Mobile and Wireless Internet Services: Putting the Pieces Together

Joseph P. Macker
macker@itd.nrl.navy.mil
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Vincent D. Park
park@flarion.com
Flarion Technologies
Bedminster One
135 Route 202/206 South
Bedminster, NJ 07921

M. Scott Corson
corson@flarion.com
Flarion Technologies
Bedminster One
135 Route 202/206 South
Bedminster, NJ 07921

Abstract

There are a variety of emerging technologies and protocol enhancements designed to extend Internet services to mobile users, including operation over more dynamic, heterogeneous wireless interconnections. Many different approaches and protocols have been proposed and there are even multiple standardization efforts within the Internet Engineering Task Force that address portions of the overall goal. This article highlights some of the emerging technology efforts and provides insight into how some of these pieces may fit together to realize seamless Internet services for users on the move or in application spaces with little to no preexisting communications infrastructure.

Introduction

Computer and communication networks such as the Internet are multi-layered, complex systems that rely on many different protocols and associated algorithms for seamless, reliable operation. One of the most fundamental aspects of such networks is how protocols dynamically find paths through the network for forwarding information between specific end systems. This routing system must autonomously adapt to changes and failures in the network infrastructure and scale to support many millions of end systems. While the routing system used in the Internet has thus far evolved to meet these expectations, the proliferation of inexpensive wireless technology, portable computing, and the information-hungry nature of our increasingly mobile society pose new challenges.

In standard practice Internet Protocol (IP) internetworks (e.g., the Internet), the address of an end system (or host) is typically related to its location in the internetwork. End systems connected to a common communication media, such as a local area network (LAN) segment, are usually assigned individual addresses from within a contiguous block of address space that can be represented as a single common network prefix address. IP addresses may be further aggregated and summarized at routing domain or autonomous system boundaries. One of the keys to the scalability of IP internetworking has been the ability to efficiently exchange and maintain routing information based on aggregated addresses. However, the nature and use of this address aggregation often necessitates additional mechanisms to support the mobility of an end system across address aggregation boundaries.

Another present characteristic of the Internet is that it possesses a relatively quasi-static infrastructure that comprises end systems, routers and switches interconnected by largely hardwired links. Extending IP internetworking for seamless operation over wireless communication technologies challenges present performance requirements and assumptions of network protocols and applications, especially if wireless technologies evolve to become a significant part of the infrastructure, Figure 1. With the advent of inexpensive, heterogeneous wireless technology options, this engineering challenge will likely become more advent in the coming years with a myriad of unexpected application areas emerging with time. As Jared Diamond states in his Pulitzer Prize winning treatise, *Guns, Germs, and Steel: The Fates of Human Societies*, “invention is often the mother of necessity, rather than vice versa.” One can view the present Internet as a case in point, for while the pioneers of the Internet Protocol suite recognized the potential power of the technology being developed, they did not envision the detailed societal applications and adaptations. As we gaze into our crystal ball, the authors also feel unforeseen widespread impacts may be the historical result of highly adaptive wireless networking technology supporting an increased degree of self-organization and heterogeneity.

If effective performance is desired for a new generation of increasingly untethered wireless networks and devices, enhanced network protocols and applications must meet the challenge of unique behavioral dynamics and resource

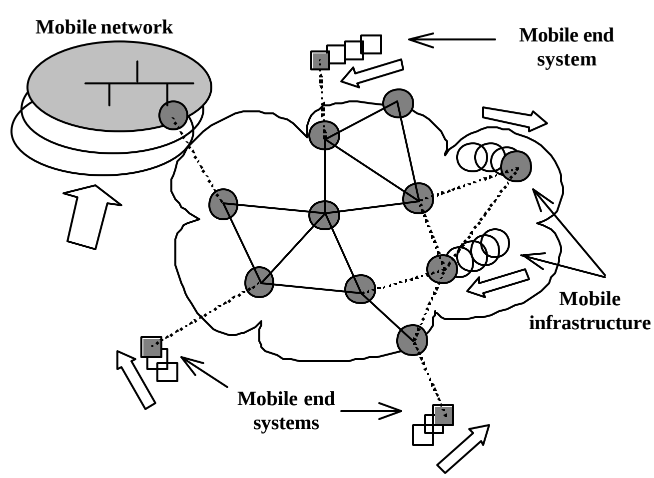

Figure 1: Mobility in IP Internetworks

This work was supported in part by the Office of Naval Research, the National Science Foundation, and the Army Research Laboratory.

constraints characteristic of wireless networks. While this topic has attracted renewed research interest in recent years, the focus has often been limited to a few isolated building blocks—e.g., developing improved routing protocols for mobile ad hoc networks or optimizing the Transport Control Protocol (TCP) for use over satellite links. Synergistic, interoperable solutions that provide a more unified approach to networking in a heterogeneous wireless infrastructure and offer a near equivalent capability to that currently available in hardwired internetworks are far from complete.

Heterogeneous Internetwork Routing

There are numerous operational factors that often distinguish mobile, wireless networks from fixed networks. Some of these include the following:

- nominally lower capacity network links (range, power, spectrum, and antenna tradeoffs)
- semi-broadcast nature of some wireless multiple access media (hidden terminal issues)
- increased likelihood of channel interference and congestion detection problems (e.g., due to bandwidth constraints, frequency restrictions or channel access techniques)
- more frequent topological changes (e.g., due to node mobility, channel propagation effects, resource failures, power control, or antenna dynamics)
- higher loss rates (e.g., due to interference, congestion or network dynamics)
- potentially higher delays and jitter (e.g., due to lower transmission rates, link layer retransmissions, use of long propagation delay links, or dynamics)
- lower physical security of media (e.g., due to lack of physical control over media)
- potential limited energy considerations (e.g., conservation of battery life)

In the past, mobile wireless networks were often looked upon as homogeneous radio frequency (RF) media problems. With time and the proliferation of numerous proprietary radio networks, the need for heterogeneous interoperability across networks is becoming a more prevalent interest area. To take a step back, one notices that the support of heterogeneity is one of the great successes demonstrated by IP technology. In the near future, computing and network routing devices may typically have multiple wireless media interfaces (e.g., Ultra-wideband, Bluetooth, 802.11, cellular). This proliferation of ubiquitous wireless devices is expected to continue with many technologies to choose from over time.

IP routing layer technology provides multi-hop relaying and dynamic internetwork connection support. In this broad sense, IP technology has and will continue to support both wired and wireless infrastructures. When facing increasingly complex or mobile wireless networks, routing performance with increasing temporal and topological dynamics is a key issue. We also wish to emphasize that increased dynamics may not always result from mobility. Dynamics due to other system effects may be quite evident in wireless networks even when the nodes are static or quasi-static. Regardless, the ability to *adapt to change* with a high degree of robustness and efficiency is assumed as a fundamental requirement. With increasing mobility, improved network device *self configuration* is also an important technology ingredient.

In homogenous wireless subnetworks (i.e., *same* underlying lower layer technology in use), a routing function at the lower layer may provide some limited multi-hop relaying and other services within that technology, and present this to the IP layer as a virtual single hop in the internetwork. While subnetwork layer mobile routing can support limited mobile networking, higher layer *mobile adept* routing for internetwork interoperability is typically still required. Or it may also be the exclusive means for multi-hop connectivity within a network, especially when the network is heterogeneous.

In heterogeneous wireless networks, we envision the IP internetwork layer interconnecting potentially different and similar wireless media segments and supporting a variety of routing protocols to provide a media independent connectionless relaying function. Using the IP layer may also be a cost benefit, as the network stack is often integrated within existing devices and operating systems or otherwise cost-effectively available in several varieties. Fundamentally, the IP protocol suite provides dynamic internetwork interoperability already; however, present off-the-shelf IP routing protocols may not be well suited for envisioned mobile wireless networking architectures.

As traditional IP routing protocols were not originally designed for the expected resource constraints and behavioral dynamics of the mobile wireless networking environment, there has subsequently been considerable renewed research interest in the development of alternative or enhanced routing solutions. The Internet Engineering Task Force (IETF), the primary standards body for Internet-based protocols and technology, has established a technical working group (WG) on Mobile Ad Hoc Networks (manet) to focus on developing and evolving IP mobile routing protocol specifications for inclusion as future IP technology standards. Note that the IETF is essentially an open standards organization; thus, much of the background research and development of the various protocols under consideration has been supported by other research organizations and projects. We will now take a step back and framework some IP routing issues before discussing some of the novel work that is progressing for mobile networks.

Broad Architectural Routing Considerations

A rough two level hierarchy is commonly used to describe IP routing in the Internet. In the case of unicast routing, exterior gateway protocols are used for routing between autonomous systems in the Internet backbone, while interior gateway protocols are used for routing within an autonomous system. The demands and requirements of these two routing functions differ significantly; thus, the preferred protocols for interior and exterior routing are typically not the same. Typically, an autonomous system comprises a set of networks and routers controlled by a single administrative authority. In addition to defining autonomous systems and routing domains based on administrative authority, architectural boundaries between portions of an internetwork with significantly differing network environments or capabilities should also be considered. When considering mobile architectures, one of these possible boundary conditions could be to determine the specific areas of significant mobility or dynamics vs. the more static areas. Defining multiple routing domains within an autonomous system allows for the use of different protocols and services within different regions of an internetwork, even if they are under a common administrative authority.

Traditional IP Routing Approaches

The Exterior Gateway Protocol (EGP) [1] and the Border Gateway Protocol (BGP) [2] are examples of exterior gateway protocols, while the Routing Information Protocol (RIP) [3] and the Open Shortest Path First (OSPF) protocol [4] are examples of interior gateway protocols. These commercially available protocols are designed primarily for operation in relatively static hardwired networks. Since bandwidth is relatively abundant in such networks, the protocol design goals primarily involve issues such as minimization of router state (to permit scalability), processing load, and protocol convergence time in reacting to topological changes when they occur (to ensure good packet routing performance). In other words, minimizing storage, computation, and time complexity are primary design goals of commercially available Internet routing protocols intended for operation over high-speed interfaces.

Emerging IP Mobile Routing Alternatives

There are many emerging interior gateway protocols that are specifically designed for mobile wireless networks [5]. While some of these mobile wireless routing approaches proactively maintain routes between all source-destination pairs, many have adopted a more reactive, on-demand design. Classification of a protocol into one of these categories is not always straightforward, as some of the protocols have elements of both or can be viewed as hybrid proactive and reactive designs. Each protocol family has its advantages and disadvantages and the appropriateness of the design type is affected by a spectrum of architecture and performance issues [6].

In pure on-demand routing approaches, routes are reactively established to a given destination when needed (i.e., traffic driven). This design choice is based on the notion that in a dynamic topology it may not be necessary (or desirable) to maintain routes between all source-destination pairs at all times. The assumption is that the overhead expended to establish and/or maintain a route between a given source-destination pair will be wasted if the source does not require the route prior to its invalidation due to topological changes. Note that this assumption may not hold true in all architectures, but it may be suitable for many envisioned wireless network applications. The validity of this design decision is dependent in part on the traffic distribution and the topology dynamics in the network. Conceptually, it would seem most advantageous when traffic patterns are relatively sparse and topology dynamics are relatively high. While these approaches have the potential to reduce communication overhead, this is achieved at the expense of some increased initial route acquisition latency.

The proactive approaches (also referred to as table-driven) are more similar in design to traditional IP routing protocols; thus, they are more likely to retain the behavior features of routing protocols presently used in practice. In general, proactive routing approaches may be better suited than on-demand approaches when routes to a large percentage of the possible destinations are typically required. When network traffic patterns include a large percentage of possible source-destination pairs, the advantage of building routes to specific destination on-demand is reduced. In short the overhead of the on-demand route request process can be saved in situations where the demand can be assumed.

Many protocols being specified within the IETF manet WG include elements of both categories mentioned above. For example, some of the predominantly on-demand approaches may cache known routes for future use or proactively maintain routes once they are initially established. In some approaches it is also possible to simultaneously support both on-demand routing for some destinations and proactive routing for other destinations.

Potential Application of IP Routing Solutions

We have presented a high level view of available and emerging IP routing protocols. A complete analysis of each protocol across a set of potential architecture scenarios is complex due to the large design space and the performance tradeoffs one might consider. As technical guidance to future system analysis in this area, there are a number of network parameters that should be considered as primary performance drivers [6]. Some of the more important anticipated system characteristics include the following:

- required network scalability (e.g., number of routing nodes, number of end systems within an area)
- typical link bandwidths and transmission rates
- percentage of application types and anticipated traffic models
- degree and class of mobility and/or wireless link dynamics expected
- network density and degree of heterogeneity expected

Network scalability is often a key factor to consider when applying candidate routing protocols to a particular scenario. In general, there may be many different system or resources constraints that limit the scalability of a protocol—e.g., available bandwidth or data storage capacity. Since routing protocol overhead typically grows as some function of the number of nodes, it is often a limiting factor given bandwidth constraints. Similarly, protocol state also typically grows with the number of nodes, and thus can be a limiting factor given memory or data storage constraints. The convergence time of a given routing protocol may also be affected by scalability with respect to the number of nodes. The potential impact on routing performance (e.g., increased delay or jitter) may be increasingly important given the proliferation of streaming multimedia applications.

Anticipated link capacities and the number of nodes sharing a common channel within a wireless network are also important considerations prior to developing any particular architecture design. Commercially available Internet protocols often assume or are pre-configured for more optimal operation in higher link rate conditions (e.g., T1 and greater). One would like to avoid the necessity to tune protocols per configured link in order to provide a design compromise between protocol reaction time and reduced overhead for application in limited bandwidth wireless networks. The lower link capacities and the increased dynamics characteristic of wireless networks beg for efficient routing approaches featuring reduced routing overhead while maintaining timely reaction to topology dynamics. This tradeoff space is a performance feature that forms a core design goal of most IETF manet protocols under consideration.

The authors' feel that the strong interaction with various traffic models presented to the routing protocols by the network applications and user/service distribution patterns is often overlooked in routing analysis of on-demand protocols. There are several areas of importance to consider. First, the number of sources and destinations and their relative distribution throughout the network is of importance. Second, the volume of traffic injected by sources and the long-lived nature vs. short-lived nature of interactions is important. Each of these elements directly affects the potential performance and overhead efficiency of some routing protocols. In addition to performance under average load conditions, the ability to maintain effective performance under network traffic surge conditions (e.g., increased numbers of source and receivers) should also be considered. In short, understanding application behavior and traffic distribution models should be included as a critical part of interpreting any protocol analysis.

Finally, the degree and type of mobility or topology dynamics anticipated is another core performance analysis factor. Many of the emerging IP routing approaches previously discussed are designed to handle dynamics more efficiently. There may, however, often be dynamics in any realistic wireless network whether there is considerable node mobility or not. Some manet protocols are designed to be able to scale with the degree of dynamics better than others. Other candidate manet protocols may work less well in high dynamics, but may target more optimal route discovery given moderate dynamics. These features need further study for a better understanding, but there are likely performance trends one can predict by understanding the nature and assumptions of individual algorithms.

As we have briefly discussed, the considerations for performance analysis of mobile routing protocols must take into account assumptions about various architectural design goals and features. In some cases, multiple protocols may be chosen for use in different parts of a network for different reasons that may include cost and complexity tradeoffs as well. This is perfectly acceptable and is not unlike the use of multiple routing domains in present day hardwired internetworks.

End System Mobility

Now we consider and discuss the challenging design issues of end system mobility as separate from or in concert with underlying infrastructure dynamics. In our terminology, end system mobility may refer to a scenario where a laptop is disconnected from a wired network at the home office and temporarily connected to a different network (e.g., during a business trip), or to a scenario where mobile wireless computing devices are dynamically associating with different points of attachment within a fixed or mobile infrastructure. In the most general case, an end system may be dynamically associated via a variety of different wired and wireless technologies to various points of attachment that are either part of a fixed or mobile infrastructure.

As stated in an earlier section, infrastructure routing and end system mobility management are not necessarily orthogonal processes. The problem of supporting end user mobility may be largely solved via network layer routing techniques or higher layer location management techniques. Even if a higher layer location management solution is utilized it may be based on the existing routing technology or integrated with a particular routing approach. Performance assessment and comparison of techniques must consider the approach to both routing and end system mobility management, since protocol complexity, adaptability, efficiency, and communication overhead can be shifted between the two mechanisms based on the overall system design. In the remainder of this section we briefly discuss some of the existing and emerging technologies that may provide support for end system mobility and how these may be orthogonal or related to improved routing algorithms.

Host Auto-configuration

The Dynamic Host Configuration Protocol (DHCP) [7] allows for a host to dynamically acquire a temporary (or permanent) IP address and other network configuration information from a server upon startup or connection to a local network. The information provided via DHCP is sufficient to allow the host to communicate over a connected internetwork, and thus provides limited support for end system mobility. That is, each time a host is moved and connected to a new network the host may acquire a new IP address and other configuration information via DHCP, and subsequently send and receive IP datagrams using the new IP address.

Used alone, DHCP has two significant limitations in terms of supporting end system mobility. First, other hosts in the internetwork cannot locate or send datagrams to the mobile host without first learning the mobile host's new IP address via some means. Second, connection-oriented data flows such as those using TCP will be disrupted each time a mobile host changes its IP address. Dynamic Domain Name System (DDNS) updates [8, 9] can be used to address the first of these issues; however, the second issue is a fundamental limitation of this architectural approach. When combining DHCP with DDNS, each time a mobile host changes its IP address via DHCP, a DNS update is sent. Thus, the mobile host's name in the DNS remains constant while its corresponding IP address changes with mobility. Other hosts in the internetwork may locate and acquire a mobile host's current IP address via a DNS request. Tracking frequent mobility changes of end systems via DNS updates has some potentially negative impacts on the DNS caching strategy and thus the efficiency of resolving DNS requests. The practicality of scaling this approach to support large numbers of mobile users may also be limited. There are also some security issues associated with endpoint authentication and key management that must be uniquely addressed when hosts are allowed to frequently change IP addresses.

A Host Routing Approach

Although IP-based datagram forwarding is primarily based on network routes, host-specific routes are possible. The use of host-specific routes provides another approach to supporting host mobility. In a host routing architectural approach, a mobile host maintains a constant IP address as it changes its point of attachment to the internetwork and dynamically updated host-specific routes are used to deliver datagrams to the host at its current location. Many of the routing approaches under consideration within the IETF manet working group could be directly applied to support host routing.

Another example of this approach is the Local-Area Mobility (LAM) software available in router products from Cisco Systems. A router configured for LAM monitors traffic on its network interfaces. When locally originated traffic from a host with an IP address that does not match the address and mask (or prefix) configured on the routers interface is detected, the router adds an ARP entry for the mobile host and installs a host route that points to the interface. The router may also redistribute the host route into a dynamic routing protocol being used in the internetwork, thus allowing the host route to propagate throughout the routing domain.

An advantage of host routing approaches over the DHCP/DDNS architecture is that transparent upper layer connectivity to the mobile host can be maintained under dynamic conditions. That is, provided that the host-specific routing is sufficiently adaptive and converges relatively quickly, connection-oriented data flows like TCP need not be terminated when a mobile host changes its point of attachment. There are practical limits regarding the extent to which host routes can or should be propagated (e.g., throughout a routing domain, autonomous system or globally). Despite the potential limitations it may still be a viable solution for certain architectures and can serve as a piece of a more comprehensive mobility solution.

IP Mobility Support—Encapsulation and Tunneling

The IP Mobility Support specification [10] defines a framework and protocols that provide tunnel-based routing to mobile hosts (i.e., Mobile IP). This specification is a product of the IETF WG on IP Routing for Wireless/Mobile Hosts (mobileip). In the Mobile IP framework a mobile host is given a long-term IP address on a home network (i.e., its "home address"). When away from its home network, a "care-of address" that reflects the current point of attachment is associated with the mobile host. Routing to a mobile host that is away from its home network is supported via tunneling by mobility agents (i.e., a "home agent" on the mobile host's home network and possibly a "foreign agent" at its current point of attachment). As the mobile host changes its point of attachment, it registers its current care-of address with its home agent. Datagrams sent to mobile host's home address are intercepted by the home agent and tunneled to the care-of address via IP encapsulation. The end point of the tunnel (where decapsulation occurs) may be either at a foreign agent or at the mobile host. IP datagrams sent by the mobile host typically use standard destination-based IP routing techniques.

As with host routing, Mobile IP supports transparent connectivity to the mobile host. That is provided that the mechanisms adapt and converge relatively quickly, connection-oriented data flows like TCP need not be terminated when a mobile host changes its point of attachment. A potential advantage of the Mobile IP approach is that it is transparent to the underlying routing approach; thus, it does not increase routing table size or in any other way impact the scalability of the underlying routing approach. The scalability of the Mobile IP architecture is primarily limited by the home and foreign agent resource constraints.

There are several other issues regarding the Mobile IP approach that merit mention and consideration. First, the basic Mobile IP approach (for IPv4) results in what is referred to a triangle routing, Figure 2. Datagrams sent from a corresponding host to the mobile host are initially sent to the home network, and then intercepted by the home agent and forwarded to the mobile host (in some cases this path may be very circuitous), while datagrams sent in the reverse direction are sent directly to the corresponding host. Note that mechanisms for route optimization (Figure 3), which would allow for direct routing between capable corresponding and mobile hosts, are built in as a fundamental part of mobility support in IPv6. There is also an Internet Draft that proposes optional extensions that would allow for route optimization in

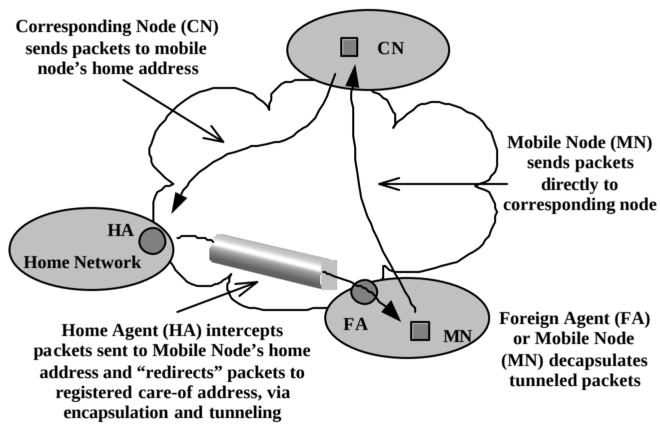

Figure 2: Mobile IP Triangle Routing Example

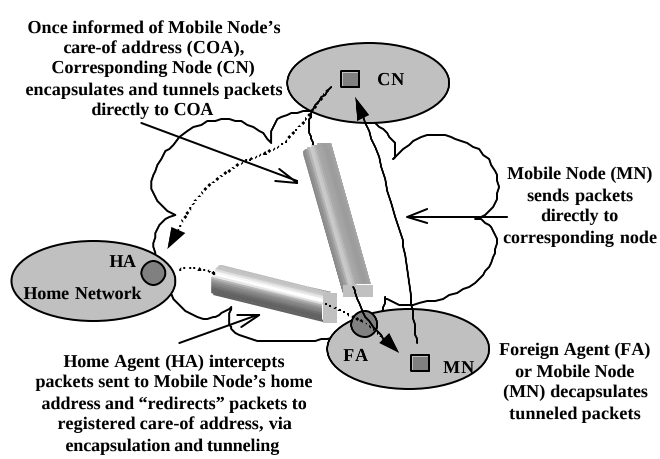

Figure 3: Mobile IP Route Optimization Example

IPv4. Second, the architectural dependence on a home agent limits the robustness of the overall system design. A mobile host must be able to communicate with its home agent in order to register its care-of address each time it changes its point of attachment. Finally, for some applications the additional per packet overhead associated with IP encapsulation may also be of concern.

Many of the mobile-enhanced routing approaches mentioned in this article can co-exist and support the overlying operation of Mobile IP techniques when and where sensible. The choice of mobile architecture protocols can be quite rich and should not be limited to protocol A vs. B, but rather what correct combination of overall mobile system components are sensible given the network operating conditions, overall architecture, and application requirements.

Emerging End SystemMobility Alternatives

There have been many different proposals and research project developments that have extended or modified the basic design of mobile IP to address particular limitations or improve it for certain architectures. We will highlight a couple of these ideas to illustrate the nature of the work. Some approaches have sought to create a more distributed variant of Mobile IP. In one approach, all routers within a the routing domain include both home and foreign agent functionality and home to care-of address binding information is disseminated among all such routers. This approach eliminates the dependence of a mobile host on a single home agent; thus, increasing the robustness of the network design and allowing for subsystem failure. It also eliminates the use of triangular routing within the routing domain. However, the advantages described above are essentially achieved at the expense of increased overhead in flooding the address binding information throughout the routing domain, which significantly limits scalability.

Another proposed and demonstrated modification to Mobile IP entails extending the concept of tunneling to a mobile host to the case of tunneling to a mobile network. This provides a framework that could be used to support transparent connectivity to a subnetwork or portion of an internetwork that is aboard a mobile platform (e.g., an automobile, aircraft or ship) and is dynamically changing its point of attachment. Performance tradeoffs of this approach should be further evaluated and compared to solutions based on the use of dynamic internetwork routing protocols.

Finally, within the IETF, private industry, and the research community there is considerable recent interest in supporting IP networking to handheld cellular end systems. The general consensus is that Mobile IP as currently specified is not sufficient to completely support this application. Within the IETF mobileip WG there are Internet Drafts that propose extensions to improve handoff performance, localize registration signaling and enhance support for third generation cdma2000 networks. Also, a new IETF WG (i.e., seamoby) was recently formed to investigate and possibly develop new protocols in support of seamless mobility. Currently the charter of the IETF seamoby WG includes investigation of routing solutions to improve intradomain handoffs, requirements for transferring context between access routers during handoffs, and the use of IP layer paging. There are several IETF Internet Drafts that propose mobility solutions based on an architecture where Mobile IP is used to support interdomain mobility and host routing is used to support intradomain mobility. Routing solutions originally developed in the IETF manet WG may prove to be scalable adaptive protocols for supporting the intradomain host routing within the fixed access network. This work is likely to continue to evolve, but is indicative of the potential widespread applicability of piecing together mobile-enhanced protocols and technology solutions in more indirect application areas other than initially envisioned.

Beyond a Basic Packet Delivery Service

Initial research on extending Internet services in mobile wireless networks has primarily focused on network-layer issues—such as the design and development of routing protocols, tailored for operation in a highly-dynamic and bandwidth-constrained networking environment. While not discussed herein there has also been considerable research on multicast routing technology for mobile environments. Although the research is still evolving, the rationale for adopting and using multicast technology is quite strong in wireless networks, so the stimulation for further development and exploration is likely to continue. The typical broadcast or semi-broadcast nature of most ground mobile packet radio and satellite downlinks makes multicast dissemination of data intended for group communications a natural engineering choice.

While connectionless routing of packets is an important first step for providing similar services to those available in hardwired networks, many applications require additional functionality, such as end-to-end reliability or flow control. These end-to-end services are typically partially or fully provided by transport-layer protocols. Providing robust, functional transport-layer services in mobile wireless networks remains a largely unexplored research area.

As with existing network-layer protocols, traditional transport-layer protocols designed for use in hardwired networks may not be well suited for scaled use in mobile wireless networks without modifications or additional system component or protocol enhancements. The underlying assumptions used in the protocol designs may not be valid due to differences in the characteristics of the networking environment and the services provided by lower-layer protocols. Ultimately, the transport layer performance would be less of an issue if the lower-layer mobile, wireless protocols could provide a service that closely emulates a fixed, wired network service. Such transparent design is not easily achieved and it is often better to reach some compromise design between lower and upper layer protocols to allow more adaptive performance and to support a larger class of applications efficiently. In short, further developments in mobile-enhanced transport and application layers are essential for providing a complete set of Internet services comparable to those available in hardwired networks.

While we do not discuss the details of upper layer mobile applications and services within this paper, it is the authors' opinion that this is an important design and consideration area in an overall mobile architecture. Resilient and adaptive applications that can continue to perform effectively under degraded conditions can significantly enhance network operations from a user's perspective. Such applications advancements can also ease the design pressure in complex engineering areas such as quality of service (QoS) and mobile routing at the network layer.

Summary

We have presented a brushstroke overview of mobility-enhanced internetworking protocols and their relative applicability to mobile and dynamic wireless communication architectures. Our focus is on emerging network technologies and architectural components that provide heterogeneous support and interoperability with existing and planned IP-based infrastructures and applications. The intent of this paper is not to provide detailed answers to a specific system design, but to provide a discussion of high level issues and challenges requiring further examination of architectural and protocol tradeoffs. We wish to reiterate that a robust, scalable, mobile network architecture is likely made up of many combinations of protocol components, a broad set of which we have discussed. We conclude this paper with a number of recommended questions to be addressed during a design and protocol selection effort for a specific wireless, mobile architecture.

- How mobile or dynamic is the network likely to be (how frequent are the topological dynamics, the wireless link perturbations)?
- Is there mostly edge system mobility or is the infrastructure itself dynamic? Is it likely some combination?
- How heterogeneous is the wireless internetwork (number of potential wireless media and link layer types)?
- Will the system support high bandwidth or mostly constrained links, or some combination?
- What is the general form and distribution of network applications to be supported?
    - What is the expected pattern and load of network traffic driven by the applications?
    - What QoS requirements are driven by the applications?
    - Is there a need to keep the design maximally flexible in this regard?
- Is the architectural design constrained or influenced by administrative, existing infrastructure, or system capability considerations?
    - Is there a natural or desired backbone?
    - Are there natural or desired routing domain boundaries?
- Is there a network node auto-configuration requirement?
- What are the security requirements and how is security intended to be managed?
- Are the systems power-constrained (e.g., portable batteries)?
- Are multicast services required/desired?

Armed with the full or partial answers to these questions, a high-level engineering tradeoff analysis can begin to set further direction and detailed focus in specific mobile system design efforts.

References

[1] D. Mills, "Exterior Gateway Protocol," *IETF STD 0018*, April 1984.

[2] Y. Rekhter and T. Li, "A Border Gateway Protocol 4 (BGP-4)," *IETF RFC 1771*, March 1995.

[3] G Malkin, "RIP Version 2," *IETF STD 0056*, November 1998.

[4] J. Moy, "OSPF Version 2," *IETF STD 0054*, April 1998.

[5] C. Perkins, *Ad Hoc Networking*, Addison-Wesley 2001.

[6] S. Corson and J. Macker, "Mobile Ad Hoc Networking (MANET)," *IETF RFC 2501*, January 1999.

[7] R Droms, "Dynamic Host Configuration Protocol," *IETF RFC 2131*, March 1997.

[8] P.V. Mockapetris, "Domain names – implementation and specification," *IETF STD 0013*, November 1987.

[9] P. Vixie, Ed., S. Thomson, Y. Rekhter and J. Bound, "Dynamic Updates in the Domain Name System (DNS UPDATE)," *IETF RFC 2136*, April 1997.

[10] C. Perkins, "IP Mobility Support," *IETF RFC 2002*, October 1996.

Biographies

Joseph P. Macker is a senior research scientist at the U.S. Naval Research Laboratory (NRL). Much of his work at NRL has focused on wireless networking and advanced protocol research. Mr. Macker's recent work and interests involve mobile networking technology, reliable multicast transport design, congestion control, Quality-of-Service (QoS) and guarantees in networks, collaborative multimedia applications, and network security. Several ongoing projects involve the research, development, and demonstration of emerging networking technologies and applications. He presently serves as co-chair of the IETF manet Working Group.

Vincent D. Park is a researcher in the advanced networking group at Flarion Technologies Inc. Previously, he worked for several years as a research scientist at the U.S. Naval Research Laboratory (NRL). The focus of Mr. Park's research is on the design, development and analysis of network protocols and architectures. Specific areas of research include: distributed algorithms; unicast, multicast, and anycast routing protocols; mobile wireless networking; modeling, simulation and analysis of network architectures and protocols; enhanced quality of service (QoS) support; and network traffic management. Mr. Park is also an active participant in the IETF manet and seamoby Working Groups.

M. Scott Corson is the director of advanced networking at Flarion Technologies Inc., where the company is building a truly IP-based cellular networking technology. He has principally worked on multiple access and routing technologies for mobile wireless networks, and presently serves as a co-chair of the IETF manet WG. He has been on the faculty of the University of Maryland at College Park since 1995, and a consulting network architect for British

Telecomm (BT) Labs working on the design of an IP-based, fixed/cellular-converged network architecture. He has a Ph.D. in Electrical Engineering from the University of Maryland.

(Sidebar)

Role of Manet Technology

What role does manet technology have to play in the coming wireless Internet? That question has been asked and will continue to be asked as new forms of wireless technology emerge and enable ad hoc networking in various forms.

The "manet pessimist" view states that there are insurmountable barriers and system constraints that will significantly impede the practical deployment of manet technology. The net effect being that manet-based communications should be viewed as an option of last resort and will likely never be used to a large extent.

The "manet optimist", on the other hand, sees the day when technological advances permit the deployment of large-scale manets spanning metropolitan areas permitting low cost deployment of wireless, wide area Internet infrastructures with little or no reliance on existing, fixed infrastructures. The "optimist" sees a very powerful technology concept and envisions a myriad of other widespread manet applications including; robotics, sensor networks, heterogeneous home networks, etc.

The "manet pragmatist" falls somewhere between these two viewpoints. The view here is that manet technology will be an enabling technology and judged on its merits once mature; filling in the seams and extending the edges where other forms of wireless communication are either relatively costly or inflexible, delivering connectivity in a complimentary fashion with other technologies to help fulfill Mark Weiser's vision of ubiquitous computing. Here, manet becomes mainstream, but indirectly, functioning in concert with existing networks, and in potentially non-obvious ways to the user.

In summary, it is apparent from the wide spread interest in manet technology that a myriad of additional networking applications of the technology will be found. Manet technology appeals to the entrepreneur, the individualist, the philanthropist, and the explorer of the next Internet generation. Individuals, organizations, and businesses will field wireless networks cheaply in their own fashion, for their own needs or for the needs of many.