MOBILE INTERNET PROTOCOL PERFORMANCE AND ENHANCEMENTS OVER ACTS

David B. Johnson
Computer Science Department
Carnegie Mellon University
Pittsburgh PA 15213-3891
dbj@cs.cmu.edu

Bernard J. Bennington
Information Networking Institute
Carnegie Mellon University
Pittsburgh PA 15213-3890
ini-director@andrew.cmu.edu

1. PROJECT BACKGROUND

In 1997, Caterpillar and Carnegie Mellon University started a three-year research program to develop a global mobile wireless communication system, one that will support all corporate, operations, dealerships, and end customers' mobile communications reliably. The system is expected to consolidate needs for all Caterpillar business units, from large mining equipment to small construction equipment, from agricultural systems to field support systems. Because of the remote location of many of Caterpillar's customers, for example large mines or quarries, the approach was to assume no common availability of public networks. Caterpillar also wanted the end architecture to be capable of incorporating many competitive IT and radio manufacturer's products, and to have open interfaces for easy usage.

The mining and construction equipment industry, like many other industries, is seeking to differentiate their hard products from those of their competition by adding value with services and applications. The wireless communication system was seen as part of this strategy. The system will support many of these initiatives, for example, it is necessary to implement improved service for machines in the field, by providing remote access to machines and people from anywhere to anywhere. Other immediate uses include, providing direct access for field service technicians to an up-to-date database of service information, hence reducing the problems of using current CDROM technology for frequently changing service information. It is also expected to allow dealers and technicians to do remote diagnosis of a machine's condition using existing and future onboard sensors and data reduction systems. Finally, Caterpillar wants to be able to field intelligent real-time applications such as information support for earth-moving operations, incorporating real time access to and updating of topographical maps, and also wants to be able to field intelligent cooperating robotic machines.

In early visits by the Carnegie Mellon team to Caterpillar's customers at major mines and quarries, and in spending time with both urban and remote service technicians, additional requirements were seen. The mine and quarry owners need a single site system serving all needs, not one for each application (including applications already ported to sites, such as voice radio communication systems and dispatch systems). End customers also stressed that the system has to be easy to install, operate, and maintain, as many sites are remote and they have difficulty getting qualified IT and wireless system staff. Caterpillar's dealerships had other requirements. They indicated the essential need for a simple technician interface and a rugged laptop they could simply turn on, not one where the network and the modem type has to be selected, where the technician has to worry about where they are. Dealership staff also indicated additional requirements in that they want access to their own service database, including manuals, work orders, history, and other technician's experience, and when performing service on a faulty machine, they want this access directly at that machine.

This is a demanding set of requirements, and it raises some difficult issues. The system has to handle a wide variety of communications, from data and file transfers, to real-time data including voice telephony, video, radar, and laser images. In addition, the system needs to handle priorities for different messages, such as for emergency stop messages for a robotic vehicle. The system also has to be able to do this anywhere within a country, accommodating the difference, for example, between downtown Chicago and the remotest areas of the Nevada desert, and has to be able to do it anywhere in the world, bearing in mind that spectrum management is different in different countries, resulting in different technologies being available in different locations. There is also a specific significant problem of being able to communicate with highly mobile machines that move frequently from site to site such as in rental fleets.

There are also a number of miscellaneous issues to be considered, for example: How do you integrate the pieces into a single system as we do with wired computers? How do we deal with robustness in challenging environments? How do we deal with security? Finally, although the mining and construction industry is large it is not large enough for a solution to its unique requirements to emerge spontaneously. The nearest similar industry is that serving the battlefield.

The approach taken in the project considered three aspects: wireless technologies, protocol technologies, and architectural scenarios.

The approach to wireless technologies began with the questions: What is available? What is coming? Where do they cover? What bandwidths are served? How do you integrate them? There is a wide variety of wireless technologies, and each represents a design compromise of bandwidth, mobility, and coverage. All have different costs to build and costs to operate. Some are real and proven technologies, and some are speculative and their adoption will depend upon markets—markets that the construction and mining industry will not greatly influence. These wireless technologies include:

- High speed, low coverage, spread spectrum wireless LANs;
- Low speed, high coverage, licensed radio;
- Packet data services that ride the public voice networks, such as CDPD or GPRS, special packet networks like RAM, ARDIS, or Metricom, and future third generation offerings; and
- A variety of satellite services, from GEOs to LEOs, with various costs and bandwidths.

All of these technologies can play a role, depending on where the equipment or technician is, what the application is, and what economics dictates is reasonable to pay, but none of these technologies can meet all the requirements. Any major application will use several overlaid networks. Consequently, the Carnegie Mellon architecture is a hybrid of fixed and mobile nodes, licensed and unlicensed spectrum, accommodating a variety of wireless technologies, integrated, just like wired systems, with common interfaces.

Examination of protocol technologies also raised questions: How do you tie unlike products and services together? How do you support migration from one to the other? How do you create simple, easily configured and maintained networks? How do you create reliability in high demand areas? To make this work, we use an IP-based architecture, with the standard TCP and UDP protocols riding on top of basic IP. Because the nodes are moving, we use the standard Mobile IP protocol [3,5,7] for migration across technologies and networks. Two other important things have been developed in the project. First, to solve the need for simplicity in creating high-speed networks in mines and quarries, we use new multi-hop wireless ad hoc networking protocols [4,5] for a simple approach to high demand areas. We also use intelligent protocols that can use knowledge of the terrain and vehicle position to enhance reliability and performance.

To focus the research and the development of ideas, several reasonable scenarios were drawn for pieces of the overall architecture. These in essence asked: What technologies are probable for use and where can they be used? How will these unfold over time? How do they relate to populated areas vs. unpopulated areas? Five architectural snapshots were settled on as being representative. These were: populated and unpopulated areas today; at a future "step one" and at a later future "step two", where "today", "step one", and "step two" were defined by existing or proposed technologies, the strength of their markets, and how they have been deployed and used, and by speculated probability of coming technology deployment and capability. In essence "today" is what is available today even if only in selected locations, "step one" is defined by expected increased bandwidth of public network services and increased speed of spread spectrum devices. "Step two" was defined by increased use of broadband satellite technologies (such as Teledesic) and their applicability to fixed and mobile elements. Possible use scenarios were seen as being composed of:

- *Mines or quarries*: characterized by having an intense population of machines, particularly high-end machines, and hence an intensity of high speed, demanding applications and services. Such sites by their nature have justification for custom-built, dedicated networks. Sites typically have work areas with clusters of cooperating machines and these are connected together by site roads. A site also has static structures such as site offices or a crushing plant.
- *Dealerships*: Fixed locations serving multiple customers which can be urban or remotely located. The dealership is the source of roaming elements, namely technicians in service trucks, who may constantly move from populated to unpopulated areas.
- *Urban areas*: characterized by having fewer (often single) machines at locations and these machines typically are smaller, lower-end machines where unit cost is critical and applications and services are hence primitive. There is no justification for a custom-built network and, hence, applications have to use what is available.

Mines, quarries, dealerships, and urban areas are usually interconnected by public roads where, once again, there is no justification for a custom built network and applications have to use what is available.

From this, we determined that satellite communication could form two important pieces of the overall architecture, namely, as a fixed link within subcompoents of the overall system, providing connection between subnetworks from fixed nodes such as site offices, and also potentially as mobile nodes providing interconnection between subnetworks. In each case, this raises the common problem of effectiveness of the fundamental IP based protocols used in the overall architecture. This is particulatrly true of the use of Mobile IP, which was adopted as the basic method of management of mobility across subnetworks.

We saw the Ka-band technology used by NASA's Advanced Communication Technology Satellite (ACTS) as having the potential for both roles, static and mobile communications, because of its available bandwidth and its use with significantly smaller ground antennas. A combination of an ACTS link and an ad hoc wireless LAN network [4,5] embedded in vehicles within a mine or quarry could be used, so that such mobile vehicles could take advantage of ACTS's large coverage area and a wireless LAN's small propagation time and high bandwidth. With the Mobile IP protocol, the mobile vehicle could seamlessly move from one network to another.

If the Internet Protocol (IP) is used as the network protocol, the ACTS network could be configured as one IP network and each ad hoc network as another. A mobile node could then move from an ad hoc network to the ACTS link, as moved away from a site, so that it could maintain a network connection all the time. Standard IP makes it very difficult for Internet hosts to move from one network to another, since parameters such as the node's IP address, subnet mask, and default router generally need to be changed, making movement both time consuming and error prone. Furthermore, a mobile nodes must also inform its new address to all nodes that want to communicate with it. The Mobile IP protocol has been developed to solve this problem in the Internet.

We conducted experiments using ACTS to observe the performance of Mobile IP. We analyzed the cross-traffic effects of Mobile IP and TCP over ACTS and implemented techniques to improve the performance of Mobile IP over the ACTS link.

2. MOBILE IP

Mobile IP [3,5,7] is a standard developed by the Internet Engineering Task Force (IETF) to provide transparent mobility for Internet nodes. Mobile IP can extend the ACTS functionality by enabling mobile nodes to use the ACTS link seamlessly without the need to reconfigure their IP address. However, Mobile IP is generally tuned to perform well in terrestrial wireless networks such as wireless LANs, which have a limited range and a negligible propagation delay. ACTS and other GEO satellite systems pose a challenge to Mobile IP due to their high propagation delay. We have conducted experiments using ACTS and different mobile node movement scenarios to observe how Mobile IP performance is affected.

We have also analyzed the cross-effects between TCP versions optimized for satellite links and Mobile IP over the ACTS link [6]. TCP, the Transmission Control Protocol, is the most commonly used transport protocol in the Internet. It provides reliable delivery of all bytes sent, by retransmitting data until an acknowledgement is received from the destination node. ACTS's high propagation delay and error rate effect TCP performance because data and acknowledgement packets take longer to reach its destination, and these packets might get lost. With Mobile IP, however, TCP also has to manage the loss of packets caused by the handoff. We have implemented and evaluated techniques to improve the performance of Mobile IP over the ACTS link, especially the performance of TCP using Mobile IP during handoff. In particular, we have shown that smooth handoff with a single packet buffer considerably improves TCP performance during handoff without modifying TCP.

In traditional IP routing, if a node has moved to another network, packets destined for it will no longer be deliverable. For a node to be able to communicate on a new network, it must change its IP address, but this makes it impossible to maintain transport and higher-level connections when the node changes locations. Mobile IP is defined as a scalable solution to this routing problem that does not require a mobile node to change its IP address when it moves. Mobile IP enables a host to be identified by a single IP address even when the device physically moves its point of attachment from one network to another, allowing for the transparent forwarding of data packets to a single address.

2.1. Basic Mobile IP

The operation of the basic Mobile IP protocol is illustrated in Figure 1. Each node is assigned a permanent IP address in the same way as any other node, and this IP address is known as the mobile node's *home address*. The IP subnet indicated by this home address is the mobile node's *home subnet*. When a mobile node is attached to its home network, traditional IP routing will deliver packets to the mobile node using its home address. When the mobile node is away from home, a Mobile IP agent on its home network known as the *home agent* keeps track of the current location of the mobile node and forwards packets to it from its home network. Any network that the mobile node visits is referred to as a *foreign network*. A Mobile IP agent located on the foreign network known as a *foreign agent* may help the mobile node register with its home agent and may also help deliver forwarded packets to a mobile node. Any node with which a mobile node is communicating is known as a *correspondent node*, and may be mobile or stationary.

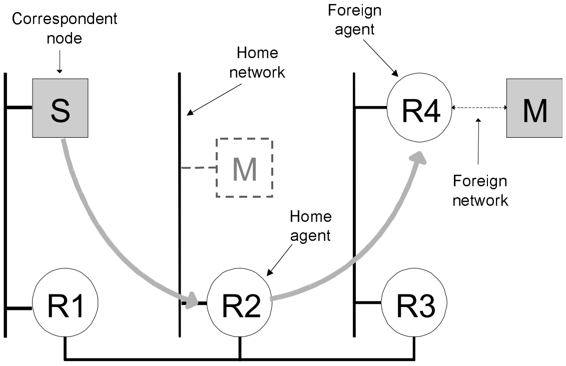

Figure 1: Basic Mobile IP

The mobile node's current location while away from home is known as its *care-of address*. The care-of address will often be the IP address of the foreign agent, and this type of care-of address is known as a *foreign agent care-of address*. Alternatively, the care-of address may be a *co-located care-of address*, which is a local address obtained by the mobile node for its own use in that foreign network; in this case, the mobile node operates without a foreign agent. In this paper, we consider only the case of a foreign agent care-of address.

Mobile IP defines an Agent Discovery mechanism to allow a mobile node to discover whether it is at home or away from home, and to discover a foreign agent in its current network with which it could register. Agent Discovery is based on an extension to the ICMP Router Discovery Protocol [2], with each home agent or foreign agent transmitting periodic *Agent Advertisement* messages, giving the agent's IP address. If the mobile node notices its own home agent's Advertisement, it knows it is at home; it then deregisters with its home agent and no longer uses a care-of address or any support from Mobile IP.

If the mobile node is away from home, it informs its home agent of its current care-of address using a mechanism called *registration*. If the mobile node is not using a foreign agent, it sends a Registration Request message directly to its home agent, which answers with a Registration Reply message; otherwise, it will send its Registration Request to its current foreign agent, which forwards the Request to the mobile node's home agent. Registrations expire after a specified period called the *registration lifetime*. The association of a home address with a care-of address and the remaining lifetime is called a *mobility binding*. When a correspondent node sends packets to a mobile node, the packets are routed normally, to the mobile node's home network. The home agent intercepts those packets and forwards the packets to the mobile node by *tunneling* them to the mobile node's care-of address. To tunnel each packet, the home agent encapsulates it in a new IP header, with the destination address in this outer IP header set to the mobile node's care-of address. If a foreign agent care-of address is being used, the foreign agent will receive the packet, decapsulate it, and deliver it locally to the mobile node. If a co-located care-of address is being used, the mobile node will receive the packet and decapsulate it itself.

2.2. Route Optimization

With the basic Mobile IP protocol, all packets for a mobile node that is away from home must be routed through its home network and home agent, possibly severely limiting the performance and reliability of packet delivery. An extension to Mobile IP, known as *Route Optimization*, enables a correspondent node to learn the care-of address for a mobile node and to tunnel its own packets directly there, bypassing the home agent [8,9].

With Route Optimization, a correspondent node maintains a *binding cache*, in which it caches the binding of one or more mobile nodes. When sending a packet, if the sender does not have a binding cache entry for the destination mobile node, the packet will be delivered to the mobile node's home network, intercepted its home agent, and tunneled to the mobile node's care-of address. If the home agent supports Route Optimization, it can then inform the original sender of the packet about the mobile node's current binding by sending it a *Binding Update*, giving the sender an opportunity to cache the binding. When the mobile node later moves to a new care-of address, packets can be forwarded to the new care-of address by the mobile node's old foreign agent, as described in the next section. When it forwards a packet, it sends a *Binding Warning* message to the mobile node's home agent, asking it to send a new Binding Update to this correspondent node to update its cache.

2.3. Smooth Handoff

Another part of Route Optimization is called *smooth handoff* [8,9]. This feature attempts to improve packet delivery during handoff by allowing the foreign agent in a mobile node's previous location to forward packets to it in its new location after it moves. The mobile node sends a Binding Update to this previous foreign agent, informing it if its new care-of address, and the previous foreign agent then tunnels any subsequent packets arriving for the mobile node to this new location. This forwarding allows any packets that were in flight to the mobile node when it moved, and any packets sent by correspondent nodes that have not yet learned the mobile node's new care-of address, to be forwarded rather than being discarded by the foreign agent. This is particularly desirable with TCP, where dropped packets are otherwise interpreted as a sign of congestion in the network, causing TCP to reduce its effective transmission rate [1].

3. MOBILE IP, TCP, AND ACTS INTERACTION

As noted in Section 2, Mobile IP is generally tuned to perform well in terrestrial networks, with limited distances and propagation delays. High latency GEO satellite links like ACTS could severely impact the performance of Mobile IP. A long propagation delay might cause Mobile IP to retransmit some control messages during registration, thereby delaying the registration process and increasing handoff time. This could cause higher-layer protocols on the mobile node to react badly, possibly even timing out and dropping the connection altogether.

The Registration Request message sent by a mobile node for registration must be protected by replay protection, which generally uses a *timestamp* in Mobile IP. When the home agent receives a Registration Request, it determines if the timestamp is within an acceptable range. A long propagation delay will increase the probability that this registration is dropped because the message is received too late. In addition, a long propagation delay will require the mobile node to wait a long time for the return Registration Reply message, delaying the mobile node's ability to recover and retransmit the Registration Request in the case that either it or the Reply message are lost by the network.

From the transport protocol's point of view, Mobile IP handoff time will be recognized as a period during which all the transmitted packets are dropped. Since TCP has been designed generally for traditional wired networks, packet drops are assumed to be caused by congestion [1]. Therefore, TCP will reduce its transmission rate and increase its timeout value, resulting in very poor link utilization.

The longer registration time could also make the Mobile IP routes obsolete or stale. For a large bandwidth-delay network such as ACTS, a number of TCP packets may be in flight as the routes are changing, degrading TCP performance. Packets may be lost or additionally delayed due to forwarding from a mobile node's previous foreign agent. Cross-traffic effects between TCP and routing protocols like Mobile IP over satellite links are areas that need further investigation.

We conducted experiments to observe TCP performance during Mobile IP registration: when a mobile node moves from its home network to a foreign network, from one foreign network to another, and from a foreign network back to its home network. Correspondent node behavior when receiving late Binding Updates due to the longer propagation delay was also observed. The experiments also measured the performance of smooth handoff, with the goal of improving performance during handoff.

4. MOBILE IP AND TCP PERFORMANCE OVER ACTS

Figure 2 shows the network configuration we used in our experiments [6]. We used wired Ethernet rather than wireless LAN subnets for local area network connections, in order to allow greater control over the local network links. Since the experiments placed emphasis on the performance difference between the low latency local area and the high latency ACTS links, the results should also be representative of the performance obtainable using wireless LAN links together with ACTS. In particular, Ethernet has a maximum roundtrip propagation delay of 51.2 μs, while a typical wireless LAN with 2 miles transmission range has a maximum roundtrip propagation delay of about 20 μs. This small difference in propagation delay is negligible compared to the ACTS propagation delay of about 512 ms (512,000 μs). Bandwidth differences between Ethernet and wireless LAN should not change the behavior of Mobile IP during handoff, because Mobile IP uses only a small amount of bandwidth for its packets.

We conducted two types of mobile node movement tests to analyze the performance of Mobile IP and TCP during handoff. One set of movements was from a subnet on one side of the ACTS link to one on the other side (e.g., from a subnet of Network A to a subnet of Network B, shown in the lower-right corner of Figure 2). The other set of movements was between subnets on the same side of the ACTS link (e.g., between different subnets of Network B).

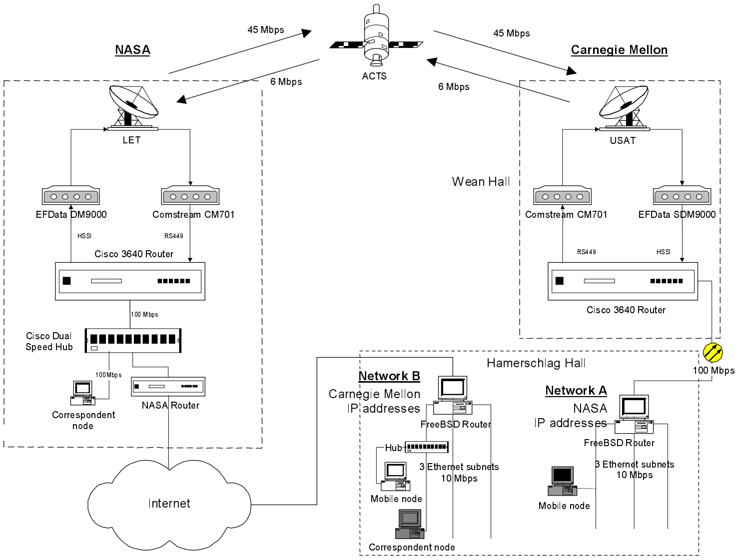

Figure 2: Mobile IP Experimental Network Configuration

4.1. Registration Performance

The first set of experiments we conducted attempted to measure the effect of the ACTS link on TCP during Mobile IP registration. We found that the high propagation delay of the ACTS link significantly reduced TCP throughput due to the additional time required for the registration to complete after the mobile node has moved to a new subnet. For example, Figure 3 shows the sequence numbers of a TCP connection, plotted versus time during Mobile IP registration. In this case, the mobile node has moved from a subnet of Network A to a subnet of Network B (Figure 2), and correspondent node is on Network B, and Mobile IP registration is done over the ACTS link. Each TCP segment transmission by the correspondent node is represented by a black diamond at the correct sequence number and time in the graph, and each acknowledgement from the mobile node is likewise represented by a gray square.

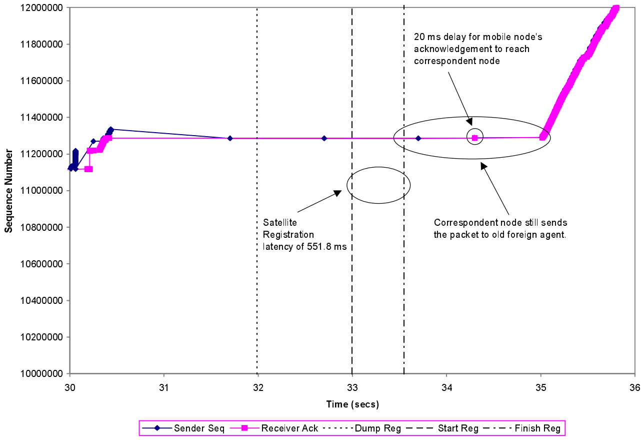

Figure 3: Mobile IP and TCP Performance during Handoff

There are three vertical lines in the graph that show the time at which specific Mobile IP events occur at the mobile node. The "Dump Reg" line represents the time at which the mobile node discards its current foreign agent because it does not receive its Agent Advertisement messages. The "Start Reg" line represents the time at which the mobile node starts the registration process with a new foreign agent. Finally, the "Finish Reg" line represents the time at which the mobile node receives the Registration Reply from its home agent, indicating that the registration process is complete. In this case, the complete registration process took 551.8 ms, from the time the mobile node initiated the registration to the time it received the Registration Reply message. This delay is due largely to the propagation delay of the satellite link over which the Registration Request and Registration Reply had to be transmitted. The first couple of packets from the correspondent node after the handoff were still delivered to the old foreign agent on the other side of the satellite link, resulting in a delay of 564 ms between the time of packet transmission and the time at which the mobile node acknowledged each. The acknowledgement from the mobile node, however, took only 20 ms, since it was sent directly to the correspondent node. Later, the correspondent node receives a Binding Update message and starts sending each packet to the new foreign agent, which is an address in the same network as the correspondent node. Although Mobile IP and TCP performed correctly over the ACTS link, the performance was significantly hurt by the long satellite propagation delay.

4.2. Smooth Handoff Performance

We conducted a second set of experiments, to observe how smooth handoff would improve the performance of Mobile IP and TCP during handoff from one foreign network to another. Smooth handoff is designed to improve performance by reducing the number of packets that are dropped during handoff, although some packets may still be dropped if they arrive after the mobile node has left the foreign network but before the foreign agent has received the Binding Update giving the new care-of address.

From our experiments, however, we found that smooth handoff did not accomplish its intended goal when using the ACTS link. By examining the sequence numbers during handoff, we found that most of the packets were dropped even before the mobile node realized that it had left the previous foreign network. Figure 4 is one example that shows this behavior.

With Mobile IP, it is common for a mobile node to recognize that it has moved to a new subnet, well after the first retransmission by TCP at the correspondent node, since Agent Advertisement messages are sent at most once per second. Even the very large roundtrip time (about 512 ms) over ACTS cannot prevent TCP from

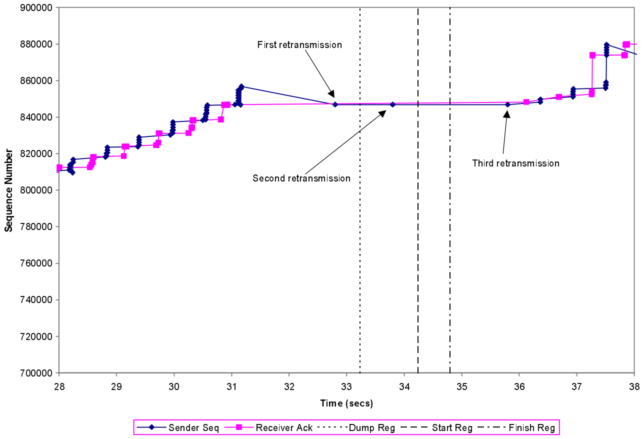

Figure 4: Mobile IP and TCP Performance with Smooth Handoff

going into slow start, since it is almost always the case that TCP will time out at least once before the handoff completes. When this happens, TCP interprets this as a sign of network congested, and will retransmit only the last unacknowledged segment, doubling the time between successive retransmission attempts using exponential backoff (Figure 4). Even after the registration process in the handoff has finished, the TCP data flow will not resume until an acknowledgement for this segment is received. As illustrated in Figure 4, 2 seconds or more may elapse between the time at which the registration is complete and when TCP actually begins to resume communication. Resumption of data flow in this case depends on TCP retransmission, although it would be better if TCP transmission could be resumed at the moment when the registration actually completes.

5. IMPROVING MOBILE IP AND TCP PERFORMANCE OVER ACTS

The performance of Mobile IP and TCP during handoff can be improved by extending smooth handoff with the capability of buffering the packets during handoff. Using buffering, packet drops during handoff can be eliminated, and therefore unnecessary TCP retransmissions can be prevented. There are several issues related to the implementation of such buffering.

The first issue is when to start buffering. In order to be efficient with both CPU utilization and memory consumption, buffering should start the moment the mobile node is unreachable or is about to move from the current foreign network. Determining when this happens is very difficult, however, since although some physical and link layers may provide some indications such as signal strength measurements, it is not standard. A method is needed to solve this problem, independent of the particular lower layers used (possibly augmented by indications from the lower layers).

The second issue is the number of packets to buffer. We would in general like to buffer all packets that may be dropped during the handoff, but it is difficult to calculate the size of buffer this would require. A small buffer might not be enough to prevent packet drops during long handoffs, whereas a larger buffer might unnecessarily waste resources and prevent the foreign agent from serving more mobile nodes.

The final issue related to the implementation of buffering is how to resend the buffered packets back into the network. If the entire buffer is put into the network at once, then these packets might create congestion, which might force some of the packets or others to be dropped, possibly defeating the purpose of buffering.

Our results for smooth handoff performance, discussed in Section 4.2, demonstrated the need to notify the sending TCP layer that the handoff is complete, so that transmission to the mobile node can be resumed immediately. In this case, buffering can be implemented to solve the problem. Instead of preventing packet drops during handoff, we implemented buffering of only a single packet, to trigger an acknowledgement from the mobile node to the correspondent node. As a result, the TCP layer on the correspondent node will receive this acknowledgement and begin retransmitting a segment immediately rather than waiting for the next retransmission attempt scheduled by the exponential backoff.

We implemented this by modifying the foreign agent to buffer the last packet that is transmitted to the mobile node. Upon receiving a Binding Update message from the mobile node's new foreign agent as part of smooth handoff, the buffered packet will be sent to the new foreign agent. This packet will arrive at the mobile node, and the mobile node's TCP stack will acknowledge the packet. This acknowledgement will be sent to the correspondent node, which will then resume its TCP transmission by (re)transmitting the next segment. Only a single packet is needed to initiate this procedure, and hence a buffer size of one packet is sufficient. This reduces the complexity of the buffer handling and also eliminates the problem on how to retransmit the buffer back to the network. Although the exact moment to start buffering is still not known, the foreign agent can simply always buffer the last packet sent to the mobile node.

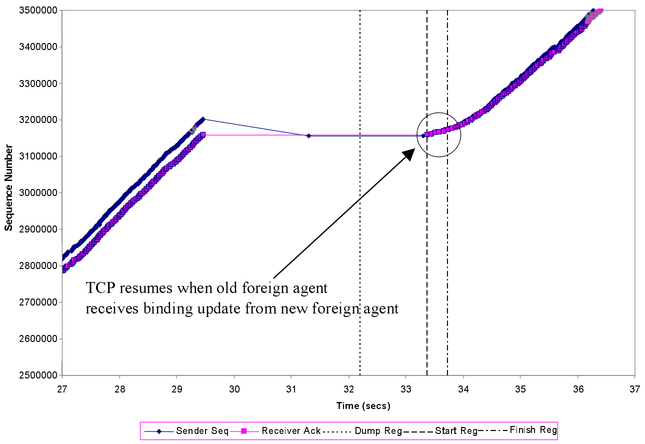

Figure 5: Performance with Single-Buffering Extension in Smooth Handoff

Figure 5 shows the TCP sequence number behavior over ACTS during smooth handoff with this single-buffering extension. In this case, TCP retransmission begins almost immediately after the mobile node starts the registration process, since the Binding Update message to the old foreign agent is sent at the same time as the Registration Request message. This Binding Update will arrive at the old foreign agent at approximately the same time as the Registration Request message arrives at the home agent. The old foreign agent will send the buffered packet to the new foreign agent, which will deliver it to the mobile node; the mobile node's TCP layer will then acknowledge the packet to the correspondent node, causing it to then (re)transmit the next segment. Each segment sent by the correspondent node will then be forwarded by the old foreign agent to the new foreign agent, until the registration process completes.

An important aspect of this enhancement is that it does not modify any part of Mobile IP or the TCP protocol. It uses messages that exist in the Mobile IP smooth handoff extension and also uses standard TCP protocol mechanisms. All security aspects of Mobile IP and all TCP end-to-end semantics are preserved.

The performance improvement achieved by implementing this single-buffering extension could be very significant, even for networks other than ACTS. For example, for a 2 Mbps WaveLAN wireless LAN network, 1 second of inactivity could be used to send over 200 kilobits of data. The improvement would be much more significant when the mobile node is frequently moving between subnets.

6. CONCLUSIONS

We have examined the performance of Mobile IP and TCP over network connections using ACTS and have analyzed the cross-traffic effects between TCP and Mobile IP in such connections. Our measurements show a substantial performance loss due to the large propagation delay of ACTS and the delays inherent in Mobile IP's movement detection mechanism using Agent Advertisements. The smooth handoff mechanism of Mobile IP Route Optimization reduces but does not eliminate this effect. We have implemented a simple technique that significantly improves this performance yet requires only buffering space at the old foreign agent for a single packet. Our motivation in this work was to examine ACTS and other GEO satellite networks as a part of the global mobile communication system we are developing at Carnegie Mellon together with Caterpillar. This work is applicable to any network connections using large bandwidth-delay networks such as these, and should also be of use in other types of networks such as wireless LANs.

ACKNOWLEDGEMENTS

The ACTS measurements and analyses presented in this paper are joint work with graduate students Satish Shetty, Roy Laurens, Parag Manihar and Ioannis Pavlidis of the Information Networking Institute at Carnegie Mellon University. This paper draws upon their M.S. thesis work and upon the paper of Shetty and Laurens published at the 1999 International Mobile Satellite Communication Conference [10]. We are also very grateful to Mike Zernic and Adesh Singhal at the NASA Glenn Research Center for helping us with the experiment logistics and for collaborating on the experiments; and to the ACTS Control Room Team, especially Rich Reinhart and Jeffery Glass, for putting up with the equipment configuration issues involved.

REFERENCES

[1] M. Allman, V. Paxson, and W.R. Stevens, TCP Congestion Control, RFC 2581, April 1999.

[2] S. Deering, ed., ICMP Router Discovery Messages, RFC 1256, September 1991.

[3] D.B. Johnson, Scalable Support for Transparent Mobile Host Internetworking, in *Mobile Computing*, edited by T. Imielinski and H. Korth, Chapter 3, pp. 103–128, Kluwer Academic Publishers, 1996.

[4] D.B. Johnson and D.A. Maltz, Dynamic Source Routing in Ad Hoc Wireless Networks, in *Mobile Computing*, edited by T. Imielinski and H. Korth, Chapter 5, pp. 153–181, Kluwer Academic Publishers, 1996.

[5] D.B. Johnson and D.A. Maltz, Protocols for Adaptive Wireless and Mobile Networking*, IEEE Personal Communications*, pp. 34-42, February 1996.

[6] R. Laurens, P. Manihar, I. Pavlidis, and S. Shetty, Simulation and Enhancements of Internet Protocols over ACTS, M.S. Thesis, Information Networking Institute, Carnegie Mellon University, May 1999.

[7] C. Perkins, ed., IP Mobility Support, RFC 2002, October 1996.

[8] C.E. Perkins and D.B. Johnson, Route Optimization for Mobile IP, in *Cluster Computing*, Special Issue on Mobile Computing, 1(2):161–176, 1998.

[9] C. Perkins, and D.B. Johnson, Route Optimization in Mobile IP, Internet-Draft, draft-ietf-mobileip-optim-09.txt, February 2000.

[10] S. Shetty and R. Laurens, "Analysis and Improvement of Mobile IP Performance on ACTS," International Mobile Satellite Communication Conference, June 1999, Ottawa, Canada.