiring rapid convergence of routing algorithms. Some first steps to consider trajectories in routing algorithms have been taken, but significant research remains to be done.

Group mobility can be exploited to aggregate nodes in teams that travel in the same directions and to develop a two-level hierarchy that lends itself to scalable routing solutions where conventional routing would simply fail because of excessive overhead. In on demand routing protocols where an expensive flood search is generally required to locate a destination and trace a path to it, mobility can help reduce the overhead by exploiting the concept of *last encounter routing*. Namely, a source can find the destination by stepping through the roaming nodes that have recently seen the destination and have drifted towards the source.

We can further consider how to *exploit* mobility to communicate when otherwise impossible. In the worst case, eventual connectivity routing will store data until a promising outgoing link becomes available (optimistic transfer). Proactive control can be used in two ways to expedite the transfer of data. *Movement control* can be used to exert control on other nodes to move them into range such that a path toward the destination exists. Alternatively, mobile nodes can *store-and-haul* packets toward their destination by

physically transporting the data, without suffering from interference and maintaining complete stealth.

3.3.3 Adaptive and Agile Networking

Even if application and mission scenarios were uniform and known in advance, mobile wireless networks are inherently dynamic, providing us with the problems of uncertainty described previously. Thus, survivable networks need network nodes and protocols that are aware of, and adapt to their environment. We therefore believe that a framework is needed for adaptively and agility in wireless networking.

Research prototype radios offer agility in terms of frequency bands of operation, modulation techniques, choice of MAC protocols, and power levels. These can be to enhance performance to the end user, to better use and manage shared spectrum, to increase stealth when needed, and to augment network layer survivability. Software radios are an important enabling technology for link and MAC adaptation. Furthermore, active networking technology provides a basis for dynamic deployment of protocol mechanisms and adaptation to traffic in the context of the wired Internet, and has been the subject of considerable research. The application of this technology to mobile wireless networking allows the dynamic selection of, not only MAC and network layer parameters previously discussed, but also the ability to dynamically provision and negotiate algorithms and select entire protocols based on application requirements and the communication environment. For example, sets of communicating nodes may wish to change from a simple efficient MAC protocol and routing algorithm to more sophisticated and survivable, as the environment becomes more challenging.

We thus believe that a framework for adaptive and agile networking must be developed, to not only understand how we can best deal with the uncertainty in wireless networks, but to allow algorithms and protocols to be dynamically deployed as appropriate to particular sub-networks, without requiring that the minimum-standard-denominator standardized solutions are imposed in all cases.

3.3.4 Airborne and Satellite Nodes

Airborne nodes (piloted or UAVs – unpiloted aerial vehicles) and satellites can serve an important role in mitigating the effects of weakly connected channels and node mobility.

The high altitude of airborne and orbiting nodes enables them to have a large terrestrial footprint, which can enhance connectivity, mitigate the effects of mobility, and supports radio silence through techniques such as datacycle.

In summary, research in efficient, survivable connectivity includes the following areas:

- Survivable information access and communication sessions
- Opportunistic routing and forwarding using eventual connectivity in weakly and episodically connected environments
- Routing and forwarding in asymmetric environments

- Open loop control mechanisms
- Expecting and exploiting movement
- Developing a framework for adaptive agile networking
- Exploiting airborne and satellite nodes to mitigate mobility and improve connectivity

3.4 Pricing and billing

The evolution of wireless networks depends not only upon the available technology, but also upon the demand for wireless applications and the ability of the network to offer these at acceptable prices. The pricing of applications in wireless networks is likely to be based on users' willingness to pay, on marketing decisions, and on the cost to the network to provide these applications. This latter portion depends on the resources that this application consumes. Pricing is therefore intimately connected with resource allocation.

The sections above have discussed future evolution of wireless network architectures and challenges in uncertainty management. We now turn to pricing and resource allocation in wireless networks, and their relationship to these issues. In addition to the typical roles of resource allocation in wired environments, resource allocation has significant characteristics that are unique to wireless environments.

First, the resources themselves are often different. In addition to processing, memory and data rate (bandwidth), in wireless networks power and energy are often scarce resources. The uncertainty surrounding these resources, and the variability in their availability, pose fundamental challenges in their allocation.

- How should resources be allocated between competing users?
- How should these allocations vary over time in respond to variations in demand and supply? What is the cost of reducing uncertainty?
- What is the value in reducing uncertainty?

Second, resource contention will exist between different types of applications. While cellular networks are architected primarily for voice applications and 802.11 networks are architected primarily for data applications, the future surely will include wireless networks that offer a wide range of applications.

- What type of resource allocation mechanisms will allow for this evolution of wireless network architectures?
- How do you inspire a flexible resource allocation?
- What is the meaning of QoS in a wireless environment in which there is no minimum performance level that can be guaranteed?
- What is the cost of QoS?

Third, in the wireless environment nearby network elements may have a stronger interrelationship due to mutual interference. Such interference causes externalities to the actions of each network element that may motivate increased coordination.

- How can this coordination be accomplished with minimal overhead?
- Can pricing be used as a method for communicating resource demands and allocations between users and the network?
- Can pricing accomplish the necessary coordination between network elements within a mutual interference region with minimal overhead?

4 Wireless Systems

In the following section we discuss open systems research issues related to applications and network evolution.

4.1 Applications

Research has focused almost exclusively on three wireless applications and architectures: high-speed Internet access, ad-hoc networks for emergency and military communications, and sensor networks. The ongoing research, while important, focuses on a small set of core issues such as routing and topology formation, but does not address many other services that are required for these networks to be useful. For example, areas such as fixed or single-hop ad-hoc networks or non-data applications, GPS for location or specialized broadcast networks, and the need for precise time information (e.g., WWV) have received little research attention.

As an example, small-scale ad-hoc networks can be created within a conference room to transfer files or to temporarily use low-speed wide-area wireless access services. While this requires no fundamental breakthroughs in routing protocols, issues of discovery, auto-configuration and security have not been addressed, so that file transfer within a group still usually reverts to the inflexible or very short-range infrared mechanism or to handing around USB memory sticks.

4.1.1 Interaction of Wireless Physical Layer and Applications

A fundamental architectural question in designing wireless systems is what aspects of the physical layer should be exposed to the application.

The information flows both up and down the stack. For example, the application may indicate that certain packets should be delivered to the other side even with bit errors (or that it wants to receive packets with errors) or that it favors throughput over bit accuracy. Conversely, the physical layer should be able to indicate that a PDU contains bit errors or convey the signal strength, noise and SNR encountered.

While much attention has focused on exporting information from a wireless NIC to upper layers and allowing upper layers to modify the behavior of those NICs, this is insufficient, as network entities that are on a wired network, but communicate with

wireless terminals may also benefit by being aware of the characteristics of the far-end wireless link. In addition, with the deployment of hybrid wireless architectures that combine technologies such as 802.16 and 802.11 or networks that are in turn nodes in ad-hoc networks, properties of interest span more than a single wireless link or wireless link type.

Unlike wired links, many modern wireless networks offer multiple services, with trade-offs involving power, bit error or packet error rate and speed. Unfortunately, there is currently no good way to perform service discovery along a concatenation of these links. This service diversity may also complicate multi-parameter routing, since link metrics are no longer simply delay and (available) bandwidth.

There are at least two reasons for providing more detailed information about lower layers to the application. First, one expects that this additional information allows applications to make better decisions and to request appropriate services from the lower layers, as a part of cross-layer optimization. A second aspect, however, may well be equally important, namely allowing applications to provide the user with indications on the source of faulty application behavior. For example, an application may notice excessive packet loss or extended drop outs and may be able to inform the user that the likely cause is the wireless link (and suggest physical relocation) rather than, say, general network congestion.

Both applications require somewhat different information from the lower layers. For diagnostics, low-level measurements such as SNR, raw bit error rate prior to FEC and similar metrics, may be more helpful, while a set available link services and their behavior are more useful to applications.

Rather than wholesale layer violation that removes the useful abstraction offered by layering, the challenge is to minimize complexity and dependence on specific wireless technologies. Tying applications too closely to one technology has increased system complexity in the past (e.g. BlueTooth). Thus, the goal is to export information that allows applications to address situations that cannot be easily handled by the lower layer, rather than tuning for minor performance optimizations. For example, signal strength or bit error information from the lower layers may allow the application to predict imminent disconnection, and thus have the application cache data, change to a different network modality or change the user interface.

A minimalist approach to exporting information from lower layers to upper layers also makes systems more readily testable as the state space is smaller.

Summarizing, open questions include:

- What information should be passed between layers?
- How far up the protocol stack should low-level, for example physical layer, information be passed?
- What is the correct level of abstraction for this information?
- Can sharing be done across nodes to give a global view of network status?

4.1.2 Wireless-aware application design

Much of the research in wireless networking has been done independently of applications. While such "pure" and "fundamental" work is of significant value, trying to solve all the hard problems at the same time is often very difficult (or even impossible), and the disconnect from applications (i.e., protocols designed in a vacuum) may lead to solutions that are unsatisfactory for any real application. A complementary approach to wireless networking research is application-driven and systems-based work. Being application-driven helps limit the scope of the problem and focus research efforts; this can both help speed up the pace of progress and improve the chances of real-world impact. A systems-based approach exposes real-world problems and offers the opportunity to run through the cycle of design, implementation, and deployment, which can feed back into an improved design.

A good example to illustrate this point is ad hoc networking. There has been a large volume of work in this field over the past decade. However, it is only recently that much attention has been paid to applications that might make use of ad hoc networking. Researchers are now discovering that the previous work has numerous limitations when attempts are made to apply it to real problems. The reasons range from the choice of metrics (e.g., mobility being the focus rather than throughput, capacity, or energy consumption) to unrealistic assumptions about the nature of wireless links (e.g., symmetry).

The relatively recent application focus in this space is leading to crisper problem definitions and grounding in reality that significantly increase the likelihood of ad hoc networks being actually deployed and used. A couple of examples:

- Community wireless networks: The idea here is to use multi-hop wireless networks to provide last-mile connectivity to the home. Capacity (throughput per home), security, and management (e.g., troubleshooting) are important. Energy efficiency and tolerance to a high degree of mobility are probably less important considerations.
- Sensor networks: Energy efficiency and tolerance to the failure of large numbers of nodes are the most important requirements. Throughput and mobility are probably less important.
- Home networks: Wireless is an attractive option for networking A/V and other devices within the home. An important consideration is guaranteed QoS (in terms of bandwidth, delay, and jitter).

This list is by no means exhaustive. There are other interesting applications such as actuator networks (which are somewhat similar to sensor networks except that security is likely to be paramount) and location sensing (where the wireless network is used to detect proximity rather than for data transfer).

We believe that such an application-driven approach to defining the problem space will lead to significant new advances in and practical impact of wireless networking.

Some effort has been expended into tuning application behavior for narrow bandwidth networks, e.g., by reducing the size of web pages or shared-applications screens by filtering in middle boxes. Less attention seems to have been paid to develop adaptive input modalities for mobile use. Another possible variable that applications may adjust to is location; for example, applications might change their input mode if they discover that the device is moving at vehicle speeds or they may automatically turn off the radio interface in secured areas or hospitals.

Expected fertile research areas include:

- Application driven networks
- Applications that adapt to their environment

4.2 Network Evolution

Relatively little attention has been paid as to whether it would be feasible and efficient to replace special-purpose wireless applications such as AM or FM radio with systems that are integrated into general-purpose cellular networks, using multicast/broadcast.

4.2.1 Designing for Evolution

Network infrastructure and technologies tend to remain in widespread use long after new technologies have been introduced. This is particularly true for wireless networks, where investments in radios, antennas, radio access networks and towers require wholesale upgrades when changing air interfaces or lower-layer protocols. For example, AMPS, the first-generation analog cellular standard, is still the primary cellular technology for pagers and in rural areas, and , offers a low-cost, ubiquitous low-bandwidth data delivery mechanism. It appears unlikely that these services will disappear in the next decade. Similarly, analog TV and AM/FM radio occupies large swaths of spectrum suitable for long- and medium-range communication. Even though more functional and spectrum-efficient means of broadcasting audio and video signals are available, there is little incentive to strand hundreds of millions of cheap receivers, as evidenced by the slower-than-predicted displacement of analog television by HDTV. Soon, 2G and 802.11b will join the set of legacy technologies.

Research has generally ignored the existence of legacy technologies, assuming a rapid green-field deployment of new technology. There are now a number of historical examples in networking and in wireless technology that indicate that we will continue to operate in an environment characterized by the coexistence of multiple generations of technology, with technology displacement cycles measured in decades. This offers both challenges and opportunities for research. The challenge is one of interworking different technologies most effectively. Such interworking can occur by using the old technology as a voice-band data channel for legacy analog voice and video technologies, i.e., a physical layer, or as a simple 7-bit one-way messaging service, as for paging and SMS. Other transition mechanisms may be more appropriate for wireless networks. For example, voice service at the application layer may be significantly more efficient in

terms of wireless channel usage, but supporting mobility and authentication services becomes far more challenging.

There appears to be little formal guidance and systematic system experience on architecting, designing and evaluating gateway architectures that are scalable, secure, preserve user services and addressing across technology boundaries and do not impede progress for the newer technology. In an era of limited investment capital, technology transitions have to make economic sense, but research into the capital and operational costs and trade-offs of interworking and transitions has been sparse. For example, we may gain a better understanding of the reasons why technology propagation is slow if research analyzes the operational incentives for certain architectures, such as the alignment of revenues and investment for individual players.

Recently, overlay networks have been proposed as means to accelerate the deployment of new network functionality into existing networks, such as better resilience to failures or multicast. However, overlay technology may add additional packet header and processing overhead that wireless links can ill afford. Therefore, the investigation of lightweight overlay technologies appears promising.

A fundamental shared assumption in the networking research community appears to be that of a natural evolution of wireless networks from special-purpose (single-service), circuit-switched, single-provider to general-purpose, service-agnostic, packet-switched and multi-provider systems. This generality comes at a cost in terms of bandwidth, power and processing. Sensor networks have started to seriously consider more restricted network technologies, but there may be value in considering the generality vs. complexity trade-offs. For example, there are successful commercial systems that provide limited information access and messaging, at significantly lower complexity and security exposure, as will as easier billing, than a general packet delivery mechanism. Successful wireless technologies seem to allow multi-modal use, i.e., acting as a content-neutral link-layer interface and offering a useful limited-complexity service such as messaging.

Since every technology will eventually become a legacy technology, it would be helpful to find out what it takes to make technologies fade away gracefully, without impeding technological progress. In essence, technologies must be designed to be forward-compatible. (This is similar to recent efforts to worry about recycling obsolete electronic equipment during the design phase and to require vendors to accept responsibility for discarded equipment.) This may involve aspects of spectrum management, the ability to extract core high-investment-cost components such as the radio interface from the remainder of the system architecture and exposure of a low-level service interface.

It appears that new technologies often introduce both short-lived and longer-lived components. For example, user identification, operational support systems and services seem to survive the original link layer technology, as exemplified by the use of E.164 numbering in the telephone system that is carried over into newer VoIP system.

Open questions include:

- Optimal design of gateways for interworking networks

- Lightweight overlay network design
- Design forward-compatible systems
- Tradeoffs between generalized network design and complexity

5 Pervasive Wireless Systems

A rich diversity of pervasive wireless systems are emerging as the wireless technology is being applied to applications that involve "networking the physical world" as opposed to traditional applications such as voice telephony and Web access. Many prototype systems have been developed, both by industry as well as by the academic community. Very few have passed the judgment of time and have made real impact in the way we do things and use the new computer data communication technology. The recent interest in sensor networks and their applications has however rejuvenated the field of pervasive wireless systems and their applications. Now is an opportune time to learn from the past and invest in defining research directions to make sure that this field has a real impact. Many issues need to be considered as pervasive wireless systems necessitate the harmonious integration and inter-working of many technologies and protocols.

While ranging from pervasive computing systems for smart workspaces to sensor networks from environmental monitoring, these pervasive wireless systems share a common set of attributes. Loosely speaking, these systems are omnipresent (to the extent possible), unconsciously used, and aware of the contextual information. Their tight coupling with the physical world leads to traffic that is low duty cycle event oriented and characterized by spatiotemporal correlations inherent in the physical phenomena that can be exploited by localized in-network computation. The scale and redundancy inherent in these systems leads to a focus on the aggregate behavior of the system as opposed to those of individual network nodes. Another attribute of these systems is the focus on energy as a key system level performance metric because of the requirements of unattended operation and high cost of deploying or modifying the system.

These shared attributes form the foundation of a research agenda to establish a foundation for pervasive wireless systems, and to help model, analyze, and generalize the solutions for them. This will help move beyond the rather applications-specific nature of these systems and help define a real notion of a generic pervasive wireless system.

5.1 Fundamentals and Analytic Models

Pervasive wireless systems such as sensor networks are distinguished from ad hoc networks in that traffic is generated through observation of the physical world, may be processed locally, and then conveyed to some end user via communication relays. Sensing, signal processing, and communication may take place in one node or these functions may be spread among specialized nodes or dealt with in a hierarchical fashion. Further, the networks may be tightly coupled to some communications or energy-distribution infrastructure after some relatively small number of wireless hops. The goal of such networks is to convey information about some physical event (or sequence of events) to some end user set, at some level of resolution. That is, spatial, temporal, and

quantization distortion are inevitable. When some portions of this network are remote from the infrastructure, a typical goal is to convey this information to the end user using the least energy so that the sensor nodes can have prolonged lifetime. Thus the traffic generated will be of particular types governed by the nature of the physical phenomenon to be observed, the energy and communication resources of the nodes, and the requirements of the particular application. In contrast, in ad hoc networks every node may be regarded as a source of a wide variety of traffic.

The important features of a sensor network are thus:

- Some nodes observe physical phenomena at the request of users
- The data are processed according to the distortion constraints imposed by the users
- Some nodes relay this information to users.

Notice that implicitly networking, signal processing and database functionality are bound up together. This opens the questions as to

- the appropriate set of abstractions to provide efficient solutions while having reasonable software complexity;
- whether one can define fundamental performance limits to the set of trades in such a large space.

To approach the latter question, we begin with capacity issues in ad hoc networks, and then discuss how the situation is different in sensor networks, viz., a combination of capacity and rate distortion problems.

5.1.1 Capacity for Ad Hoc Networks

For point-to-point links over the Gaussian channel, the Shannon capacity is

$$C = W \log_2(1 + S/WN) \text{ bits/s}$$

where W is the bandwidth, S is the signal power, and N is the one-sided noise power spectral density. The capacity represents the highest bit rate that can be achieved over the channel without transmission error, assuming perfect synchronization and the use of the most powerful error control coding. We can in fact now achieve rates very close to the Shannon capacity for error rates of practical use. Notice that capacity expands roughly linearly with bandwidth, and logarithmically with signal to noise ratio. Gaussian capacity is important because many channels can, through various signal processing means (equalizers, diversity combining, etc.), be reduced to Gaussian channels, and because this general behavior applies to many situations.

For spread spectrum systems, the bandwidth of significance is the information bandwidth, that is, the bandwidth following the correlation operations in the receivers. In a Gaussian channel, assuming perfect correlators and synchronization, the capacity of

spread spectrum link is exactly that of a non-spread transmission. Over dispersive channels, one can do better with OFDM (orthogonal frequency division multiplexing) or other techniques that specifically allocate bits and power to those portions of the spectrum with better SNR according to the prescription of Gaussian water-filling. Spreading merely gives an averaging operation (in effect, a blind bit and power allocation); one can do better with specific channel knowledge.

Relatively few capacity results are available for multiple access situations, although there are numerous bounds. The highest capacity is achieved if the different users can coherently combine their signals at the receiver, but this can be achieved in practice only with significant resources expended on synchronization. Spread spectrum multiple access that is coordinated in this way is very efficient. However, when tight synchronization and coordination among users cannot be achieved due to channel dynamics or multi-path, then orthogonal channelization such as TDMA (time-division multiple access) and FDMA (frequency-division multiple access) perform better than many CDMA (code-division multiple access) techniques, since self-interference is reduced. Nevertheless, if other interference may be present (e.g., from nearby clusters or cells) then spread spectrum techniques can be competitive or even superior to TDMA depending on the applications and signal processing techniques employed (voice activity management, multi-user detection, dynamic channel and power assignment, etc.).

One question of considerable importance is whether given the maximum signal processing effort an ad hoc network can scale. That is, given that each node added to the network generates more traffic, can that additional traffic be conveyed to any random node within some bounded region as more nodes are added? The answer to this question is no. Under the Gupta-Kumar model whereby nodes simply relay each other's traffic, the per node transport capacity grows linearly with bandwidth but drops as the square root of the number of nodes n in the planar region. Roughly, as n grows, the expected number of hops in a multi-hop link (and thus traffic to convey) grows as $\sqrt{n}$.

Therefore, each node added progressively increases the fraction of other nodes' traffic that must be relayed. It might however be argued that if the nodes were to cooperatively send traffic as in capacity-achieving solutions for simple multiple access then the situation might be better. Unfortunately, even for the situation of a single traffic stream across the network the improvement offered by having all nodes help is at best O(log n), which grows more slowly than $\sqrt{n}$. A similarly depressing situation arises with network delay.

Of course, such results apply not just to ad hoc networks but to general telecommunications networks also. This is one reason why these networks are designed hierarchically and to exploit spatial reuse. Each level within the network has a bounded number of traffic generating nodes, with successive levels having increased resources (bandwidth) to deal with the aggregated traffic. As higher layers typically also have longer hops, hierarchical networks have the further benefit of reducing end-to-end delay. Notice that in this scenario there is a separation of functions: only the lower level nodes generate traffic, while all higher layers function only to provide a communications relay.

Is hierarchy the only solution to the scaling problem of ad hoc networks? This depends on the goal. Directional antennas decrease mutual interference and increase ranges so that in fact within a given geographic region fewer hops would be required. While ultimately n going to infinity will win out, for many applications there may be sufficient capacity to not have to resort to wired infrastructure or relay nodes using additional spectrum. It might be argued that if the number of antenna elements per node grows as the rate of adding nodes to the network then in fact per node transport capacity will not actually decline, but then of course one runs into cost and size issues. A good research question is how far different wireless communications techniques can extend flat architectures.

Open research questions include:

- Tight bunds for capacity in multi-access environments
- Bandwidth relative to application environment
- Optimal hierarchical designs

5.1.2 Scaling for Sensor Networks

Scalability in sensor networks rests on the separation of functions among observation (traffic generation), signal processing (traffic compaction), and communications relays. Clearly, as the number of communication relays goes to infinity, the interference radius for each transmission tends towards zero using appropriate power control, and thus the number of independent messages that may be conveyed across the network goes to infinity. This would be for naught if the traffic generated increased at a faster rate. However, in observing any physical phenomenon, there is a finite amount of information that must be extracted for any given application. For example, consider a mapping application (e.g. of a chemical plume). Values need only be known to some number of bits of resolution and according to some spatial and temporal granularity. Once the sensor density is high enough to achieve this, adding more will not produce further traffic, under appropriate local processing to eliminate redundant information (e.g., selection of only one sensor per coordinate grid). Thus, the data to be extracted is finite for a given study region while the potential communications capacity is infinite. Viewing the nodes selected for observation as the lowest tier, and the relays as the upper tier, the scalability situation is remarkably similar to that of telecommunication networks: we keep adding telecommunication resources until the traffic demand is met.

What allows this to work is local decision making that determines what information is to be conveyed beyond some local region. This also promotes conservation of energy, as long-range communications transport is generally much more power-hungry than signal processing, being governed by both Shannon capacity limits and Maxwell's equations.

The fundamental performance limits to explore include the relative densities of sensors and communications relays assuming optimal local signal processing. This is a combination of rate distortion theory and multi-user capacity. In the regime of very high densities it is clear that the questions can be largely separated in the sense that most traffic will be local and so standard techniques will suffice for the long-range transport even if they are not optimal. The research problems become very complicated when

sensors are barely able to observe phenomena to within the desired granularity, requiring large-scale cooperation. A large number of interesting research questions remain to be solved both from the point of view of fundamental limits and practical algorithms that may approach such limits.

Open research issues include:

- Determining the optimal local signaling processing given sensor densities, information characteristics (e.g. uniformity), and network connectivity
- Determine the optimal distribution of processing (including fusion and data reduction) and communication given large-scale cooperation
- Determine the optimal local storage of sensed data for future retrieval
- Interaction of actuators in the sensor network

5.2 System Issues

5.2.1 Architectural tradeoffs

Pervasive applications can be broadly classified as either *user-centric* or *task-centric*. User-centric applications require a plug-and-play environment that allows the user to be oblivious of the system. *Smart space* applications belong to this class. Task-centric applications are the ones that are used to collect data, process data, and collect them in a central and/or distributed storage repository for use. Sensor networks support these types of applications.

Environmental information (the term environment is used here very loosely – the application defines the type of environment) is collected by sensors to be relayed to a repository. A sensor can communicate with other sensors or with an access point to the (wired) network to relay its data to the repository. A sensor device may have a limited processing capability to compress and/or filter the data before transmission. Processing of the data at the sensor consumes energy and requires storage (both energy and storage are scarce resources). On the other hand, transmission of uncompressed data without any processing consumes a larger amount of bandwidth, which is, again, a scare resource. It is then of vital importance to understand the tradeoff between processing and communication to decide on the alternatives. An important aspect of sensor networks is that they have a *goal* to which all sensors must contribute. Thus, the paradigm is intrinsically one of cooperation between sensor nodes. Another issue of great importance in sensor networks is that of deployment. In many cases, a rigorous placement of the nodes will not be possible and only a random deployment will be an option. This brings to the forefront a host of issues related to the structural characteristics, the deployment, and the maximization of the lifetime of sensor networks. There is a need to develop an appropriate modeling framework and provide insights on how sensor networks should be conceived, deployed, and run. This is with a view to designing robust and efficient networks in which nodes work collaboratively to achieve the common goal. The technological impact of this type of research will be to provide rules for the design and deployment of reliable, heterogeneous, energy efficient sensor networks.

This type of research is likely to impact network architecture, the design of energy efficient communication paradigms, the understanding of the impact of mobility in wireless systems, and the ways we can compress and aggregate information to make the network energy efficient.

Open research questions include:

- Determining trade-offs between processing and communication alternatives
- Determining the affects of structure and deployment of sensor networks on lifetime

5.2.2 System Management

One main appeal of a pervasive computing system is its potential to make our lives easier, simpler, and/or safer. Based on this promise, the use, configuration, and management of such systems should be simple, natural and effortless, especially for the user-centric systems. On the other hand, wireless systems, by nature, are very complex. The disparity between the ease of use and the complexity of the system can be bridged by a powerful (and still simple to use!) management system. To be effective, the management system (which essentially runs at the application layer) should be aware of the intricacies and events at the lower layers of the communication stack. In essence, there is a need for provision of a cross-layer signaling and protocol in the communication stack. These functions should be easily available to the pervasive application, so that it can reconfigure and tune itself, based on the feedback from the environment (carried across the layers), with the user interaction.

Ad-hoc networking enables number of computing devices to establish (and maintain) a network among them as discussed previously. While the use of ad-hoc networking among a large number of mobile stations has not materialized, there is a high potential of use in environment in which nodes are not mobile, but are randomly deployed. In addition, certain configurations might be more likely to be of use. One such configuration is where a node can either communicate with an access point directly (one-hop path), or through only one intermediate node. Research on the management and use of these networks is important.

5.3 Energy-efficiency and Energy-awareness

Energy has been well recognized as an Achilles' heel for pervasive wireless systems, and has led to its emergence as a system-level performance metric that is at least as important if not more than traditional metrics such as capacity, latency etc. Unfortunately, there is no equivalent of the Moore's Law with its fast exponential growth when it comes to the battery technology. As important as reducing the energy consumption is to manage the available energy via energy-aware management of the system resources. Moreover, the incremental energy cost of sending a bit even over short distances is several order of magnitudes greater than the energy cost of executing a primitive digital operation. Therefore, usually, the energy consumed in wireless communications dominates the energy cost of application related computation. Unfortunately, wireless network models used in WLAN, WWAN etc. are inadequate for the wireless embedded devices that

constitute pervasive wireless systems. Traditional wireless network models rely on radio and protocols designed for high-persistency and high duty-cycle operation and optimized for bit-transport instead of the collaborative in-network processing that characterize pervasive wireless systems. Excessive protocol layering and indirections, and sophisticated management of network connections and nodes are other reasons why one cannot just adopt traditional wireless network models.

5.3.1 Physical-layer Considerations in Protocol Design

What is needed is an understanding of the coupling of protocol design with radio and application characteristics and therefore a proper understanding and modeling of the latter two and their differences from traditional wireless networks. Focusing on the energy aspects of radios, it is worth noting that the energy consumption of wireless communication subsystem consists of three components with substantially different characteristics: transmit electronics, transmit RF power amplifier, and the receive electronics. The relative importance of these depends on the transmission range of the radio. In particular, the energy consumption in the RF power amplifier increases with transmission range, and dominates the overall power consumption in the case of devices such as cell phones. In pervasive wireless systems, the ranges are much shorter and therefore the electronic power consumption is significant or even dominant.

For example, a good first order radio transmitter model for energy per bit that captures the effect of electronics is ETx = a + bdn where the coefficient a corresponds to the radio electronics power consumption and the coefficient b captures the energy spent in the RF amplifier and is a function of the RF transmit power setting. Usually the first term is considered to be much smaller than the second term, and such is indeed the case in cellular systems. But in radios commonly used for pervasive wireless system the ratio a/(bdn) is currently in the range 1 to 5 thus indicating that the electronics related energy consumption is at least as important as the RF transmit power related factors, and often much more. Moreover, while progress in microelectronic technology will help, it will not eliminate the problem as the power consumption in the analog electronics in radios is not expected to benefit nowhere near as much as the benefits digital electronics derives from shrinking technology and reduced voltages.

The simplistic radio models commonly used in current networking research ignore the significant impact of electronics on energy consumption and radio performance, and thus lead to rather misleading results about protocol performance. For example, some of the assumptions underlying traditional wireless protocols are that transmission is costlier than reception, that idle listening is cheap, and that radio energy consumption scales as dn (which implies that multi-hop routing and the use of stronger codes and low-valued M-ary modulation with lower Eb/N0 saves energy). But these assumptions breakdown in pervasive wireless systems such as sensor networks, and thus render the traditional wireless protocols based on them unsuitable. For example, the event-driven nature of sensor network traffic means that much of the time the radios do not have any traffic.

If the MAC protocol were to just leave the radio idle during such time (as protocols such as CSMA/CA do), then the energy consumed during idle state overwhelms the energy

spent in actual packet transmission or reception so that higher layer optimizations have no benefits (e.g. a smart multicast will not be any better than simple flooding). As another example, the significant energy consumed by the receiver electronics (which is comparable to and often more than that consumed by transmitter electronics) means that multi-hop is often less energy efficient than direct transmission at distances typically found in pervasive wireless systems.

Open issues include:

- Coupling protocol design with radio and application specifics
- Study of the interaction of system design and electronics
- Development of realistic radio energy model

5.3.2 Radio Requirements for Pervasive Wireless Systems

A key implication of the radio electronics energy dominating the RF energy is that classical communication techniques to lower power consumption by trading energy against latency do not work. For example, were RF energy to be dominant, one could reduce the energy by stretching the data transmission over the available time by using FEC codes which provide coding gain or using lower rate modulation schemes with lower Eb/No. However, when the electronic energy is significant, as is the case in pervasive wireless systems, these techniques are ineffective and instead the best strategy is to transmit as rapidly as possible and simply shut down the radio for the remaining available time.

While radio shutdown is the preferred strategy for managing radios in pervasive wireless systems, unfortunately traditional radios are not designed with this in mind. One significant issue is the transient start-up time that the radio takes when going from shutdown to active (or idle) state. The various circuits such as PLLs and frequency synthesizers take time to settle, and often this time is comparable to the duration of the packets in systems such as sensor networks where the packet sizes are small as they primarily carry event information. The energy spent in the transient startup time constitutes energy overhead for each bit communicated. Clearly, radios with fast start-up and acquisition are needed for pervasive wireless systems such as sensor networks.

The key to effective shutdown-based energy management of radios is the ability of a sender radio to wake up the destination radio in a data-driven fashion. No such effective wake-up mechanism currently exists, and protocols rely on duty cycling whereby a shutdown radio periodically wakes up to poll. An energy-latency trade-off manifests itself as a function of the polling period. An alternative to consider is a separate wake-up radio module that is designed to have an ultra low idle mode power and handles very low-data rate sporadic wakeup messages with a small number of bits. While there have been a few attempts in this direction (e.g. using a simple energy detector), no satisfactory solution that is also robust to false wakeups currently exists. In addition, related technologies can help to eliminate false wakeups and minimize their impact. Examples include cheap directional antennas that can focus wakeup messages towards the desired node, and smarter radio basebands that not only match the packet preamble but also the

address field to decide whether to keep the radio up for receiving the remainder of the packet.

Another alternative to consider is the scheduling of receivers when data transmission patterns are deterministic, or can be inferred with high probability (e.g. periodic sensor-to-monitor transfers.

Key open research questions include:

- Optimal radio design considering energy management requirements and application characteristics
- Alternative techniques for receive power control, including event-driven wakeup and scheduled

5.3.3 Beyond Energy Reduction: Energy Awareness and Energy Harvesting

Much of the focus has previously been on energy reduction in the form of either reducing the energy to send a bit, or to reduce the number of bits that need to be sent to perform the task. However, pervasive wireless systems, by their very nature, offer other possibilities that may be exploited. First, a key function of these systems is to answer a query by estimating some spatio-temporal function of the observed sensor values of the physical world. This opens up the possibility of energy-aware protocols that trade-off energy against other metrics of system performance such as accuracy, coverage etc. by intelligent adaptation of radio and protocol control knobs. Second, with the reduction in electronic power consumption it will soon become practical to have sensor nodes driven by energy sources that harvest energy from the environment such as solar, wind, salinity, temperature differential, and vibration.

A conventional energy source may still be present to act as a buffer or reservoir of energy to smooth out the temporal variations. However, environmental energy availability is not the same at all the nodes. For example, the level of wind or the intensity of sunlight is a function of the node location. This opens up the possibility of protocols that learn the spatio-temporal characteristics of the environmental energy availability, and then factor that in allocation of networking tasks to nodes, such as selecting paths for multi-hop routing. Such protocols would result in better performance than those that simply minimize energy or take only the remaining battery energy into account.

In summary, energy efficiency in the context of pervasive wireless systems requires protocols that are keenly aware of the bare metal and factor in radio electronics characteristics, channel characteristics, and energy availability. This requires radio models that are much more sophisticated than the ones currently in use by networking researchers. In addition, the radios themselves need to be optimized for the distinctive traffic characteristics of sensor networks. Reduced start-up time and wakeup radio modules are desired features. Protocols that go beyond energy reduction and exploit environmental energy intelligently, and which eliminate excessive layering and indirections are also desirable.

6 Wireless Security

Wireless, pervasive systems carry several unique characteristics. They operate over an open medium that is inherently shared, incorporating devices of modest capabilities. These systems are expected to operate at scale, with constituents coming and going at will. Together, these characteristics impose a number of interesting security challenges: negotiable, adaptive security policies, defense in depth, and support for transient relationships.

6.1 Adaptive Security Policies

There are a wide variety of applications envisioned for ad hoc networks. These include emergency search-and-rescue, military operations, conferences, data collection in remote locations, classrooms, and general spontaneous networks. The security requirements of these applications are dictated by the location and type of data exchange, and consequently are highly varied. For instance, in a network formed by military nodes in a battlefield, there are security threats to both the physical safety of the nodes, as well as to the communication between the nodes. Every node in a military environment is vulnerable to physical capture and over-run of equipment. Once compromised, hostile entities can then pose as friendly entities at the compromised node. Therefore, exposure of node location is undesirable so that enemies are not able to locate military personnel. In addition to threatened physical safety, the military nodes are prone to fabrication attacks against both control and data messages. These networks must ensure that sources, destinations, and all intermediate nodes can be authenticated, and that the integrity of the messages can be verified.

Ad hoc networks on college campuses or at conferences, on the other hand, do not suffer the physical security threats experienced by military networks. While nodes in these more friendly environments can be stolen, their capture does not represent a network vulnerability since these networks are typically open to general participation. Instead, users communicating on a college campus are likely to want data security and message integrity, as well as authentication of sources and destinations.

The provision of network security adds communication and processing overhead and complexity to a network. Mechanisms used to provide network security, such as digital signatures, encryption and hashes, add extra bytes to control and data messages. They also require additional processing at source and destination nodes; intermediate nodes are often impacted as well. Nodes in an ad hoc network are typically resource poor; they have limited processing ability, limited battery lifetime, and limited storage. Further, wireless networks suffer from low bandwidth and high channel error rates, making the transmission of large packets difficult. Because of these factors, the incorporation of security mechanisms into the ad hoc network can result in severely degraded network performance.

The negative impact of the security mechanisms can be mitigated by the inclusion of only those security mechanisms essential to the operation of the network. For instance, while it is essential that the location of military nodes be undisclosed (LPD – low probability of

detection), users of civilian networks do not have the same concerns. Hence, civilian networks should not deploy security mechanisms to hide location because it will result in unnecessary security overhead. The security requirements of users in an ad hoc network are influenced by a number of factors. These include the location and application of the users, as well as the domain of network use. By deploying only the essential security requirements, the additional overhead is minimized while the needed security is obtained.

The composition and application of a network influences whether initialization of security parameters can occur before the nodes enter the network. For instance, in military, search-and-rescue, and classroom environments, the ad hoc network will typically be formed of specific, known users. This enables participants in these networks to exchange security information before entering the network. This a priori exchange of information allows the use of session keys and the establishment of security associations. In other networks, where the identity of the participants is not known ahead of time, it is difficult, if not impossible, for key exchange to occur in advance. This type of scenario may occur, for example, for a user walking through an urban environment or driving on a highway.

Systems can be classified into three distinct security environments: open, managed-open, and managed-hostile. These environments are distinguished both by their security requirements and by their opportunity for pre-deployed exchange of security parameters. In the open environment, random nodes establish and maintain connectivity to other random nodes without any trusted third parties in common. Because nodes are often communicating for the first time, it is unlikely that key exchange can occur in advance. In the managed-open environment, however, nodes participating in the network do have the opportunity for the pre-deployment exchange of security information. This type of network can occur on a college campus or by colleagues at a conference. Finally, the managed-hostile environment also offers the opportunity for the exchange of security information. However, this environment is distinguished because nodes are vulnerable to physical capture and location information must be concealed.

In summary, the security requirements and solutions are influenced by a number of factors, including the application and location of the network. Because security mechanisms are heavyweight, they result in early depletion of node battery and increased consumption of network resources. To reduce the detrimental impact of the security solutions on the network while still providing the needed security, networks should only include the security solutions essential for network operation. Each node should include support for a variety of security techniques, but should only utilize those that are appropriate to its current environment.

Open questions include:

- Defining appropriate structure for adaptive security systems
- Designing adaptive systems to be interoperable
- Designing mechanisms and protocols that are adaptive

- Understanding the tradeoff between needed levels of security and the resource cost of implementation

6.2 Defense in Depth

Wireless and mobile networks have several characteristics different from wired networks. These characteristics make a case for building multiple-fence security solutions. In a wireless network, users access the network through the wireless channel. The wireless channel is open, and is thus accessible not only to legitimate users, but also to eavesdroppers and malicious attackers. The boundary that separates the inside network from the outside world becomes blurred. In fact, some popular wireless networks, e.g., a mobile ad-hoc networks or a large sensor network, do not even offer a clear line of defense from the security design perspective. Each node may be a host, but serve as a router for others at the same time. There is no well-defined place in the infrastructure where we may deploy a single security protection scheme to guard against the possible threats. The overall security solution will spread across many individual components and rely on their collective protection power to secure the entire network. Moreover, individual network devices, ranging from desktops, laptops, to PDAs and smart phones, may have different degrees of resource constraints that prohibit them from deploying a single, powerful security solution.

The security scheme adopted by each device has to work within its resource limitation. This also calls for a multi-fence security solution, each of which adapts to the individual device's capability in terms of computation, memory, communication capacity, and energy supply. Furthermore, portable devices used in wireless networks may be more vulnerable to compromises or theft. It is quite possible for attackers to target low-end devices, which could not deploy a high-end security solution due to constrained resources, and subvert a few of them. These devices may pose as the weakest link of the entire system and incur domino effect for security breaches. A multi-fence solution can also help to localize the threat damage and guard the system from collapse. In addition, a multiple-fence solution offers the benefits of extensibility, modularity and portability, and improved opportunities for incremental deployment. These features fit well with the device and network heterogeneity in the emerging wireless Internet that spans indoor WLAN, outdoor WWAN, to global satellite networks. Finally, it is expected that the ultimate solution to system security has to span different layers of the protocol stack, where each layer may contribute to a line of defense. No single-layer solution is possible to thwart all potential attacks. This also calls for a multiple-fence design approach that spans the protocol stack and distributes among different devices.

In the envisioned multi-fence security system for wireless networks, security is built into possibly every component, resulting in an in-depth protection solution that offers multiple lines of defense against many possible security threats. The individual fence adopted by a device may vary in security strength depending on the available resources, deployment cost and complexity concerns of the device. The system does not stipulate the minimum requirement that a component must have. Instead, it expects a "best-effort" approach from each component. The more strength a component has, the higher degree of security it has. The design of each individual fence may again take new approaches. One possible

approach is to build all function elements of prevention, detection/verification, and reaction into a designated fence. The prevention element, e.g., through authentication or encryption, guards against some popular security threats. The detection/verification element further monitors the system for suspicious or malicious operations. This helps to protect from insider attacks possibly due to compromised or stolen devices, or un-anticipated attack at the design phase. Once such threats are detected, the reaction element may invoke certain actions to localize the damage, warn the rest of the system, and minimize the system performance degradation.

Another potential direction is to take a resiliency-oriented approach for protocol design. The state-of-the-art protocols are typically designed for functionality, rather than for security or resilience. Therefore, they typically carry only minimal information and perform only necessary operations for system functioning. In a resiliency-oriented design, a protocol may incorporate additional information and perform additional operations for security purpose. At each step of the protocol operation, the design makes sure that what it has done is completely along the right track. Anything deviating from valid operations is treated with precaution. This way, the protocol tells right from wrong because it knows right with higher confidence, not necessarily knowing what is exactly wrong. The design strengthens the correct operations and may handle even unanticipated threats at the runtime operations. At the system level, the design may also take a paradigm shift from conventional intrusion prevention to intrusion tolerance. In a sense, certain degrees of intrusions or compromised/stolen nodes are the reality to face, not the problem to get rid of, in wireless network security. The overall system has to be robust against the breakdown of any individual fence, and its performance does not critically depend on a single fence. Even though attackers intrude an individual fence, the system still functions, but possibly with graceful performance degradation. A possible way to achieve this goal is to let the system rely on the collective behavior (e.g., through consensus or majority voting) offered by all individual components.

Open research issues include:

- Defining the proper security fences
- Evaluating the effectiveness of each fence and the minimal number of fences that the system has to possess to ensure some degree of security assurances should be evaluated through a combination of analysis, simulations and measurements in principle.
- Development of effective analytical tools, particularly in a large-scale wireless network setting
- Exploring multi-dimensional tradeoffs among security strength, communication overhead, computation complexity, energy consumption and scalability
- Developing effective evaluation methodology and toolkits through interdisciplinary research

6.3 Transient Relationships

Encounters between devices in wireless, pervasive systems are typically transient. A security infrastructure cannot impose undue friction on these encounters without limiting the applications. Such short-term encounters are complicated by the fact that devices and services come from many different administrative domains, and any principal is likely to have pre-established relationships with only a handful of these domains. In all of these concerns, usability for the end user must be paramount.

As devices encounter one another, they must negotiate for any services provided. To do this, one must reliably name the devices and services in question. Unfortunately, this presents a problem. Given the scale of devices expected, and the diversity in administrative domain, it is impossible to expect each principal to know the names of all other principals and devices in advance, and in fact some of them may be anonymous by design. Thus, naming itself is a first-class problem in pervasive security infrastructures. The research community will need to explore a variety of ways to provide meaningful names at this scale, and to ensure that names can be strongly bound to the services, devices, and principals that they purport to identify.

Once a set of collaborating devices is found, they must negotiate for services, including any required security infrastructure. In order to support short-term services, this negotiation must be lightweight, preferably requiring no user involvement. Users often are unwilling to bear even minor inconvenience in the name of security, and are likely to be even less tolerant when the protected interaction is a short one.

These negotiations are complicated by the expected diversity of administrative domain. To see why this is a problem, consider the difficulties one encounters when moving to a new city. The utility providers: phone, electric, gas, water, etc. are all likely to be different from those in one's old home. The new resident must establish credit and accounts with each new provider rather than carry over old relationships. The same diversity is likely to be present in pervasive settings. However, a heavyweight establishment protocol will make these services unusable, limiting the value that pervasive services can provide to users on the move. Instead, some notion of hierarchical service establishment or resource brokering will be required to enable these services.

A related problem is the inability to properly punish greedy or malicious parties. Returning to our utilities example, a customer who does not pay can be cut off. Since most utilities are geographically determined, this customer will not be able to obtain services elsewhere. However, this protection is unlikely in wireless relationships, which are transient and involve potentially many different domains of control without geographic ties. Solving this problem is critical to the practical success of any pervasive infrastructure.

Open research issues include:

- Defining effective naming conventions
- Security negotiation mechanisms

- Defining hierarchical structures for security systems

7 Evaluation

7.1 Networking and Communications Tools

A systematic and scientific research methodology is important for any discipline to thrive. Workshop participants felt that a national facility, or facilities, that provide a set of common core service, would be very useful for advancing wireless networking research. The workshop participants envisioned that this national facility would provide key support for academic research and would facilitate interactions between academia and industry.

This national facility would provide hardware and software support for experimental tools. For instance a networking protocol that addresses all of the important issues in wireless networking would take significant resources to develop under a research grant. Similarly a programmable radio would be a valuable resource. The work force that might have developed this protocol or radio is likely to be transient (a graduate student or post-doc) and not likely interested or available to support future. A national resource could take a hardware or software tool into its inventory and assume support for this tool if it was proving useful to the research community. This kind of support would remove the non-academic burden from the university research groups and allow a significant number of software tools to be reused by the community. This would allow significant infrastructure for experimental research to be developed and to be reused by a wider academic community. The key idea would be to provide a centralized institution that would provide "corporate" memory and technical support for realistic wireless experimentation.

It is envisioned that this national facility would also support significant prototype development. Network processor cards, radio cards or multiple antenna testbeds could be a major component in an experimental research program and yet there might be no interest or expertise in building these realistic and fully operational building blocks in university research groups. The national laboratory would be envisioned to follow a model like Fermi or Argonne Laboratories in physics. Significant experimental equipment development would be concentrated in a national center that would share these resources to research groups nationwide. Few universities can generate enough research to keep a full time radio engineer, or a full time software staff , or a high speed processor card developer employed. The national research effort has significant needs for such engineering talent if we really want to deploy and experiment with state of the art wireless networks.

The workshop participant felt strongly that this national testbed facility should not be funded out of current program funds but should be a part of a consistent national effort to bring more research funding to bear on the important problems in information technology. It is important to have both theoretical work focused on making fundamental contributions and experimental work to test out theories and reveal the flaws in current approaches. We need to come together as a community to lobby for such efforts with

national organizations and funding agencies (not just NSF), and reduce the hypercritical nature of the review process as this hurts networking communities chances in larger multi-disciplinary programs.

In the following subsections we describe some of the key services this facility might provide in more detail.

7.2 Tools

Radios. In recent years, there is a growing realization in the research community that, to optimize performance of wireless networks, it may be necessary to improve cooperation between the physical layer, and the upper layers. Several protocols have been developed that attempt to utilize physical layer information to adapt upper layer behavior, and vice-versa. Evaluation of such protocols is often handicapped due to the unavailability of suitable radios. Often, the wireless devices are incapable of providing necessary information to the upper layers; similarly, often the upper layers are unable to exert adequate control on the behavior of the wireless devices. To alleviate these shortcomings, two approaches may be used:

(a) develop a programmable radio, whose behavior and interface to the upper layer can be customized relatively easily, or

(b) nurture a supplier for radios designed to researchers' specifications at low cost. In the absence of suitable devices, researchers will not have adequate resources to performance experimental research with wireless networks.

Network Controllers for Testing. The running of actual wireless experiments with laptops, PDAs, etc have been done many times by researchers, but due to the difficulty in scheduling repeatable movements in a realistic environment, these tests are usually done with just a few nodes in just a few selected areas. One recent direction numerous groups have undertaken to improve this situation is the construction of static testbeds of machines that individually run the actual designed protocols as if they were actually in the target environment. To simulate movement or the wireless channel, these testbeds use either two Ethernet sub-networks – one for control information, and one that simulates the wireless channel by appropriately dropping packets from nodes that are out of range; or the testbeds use actual radios connected to an automated radio attenuation device. This second option is extremely expensive and is typically only used by researchers in industry.

Testbeds running actual networking software are very useful for researchers, but so far there have been three major problems. First, there has been no concerted effort to unify and agree on a particular Ethernet controlled testbed. There are multiple efforts, each with their benefits, but without a single group coordinating the efforts, there is much duplication. Second, effective and realistic RF propagation loss models are not used. This is partly due to a lack of understanding from the networking community, and partly due to a lack of appropriate mobility and radio models for realistic situations. Third, these systems do not allow for realistic models of MAC performance. They maintain simplicity by simply using a path loss matrix broadcast to all nodes. Nodes therefore drop traffic

only when they are out of range, or their queues back up. They do not loose packets due to collisions or accumulated noise. Implementing an environment that allows reasonable MAC evaluations in such a situation is tricky because one would like to maintain a completely distributed system, when in fact the wireless channel is really a single shared resource. Investing research funding into this area could provide significant benefit to the community.

Simulators. The state of wireless network simulation is better than it ever has been, but has hit significant issues that need to be resolved before the use of simulation in research and design can be increased. Detailed simulation tools that allow both wired and simulated wireless environments are freely available. Additionally, the research community has donated dozens of open-source models that allow researchers to study and compare against existing established protocols. Unfortunately, the wireless channel is a particularly complicated physical environment. Radio propagation is affected not only by the static objects in the environment (buildings, trees, etc), but also by other moving nodes. Additionally, the detection and decoding of the radio signal in the presence of other interfering signals can be quite complex.

There are three major simulation systems used within the research community -- OPNET, Qualnet and ns-2. OPNET and Qualnet are commercial products. They are free to universities, though there is no guarantee that they always will be. ns-2 is an open source package that has been well received and used by the research community. However, many in the community have noted that ns-2 is beginning to show many limitations. Specifically, 1) It was originally designed to be a simulation tool for wired networks, so it's support of complex wireless propagation and radio simulation is limited, and 2) Limited tools are available for creating mobility traffic, and performing analysis, and; there is no central maintainer of the models for ns-2 and there seems to be no funded group that is actively maintaining/expanding/improving ns-2.

The workshop felt that a significant simulation capability that improves upon existing tools and does not completely rely on a commercial entity was extremely important. Funding for significant improvement and maintenance of ns-2 would be quite beneficial. Alternatively, some group may wish to start from a ground up approach for an “ns-3” simulator. This would need to maintain backward compatibility or have very broad support within the community to be successful.

Engineering Support. Not all research institutions have the expertise or resources to be able to assemble systems, particularly the hardware components, such as processor boards, etc., which are often required for experimental evaluation. To facilitate experimental research at a larger number of research institutions, therefore, it will be useful to develop a shared engineering facility (or facilities) that will house expertise as well as equipment that may be necessary for assembling necessary hardware. Similar facilities have been developed to support other areas of research; for instance, the MOSIS facility that provides an integrated circuit fabrication service.

National Testbeds. The networking research community also felt it important to establish some national testbeds. These testbeds would be national resources for large

controlled experiments. The testbeds might be in relatively benign interference environments where the wireless experiments could be carefully controlled. Alternatively a large collection of radio channel emulators might be established to provide a repeatable test environment. Either of these types of facilities would provide for a large scale and repeatable test environment. This environment could be time multiplexed by researchers whose research required verification in large-scale experiments much like national telescopes are now.

7.3 Modeling for Simulation

7.3.1 Measurements to Models

A mobile network is inherently subject to uncertainty owing to mobility, bursty traffic patterns, varying radio propagation and cross-layer protocol interactions. As a result, a complete model of a mobile network must account for mobility, traffic, radio propagation, and protocols. With respect to propagation, there is a vast body of literature on the measurements and modeling of propagation in various environments at various frequencies. The characterization of the channel response depends on performance assessment needs: it may be site-specific, that is, based on actual measurements or a ray-tracing approximation; or it may employ stochastic processes with values and correlations mimicking those measured in a class of similar environments.

There are fewer measurements and less agreement regarding models for mobility and traffic, in large part, because of the great sensitivity of these models on the network application. For example, mobility models for sensors are likely to be quite different than those for cellular voice networks. Finally, there is very little understanding of the cross-layer interactions and its impact on relevant performance metrics at both the user and the network level. Given the great number of dimensions along which characterization of mobile user and network models is necessary, there is a need for a systematic design and evaluation process that will seamlessly transition from measurements to models. The range of approaches to make this feasible include: a full-blown large-scale life-size testbed with detailed implementation of traffic, mobility and protocols; a miniaturized version of the same with suitable abstractions that still retain a holistic value; measurement based simulators and emulators at various granularities (and protocol layers) for testing and verification of models and their abstractions.

7.3.2 Model Verification

Modeling of radio networks is very distinct from modeling of wired networks in that the physical channel properties, i.e., radio propagation and interference, cannot be separated from the higher network protocols because of strong interactions at all levels. To verify a model, a wireless system simulator/emulator must have the ability to model radio phenomena at different temporal and spatial scales, as well as the ability to model multiple protocol layers. The levels of granularity are dictated either by the hierarchy of transmission units in the system (e.g., sessions, flows, frames, packets, bits, or chips), and by characteristic time-scales for the physical processes with which transmissions interact (e.g., short–scale Rayleigh fading, or long-scale shadow fading as well as interference).

Moreover, a wireless simulator or emulator should support research studies across levels and time-scales. Such a system would facilitate the incorporation of interactions between layers in the design and optimization process.

7.3.3 Test Vectors and Traces

For radio receiver simulation, the only test trace needed is an independent symbol sequence. For wireless network simulation, a test trace must specify the mobility, traffic and radio propagation. Test vectors and traces simply do not exist for most cases of wireless networks at least in the context of non-cellular like systems. Generation of such test vectors and traces would again require a range of approaches similar to that outlined above for transitioning from measurements to models.

7.3.4 Appropriate Abstractions

The performance of a radio receiver depends on receiving sufficient signal energy from a desired transmitter while not receiving too much interference from other sources. Still, an abstraction for the physical layer must capture the key elements of a given transceiver technology. For a large class of packet based systems and protocols, a suitable abstraction for link quality is the signal to interference ratio (SIR).

Radio resource management includes any actions that are necessary in a wireless system to provide acceptable quality over the radio link. In wireless voice networks, these actions include protocols that control call admission, channel allocation, power control, and handoff. In a wireless data network, there are also transport protocols acting above the resource management activities. Almost all aspects of radio resource management depend on accurate measurements of the quality of the wireless link. Further, the evaluation of a radio resource management protocol dictates SIR measurements on a time scale consistent with the operation of the protocol. For example, simulations of inter-cell handoff protocols employ SIR measurements averaged over seconds whereas packet-level simulations need multiple SIR values on the short time-scale of a packet transmission to determine if the packet's data have been corrupted.

We note that in the widely used ns-2 simulator, abstractions for the physical layer are perhaps a decade behind physical layer research. In the past decade, there has been tremendous development of methods of opportunistic communication that exploits the variability of the channel in time, frequency and space to achieve higher average data rates. There has also been a concurrent advance in multi-antenna transmission. Although physical layer abstract models based on SNR/SIR measurements have emerged in integrating resource allocation with transceiver technology, these models do not exist in network layer simulators. Thus the benefits of recent advances in physical layer communication with respect to higher layers are yet not well understood and must be determined in a systematic fashion.

Open issues include:

- Experimentally validated mobility models for various wireless applications

- Flexible simulators that allow multiple levels of abstraction and variable degree of granularity

8 Summary of Workshop and Acknowledgements

This report reflects the views of the participants of the NSF Wireless Networking Workshop. It contains high-level recommendations, such as the creation and support of a wireless networking community, creation of national facilities, and guidance on the types of research and education to be fostered. It also contains detailed descriptions of the technical areas in which it would be worthwhile for NSF to invest in the view of the participants. The workshop participants stress the importance of wireless networking research and feel a critical set of resources must be dedicated to solve these research problems.

We would like to acknowledge Prof. Joseph B. Evans and Dr. Mari Maeda who supported and helped organize this workshop. Dr. James P. G. Sterbenz provided valuable comments to an early draft of this report. We would also like to acknowledge Gary Minden for his great assistance with logistics in support of this workshop.

Appendix A: NSF/FCC Workshop on The Future of Spectrum

May 2003

Summary of Radio Discussions

In May 2003 the National Science Foundation and the Federal Communications Commission sponsored a workshop on The Future of Spectrum: Technologies and Policies. The notion of a cognitive radio, and policy implications, were discussed at length. A radio that can learn its surroundings and be nimble enough to adapt to it is of tremendous value in improving network throughput, and opportunistically exploiting unutilized spectrum.

In general, a cognitive radio cannot be separated from the networking aspects that go along with a highly agile communication system. Any future research that uses the cognitive radio concept must simultaneously develop novel data link layer and networking concepts to fully exploit the potential of high throughput opportunistic wireless communications.

Some discussions regarding the research topics associated with a cognitive radio resulted in the following topics:

- Bandwidth, frequency and waveform agility
- Power efficient, programmable base-band processing engines
- Multiple antenna capable
- Novel RF architectures
- Interference level or Interference temperature measurements
- Learning algorithms that will help the radio adapt to e
- A common API to higher layers

What follows is a summary of different presentations throughout the day.

The future of spectrum and radio design

Telecom applications have an insatiable thirst for increasing data rates. This appetite has been quenched for wireline systems where Gbps connectivity within a Local Area Network is now a thing of the past and companies are looking towards 10 Gbps connectivity over longer distances. In the wireless space, however, the state of the art is three orders of magnitude lower than the wireline space. 802.11 wireless LANs are struggling to deliver 10's of Mbps connectivity at distances of greater than 200 feet. This gap is one of the fundamental challenges of wireless data communication systems.

Traditional approaches to increasing throughput rate have been to increase bandwidth. However, the scarcity of bandwidth at lower carrier frequencies due to regulatory issues that prevent sharing of bandwidth are forcing the wireless industry to look at alternatives.

Two such alternatives are the use of multi antenna technologies, namely MIMO, to increase spectral efficiency, and migration to higher carrier frequencies where regulations are less strict and more bandwidth is available.

Multi Input Multi Output (MIMO) technology creates independent spatial channels by modeling the wireless channel by a matrix impulse response. Exciting the channel with a vector and then sampling it at the receiver with a vector allows us to transmit different information along spatially unique paths. Preliminary experimental results have shown that MIMO systems can easily achieve spectral efficiencies of 10 to 15 bps/Hz. In one instant, a 15x15 system developed at Bell Labs has demonstrated 40 bps/Hz. The challenge in MIMO systems is the increased number of RF and antenna elements needed, as well as the order of magnitude higher processing needed to invert the matrix-channel.

Migration to higher carrier frequencies is a second approach to achieving higher bandwidths. This requires more advanced semiconductor processes that can provide gain and functionality at higher frequencies. Currently researchers are pushing the envelope with regards to CMOS technologies and carrying out preliminary studies to evaluate the utility of CMOS technologies at higher frequencies. Apart from the electronics, another challenge to higher frequency communications is the higher attenuations associated with wireless propagation above 10 GHz. This requires rethinking of the overall wireless system architecture. Typical applications at these frequencies are: traffic back-haul, curb-side to home “last 100 meter” connectivity, and building to building communications.

In regards to the frequency and bandwidth agility required by cognitive radios, the emergence of software defined radios and direct conversion receivers have provided the basic building blocks necessary to make cognitive radios a reality. Under the DARPA GloMo program, Rockwell Collins developed a unique direct conversion radio with the following parameters: tunable frequency from 20 MHz to 2.5 GHz; up to 10 MHz of bandwidth; 1 Hz tuning steps; 100 micro-second tune time; 20 cubic inches.

Aside from the RF and transceiver issues, a cognitive radio must be waveform agile. This requires flexibility of the baseband processing engine. The immediate answer to the need for reconfigurable baseband engines is to use a general purpose programmable device such as a micro processor, a DSP or an FPGA. These devices provide varying degrees of flexibility, however, there is a cost associated with this level of flexibility. For wireless communications, power consumption is a major concern and must be factored into the flexibility equation. An appropriate cost metric will thus look at the amount of energy needed to deliver one MOPS (Mega Operations Per Second) of processing power. Plotting this metric for a class of microprocessors, DSPs, and dedicated ASIC solutions is shown below.

Chip #	Year	Paper	Description
1	1997	10.3	?P - S/390
2	2000	5.2	?P – PPC (SOI)
3	1999	5.2	?P - G5
4	2000	5.6	?P - G6
5	2000	5.1	?P - Alpha
6	1998	15.4	?P - P6
7	1998	18.4	?P - Alpha
8	1999	5.6	?P – PPC
9	1998	18.6	DSP - StrongArm
10	2000	4.2	DSP – Comm

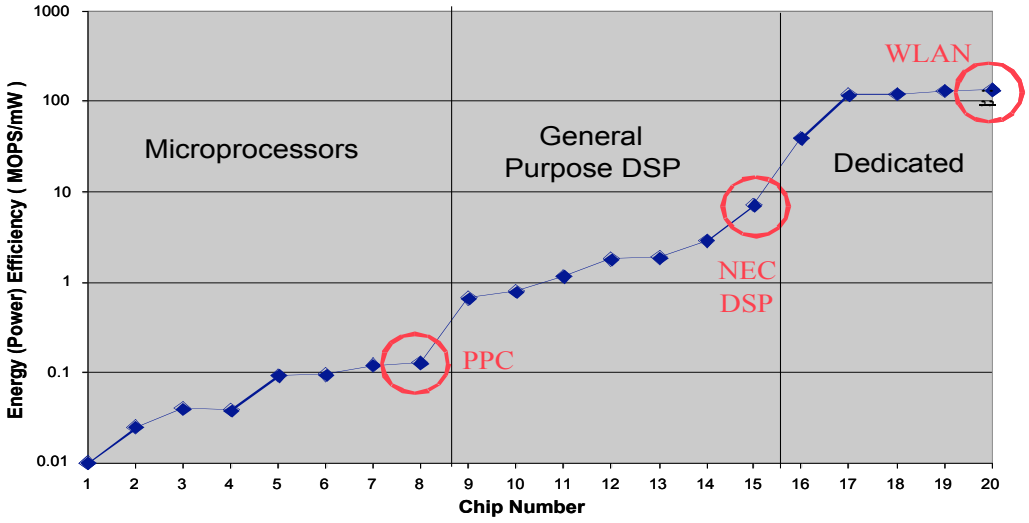
Microprocessors

DSP's

Chip #	Year	Paper	Description
11	1998	18.1	DSP -Graphics
12	1998	18.2	DSP - Multimedia
13	2000	14.6	DSP – Multimedia
14	2002	22.1	DSP – Mpeg Decoder
15	1998	18.3	DSP - Multimedia
16	2001	21.2	Encryption Processor
17	2000	14.5	Hearing Aid Processor
18	2000	4.7	FIR for Disk Read Head
19	1998	2.1	MPEG Encoder
20	2002	7.2	802.11a Baseband

DSP's

Dedicated



It is interesting to note that on average, general purpose microprocessors are two orders of magnitude more power hungry than general purpose DSPs which in turn are an order of magnitude worse off than dedicated ASIC solutions. It is not clear as to where exactly the line is to be drawn in terms of satisfying the base-band processing needs of cognitive radios.

Conclusion

The next frontier in radio research is the development of cognitive radios, i.e., radios that can sense and adapt to the electromagnetic environment around them. To do this, cognitive radios must show tremendous versatility. They must be bandwidth agile,

frequency agile, and waveform agile. Additionally they must be able to sense their environment and opportunistically send information across. Although a challenging research problem, cognitive radios could significantly improve spectrum utilization by enabling opportunistic communication across many bands.

Appendix B: Participants

Report Contributors:
Babak Daneshrad, UCLA
Brian Noble, Univ. of Michigan,
Catherine Rosenberg, Purdue
Christopher Rose, Rutgers
Elizabeth Belding-Royer, UCSB
Greg Pottie, UCLA
Henning Schulzrinne, Columbia
James Sterbenz, BBN
Jason Redi, BBN
Lixia Zhang, UCLA
Mani B. Srivastava, UCLA
Mario Gerla, UCLA
Mike Fitz, UCLA
Narayan Mandayam, Rutgers
Nitin Vaidya, UIUC
Parviz Kermani, IBM
Ram Ramjee, Bell Labs
Rene Cruz, UCSD
Roy Yates, Rutgers
Sajal K. Das, UT Arlington
Scott Jordan, UC Irvine
Songwu Lu, UCLA
Thomas La Porta, Penn. State
Gary J. Minden, KU

Additional Participants:
Anant Sahai, UC Berkeley
Darleen Fisher, NSF
David N. Tse, UC Berkeley
H. Vincent Poor, Princeton
Joseph B. Evans, KU
Mari Maeda, NSF
Patrick White, Steven Institute of Technology
Rory J. Petty, KU
Timothy Shepard
Venkata Padmanabhan, Microsoft

Appendix C: Meeting Agenda

NSF Workshop – July 29-30 Chicago, Illinois

AGENDA

July 29

8:00 AM - 8:30 AM	Welcome and Workshop Plan Joe Evans, Mari Maeda, and Tom La Porta
8:30 AM - 10:30 AM	Agile Radios and Spectrum Policy Chair: Joe Evans Speakers: Tim Shepard Babak Daneshrad, UCLA
10:30 AM - 11:00 AM	Break
11:00 AM - 12:30 PM	Wireless Architectures: Ad Hoc, Cellular, and Hybrid Chair: Tom La Porta Speakers: Ram Ramjee, Bell Labs Nitin Vaiyda, UIUC
12:30 PM - 1:30 PM	Lunch
1:30 PM - 3:00 PM	Management of Networks of Agile Radios Chair: Mario Gerla Speakers: James Sterbenz, BBN Mike Fitz, UCLA
3:00 PM - 3:30 PM	Break
3:30 PM - 5:00 PM	Pervasive wireless systems Chair: Mario Gerla Speakers: Mani Srivastava, UCLA Greg Pottie, UCLA Parviz Kermani, IBM
6:00 PM -	Dinner

July 30

8:00 AM - 9:30 AM	Wireless Systems Chair: Tom La Porta Speakers: Henning Schulzrinne, Columbia Univ. Scott Jordon, UC Irvine
9:30 AM - 11:00 AM	Wireless Security Chair: Tom La Porta Speakers: Brian Noble, Univ. Michigan Songwu Lu, UCLA
11:00 AM - 11:30 AM	Break
11:30 AM - 12:30 PM	Writing Breakout
12:30 PM - 1:00 PM	Wrap-up