Analysis and Comparison of Different Host Mobility Approaches

Mevlut Erdem Kurul, Rommel Pesigan and Brad Weekes
Department of Computer Science
University of California, San Diego
La Jolla, CA 92193-0114
{mkurul,rpesigan,bweekes}@ucsd.edu

Abstract

Host mobility is becoming an important topic in networking due to the recent growth of wireless networks. There have been a number of different approaches in literature describing support for this topic. In this paper, we survey and compare four different approaches, Mobile IP, I-TCP, the End-to-End Approach, and Cellular IP, based on a set of requirements for host mobility. Between these four, we observed that there is no best solution because each approach has some deficiencies in fulfilling these requirements. Finally, we describe an alternative "hybrid" solution for host mobility by examining the tradeoffs between performance and network transparency.

1. Introduction and Problem Statement

The growth of mobile computing devices and wireless networking products introduces problems in host mobility, where users expect to be able to communicate and access information from anywhere. The problem becomes even more complicated if the user moves during a live connection, which is defined as migration.

Different approaches have been proposed for supporting mobility from different points of view. All these approaches should be considered based on the following requirements: ease of deployment and scalability, handoff or ease of migration, routing, network breakage, and bandwidth utilization.

When dealing with mobility, it is important to consider each of these requirements with sufficient attention. First of all, the success of any proposal depends on how backward compatible and scalable it is with the current technology. Another issue is that host migration should be handled efficiently and transparently to the user. The cost of migration is an important issue in host mobility where performance is defined by how frequent a mobile user moves. Also, the unreliable nature of wireless networks raises the question about network breakage. What happens if a home agent or a mobility support router crashes? A good protocol must handle such crashes without causing severe loss in performance. Host mobility routing is another important issue. How optimal the routing is has a direct effect on performance and network bandwidth utilization. Finally, the efficient use of bandwidth is always important to consider because of its limited availability. In some protocols, broadcasting is used during migration or for locating a mobile host. Such broadcasting scheme makes inefficient use of bandwidth. All these requirements extracted from [6] define our problem domain.

The goal of this paper is to analyze and compare four different protocols by evaluating how each fulfilled some of the important requirements of host mobility. Mobile IP for IPv4 [5] is Internet Engineering Task Force's proposal, which is based on a network-layer solution. I-TCP [2] addresses the performance problems of Mobile IP by relying on an indirect transport protocol model. The third involves an End-to-End solution [8], which uses DNS queries and offers TCP migration during a live connection. Finally, Cellular IP [3,4,9] is a micro-mobility protocol, which provides flexibility and robustness in small networking areas. We chose these four protocols because each of them presents a fundamentally different attempt in supporting mobility, providing solutions in the network layers and above.

The rest of this paper is organized as follows: Section 2 briefly describes each of the protocols. In section 3, we analyze and compare the protocols based on a set of important requirements for host mobility. In section 4, we further discuss the issues of mobility and describe an alternative hybrid solution, where we state our preference between performance and network transparency. Section 5 presents our conclusion.

2. Protocols

This section surveys four different approaches to host mobility. Mobile IP for IPv4 [5] is a network-layer solution whereas I-TCP [2] offers an indirect transport protocol solution. The End-to-End host mobility approach [8], which uses DNS lookups, is done at the transport and higher layers. Cellular IP extends Mobile IP to support seamless mobility on top of global mobility.

2.1 Mobile IP

Mobile IP [5] falls into the category of mobility protocols that keeps the mobility of the source and/or receiver transparent. Transparency is achieved by allowing mobile hosts to change network subnets, obtaining new IP addresses, without having to update the routing tables throughout the network. A major advantage of Mobile IP's transparent approach is that it allows the network to adopt mobility with little modification and overhead.

The implementation of Mobile IP's transparent approach is done in the following manner. The mobile host is represented by a permanent IP address, called the *home address*, and it is recorded on the home agent, which resides in the mobile host's home network. The home agent essentially masks itself to the rest of the network as the mobile host, and therefore all packets intended for the mobile host go through it. While the mobile host changes network subnets, it obtains a new, temporary IP address, called a *care-of address*, from a *foreign agent* associated in this new subnet. The foreign agent then updates the home agent with this information, allowing it to keep a record of the mobile host's current location. When a packet arrives to the home agent from a correspondent host in the Internet, it looks up this care-of address for the particular mobile host, constructs an IP header with the care-of address as the source address, and encapsulates the received packet within this new IP header, known as IP in IP or IP encapsulation, and transmits the packet. By adding the extra IP header it enables the existing infrastructure to route the packet from itself to the mobile host using the destination address of the outer IP header – a technique known as tunneling.

When a mobile host enters a foreign network it must discover the foreign agent within this network – this process is known as *agent discovery*. The agents send out beacon messages that the mobile hosts listen for to tell if they have migrated to another network. Once the mobile host detects that it has migrated to another network and has found the foreign agent, the mobile host needs to register with the agent. This process consists of the mobile host initiating a registration request to this new foreign agent. The foreign agent then forwards the request to the home agent, allowing it to update the care-of address. The home agent then replies to the foreign agent, which the foreign agent forwards to the mobile host.

2.2 I-TCP

Network-layer mobility, which attempts to handle host relocation within the IP layer, suffers from a significant cost of operations, complexity, and processing problems, resulting in poor performance. I-TCP [2] is an indirect approach to mobility, which is handled at the transport layer. It is proposed to minimize unreliability of the wireless media and address the performance problems of Mobile-IP. In I-TCP, a Mobile Host (MH) communicates with a Fixed Host (FH) via its Mobile Support Router (MSR).

Performance cost in Mobile-IP generally results from cell crossovers. Lost data packets forces TCP to trigger exponential back off. Then, congestion control and recovery mechanisms are put into effect, which may last several seconds even after the connection is reestablished. I-TCP tries to minimize this problem by using indirection at the transport layer.

I-TCP basically splits any interaction between a MH and a machine on the fixed network (FH) into two separate interactions. The first interaction is between the MH and its MSR over the wireless connection, while the second interaction involves the MSR and the FH over the wired connection, making the MSR the central point. It is important to consider that I-TCP still relies on the underlying Mobile-IP on the MSR to support communication with the mobile host over the wireless media. According to [2], I-TCP uses a variant Mobile-IP called Columbia in the network-layer. The separation of wireless and wired networks is used to accommodate the special requirements of a mobile host communicating through the low-speed and unreliable nature of the wireless connection. High-speed and reliable wired connections do not require any special modifications. Therefore, the problems related to mobility and the unreliability of a wireless network can be managed separately from the wired network, where the structure of the TCP/IP implementation remains unchanged.

This separation also allows I-TCP to identify the proper protocol to manage communication over the wireless link, such as notification of events and efficient bandwidth utilization, in order to support link aware and location aware mobile applications. Also, the indirection through the MSR is used to manage the communication overhead of the mobile host better.

The operation of I-TCP is best understood by an example where an MH wishes to communicate to a FH through a MSR (MSR-1). First, the MH sends a request to MSR-1 to open a TCP connection to the FH on behalf of the MH. MSR-1 establishes a socket with the MH address and port number for the FH, whereas the MSR-1 uses its own address and a suitable port number for the MH. Once all these parameters are set, the MH can communicate with the FH through MSR-1. As a result, the FH only sees an image of the MH residing on MSR-1. Furthermore, consider the situation where the MH moves to another cell and another mobile support router, namely MSR-2. Now, the states of the two sockets from MSR-1 are handed over to MSR-2. The sockets on MSR-1 are then deleted after MSR-2 creates essentially the same sockets and parameters. This allows the connection to not be reestablished again since the connection endpoints remain the same. All the operations are hidden from the FH. However, special I-TCP calls might be needed by the MH to establish an I-TCP connection with the FH.

Briefly, the I-TCP approach minimizes the performance degradation of wireless communication on the wired network by separating the interaction into two forms of communication. All the registration, hand-off, and other changes are performed on the wireless network by using the underlying Mobile IP and an indirect transport protocol model. Although, it does not require any changes on the TCP/IP structure on the wired network side, both the TCP and Mobile IP code on the MSR need to be modified. In addition, the I-TCP library must be adopted by the MHs.

2.3 End-to-End Approach

The End-to-End host mobility approach [8] is based on providing seamless connection like the other host mobility protocols, but unlike these protocols, E2E does not attempt to make mobility transparent. The argument for this is the typical E2E argument where the upper layers, including the application layer, are involved in establishing a connection because these layers know the most about how the data are being transmitted.

The key concept of the E2E approach is that it uses the already existing DNS as an aid in host lookup. When a mobile host leaves a network, enters another, and obtains a temporary IP address, the mobile updates the DNS about its current location. This approach is natural because it uses an existing component of the network, namely DNS. Another distinct feature of E2E is how migration is done, which occurs when the mobile host changes attachment points in the network. E2E proposes a new migrate TCP option that is included in the SYN segments. This new option identifies a SYN packet as part of a previously established connection rather than a request for a new connection. This is used when the mobile host attaches to a new subnet and the connections from the previous subnet need to be migrated over to the new subnet seamlessly. A token is negotiated at the beginning of a connection setup and this can be used to restart a previously established TCP connection from a new address by sending this new migrate option. This migrate option contains the token that identifies the previous connection. The migrated connection maintains the same state as the previous connection, which is necessary in order for TCP congestion control and reliability mechanisms to function properly.

2.4 Cellular IP: (Diagram provided by [3])

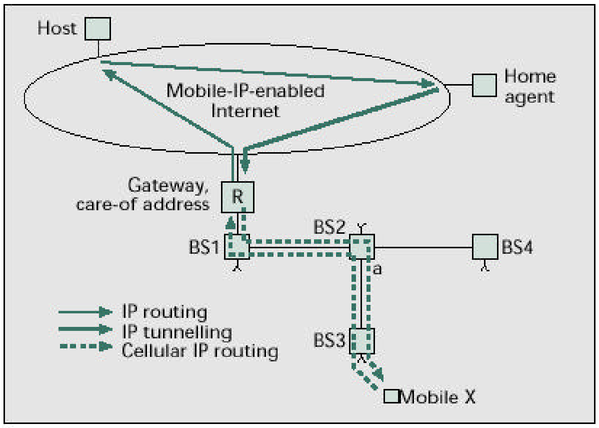

Figure 1. *Cellular IP access network.*

Cellular IP [3,4,9] combines the benefits of wireless cellular communication to the Mobile IP protocol to support seamless mobility and passive connectivity of mobile hosts within a small networking infrastructure. It essentially extends Mobile IP so that host mobility can be done effectively in both the global and local networking environment. Data packets from the Internet are sent to the mobile host using regular Mobile IP. However, the foreign agent representing the host's care-of-address does not forward packets directly to the host's home address. Instead, these packets enter a local *wireless access network* where they are eventually forwarded to the host's location by *base stations.* In Cellular IP, base stations function as wireless access points that hosts associate with as well as IP routers that forward packets from the host to the Internet. Base stations maintain routing caches that describe a hop-by-hop routing path for packets being sent from the Internet to the mobile host and vice versa. Routing caches are updated either when base stations encounter data packets being sent from the host, when the host transmits special routing-update packets, or when they receive broadcasted packets from other stations.

Cellular IP's *semisoft handoff* allows for fast host migration from one base station to another with almost no packet losses. Semisoft handoff takes advantage of the fact that some mobile hosts can listen to two base stations at once, namely its current station and the station it wants to migrate. Before the actual handoff process, the mobile host transmits a special packet, while still associated with the old base station, to the new base station to establish a new routing cache for the host. After a certain delay, regular handoff can proceed. This delay is to ensure that data packets are being sent to both the old and new stations as the host prepares for migration.

When migrating hosts can only listen to one base station at a time (namely the old base station), this form of handoff is known as *hard handoff*. Like semisoft handoff, the migrating host transmits a packet to the new base station to establish a new routing cache for the host; however this is done *after* the host has migrated to this new station. Packets transmitted to the old base station will cease as soon as the new routing cache has been established.

Another attractive feature of Cellular IP is its support for passive connectivity using *paging*, which allows mobile hosts to be easily located whether they are actively transmitting data packets to the Internet or not. To support this, base stations maintain paging caches, in addition to routing caches, which are updated when idle hosts periodically transmit special paging-update packets. With this additional support, packets addressed from the Internet to the mobile host can easily reach the host despite its current state of connection.

3. Analysis and Comparison

This section analyzes and compares the four protocols summarized above based on the following requirements for host mobility: deployment/scalability, handoff/ease of migration, network breakage, routing, and bandwidth utilization. We emphasize that these are not the complete requirements and that other factors, such as security, are important as well. But we feel that the five requirements we will soon discuss are the most representative.

3.1 Deployment/Scalability

Host mobility protocols must support ease of deployment and scalability with the current networking infrastructure. It is desirable that these protocols make very minimal changes to this infrastructure, be backward compatible with existing protocols, and scalable to support wide area networks.

We first discuss two different Mobile IP approaches – one for IPv4 and one for IPv6 [7]. Although the Mobile IP protocol can be deployed with virtually no changes to the current infrastructure, the resulting performance can be very poor if there is no support for any routing optimization techniques. Today's networking environment is based on IPv4 and [5] discusses how routing optimization can be supported using binding caches and updates. However, mobility was never considered in the initial design of IPv4. As a result, hardware and software modifications are necessary in order to support binding caches and updates, which can be expensive. Without support for these optimizations, Mobile IP will be limited to indirect packet routing, which wastes network resources and hinders performance.

Mobile IP in IPv6 does not have these compatibility problems because Mobile IP support is embedded in the current IPv6 design [7]. Headers in IPv6 contain fields specific for binding update information. With this support, binding updates can be sent along with data or acknowledgements, allowing for better network utilization. As a result, there will be minimal changes to hardware or software to support Mobile IP in IPv6, which means no additional cost for deployment. However, it is still not clear how transparent IPv6 deployment is to the current infrastructure that is entirely IPv4 based. Until then, Mobile IP for IPv4 will remain.

I-TCP has similar compatibility issues as Mobile IP with routing optimizations. Changes are required in the existing routers in order to support the wireless communication portion of the I-TCP connection. These routers, (essentially the Mobile Support Routers, or MSR), must be able to adapt with the underlying network-layer mobility support of I-TCP as well as with transport-layer changes due to the I-TCP design. This is a major disadvantage of I-TCP because it will not work without these modifications.

In the wired portion of the I-TCP connection, compatibility support is not necessary since it will simply use the existing protocols for host communication. This is because the Fixed Host (FH) only sees an image of the Mobile Host (MH) residing at the MSR, making the wireless communication transparent from the FH.

Cellular IP does not have any major compatibility issues since the protocol is implemented around the existing IP model. Cellular IP can also be scaled and deployed easily since the design is not overly complex. However, Cellular IP still requires modifications to existing base stations, but these software changes are minimal.

A potential problem for Cellular IP is how base stations manage increasing route and page cachings as the number of mobile hosts increase within the access network. This inherits the overhead problem from routing table lookups that currently exists in IP routers today. However, this can be avoided by applying fast routing table lookup algorithms in the base stations.

Cellular IP also assumes that the most common mobile hosts in the network are *wireless* mobile hosts. This means laptop users that needs a physical modem would not find Cellular IP very useful, unless it uses a wireless modem card.

The E2E host mobility approach requires the greatest number of changes in the current infrastructure out of the four protocols. The wide scale modifications in order to support the E2E solution consist of changes in the transport and application layers of all hosts in the Internet. This is needed for the new migration option in the protocol to identify a previous connection in order to accomplish seamless mobility. A protocol that requires wide-scale modifications to solve wide-scale networking problems is not desirable. As a result this severely limits its chances of being deployed widely.

There is no favorable host mobility protocol that will completely solve the issue of deployment. Cellular IP appears to be better, but its design uses and extends Mobile IP. So, based on our analysis, Mobile IP without routing optimizations seems to provide the best approach.

3.2 Handoff/Ease of Migration

One of the main challenges of mobility is how efficiently hosts move from one base station to another without disrupting the current connection. It is desirable to have transparent migration between the correspondent and mobile hosts, but not at the expense of performance.

Mobile IP handoff requires that the mobile host always register its current location to its home agent every time it obtains a new care-of-address (or associates with a new foreign agent). However, the host does not inform its new location to its previous foreign agent, causing any packets forwarded to this agent to be lost until the host registers its new location to its home agent. Allowing the mobile host to notify its previous foreign agent of its new location so that the old agent can forward any incoming packets or old binding-caches to the new agent solves this problem. This will also support a smoother and more efficient handoff.

However, Mobile IP handoff does not perform well if mobile hosts are moving frequently, which might be the common case for a smaller networking environment. Cellular IP deals with this problem by essentially managing all handoff procedures *within* an access network. In Cellular IP, the host does not need to register its new location every time it moves. Instead, the base stations within the access network will do the location management, which is transparent to the home agent. This approach, however, comes with a price. There will be times that a host migration will cause packet drops when hard hardoff is in effect, much like the initial problem stated above in Mobile IP. Cellular IP does not allow new base stations to notify previous base stations that a new host has just migrated to it. However, this problem is non-existent during semisoft handoffs. Semisoft handoffs virtually eliminate the occurrence of packet losses during migration, but at the expense of increasing the network load.

I-TCP handoff inherits the handoff procedures from the underlying Mobile-IP variant, which, according to [2], uses the Columbia Mobile-IP. [6] discusses some of the migration issues of Columbia, such as migration detection support in a local area network only.

I-TCP handoff requires a state transfer from one MSR to another. This state normally includes all active I-TCP connections at the MH and the previous MSR address. In addition to the state transfer, the migration process involves numerous operations (ten steps to be exact [2]) such as sending different requests and acknowledgements between the MH and the MSRs involved. These message exchanges and state transfer can be a performance bottleneck for I-TCP migration. In addition to the inherent migration issues from Columbia, this bottleneck can really affect the overall performance of the mobile host, especially if the host is frequently migrating. Even if the mobile host hears the beacons from the two MSRs simultaneously (analogous to Cellular IP’s semisoft handoff), the migration process still requires six steps due to the state transfer procedures [2]. Cellular IP’s semisoft handoff does not have this problem since Cellular IP is implemented in the network-layer; therefore, no state transfers are required.

Like I-TCP, the handoff procedures in the E2E host mobility approach are complex, resulting in large computational overheads. The handoff process requires halting the current TCP connection from the mobile host, which is then migrated to another network attachment point. This requires efficiently managing and synchronizing the state of the connection between the host and the new attachment point, such that acceptable handoff performance can be attained. If synchronization is not handled well, many packets can be lost, which is undesirable if the connection is a TCP session. These packet drops will result in frequent timeouts and packet retransmissions that will affect the overall migration performance as well as increase the network load.

Despite the complex handoff process, E2E has the advantage of using knowledge from the higher layers about the current connection state. When a handoff is completed, it is desirable for a TCP connection to perform any congestion control mechanisms since the bottleneck router is unknown. This is not possible if handoff is transparent to the correspondent host. With E2E, the state of the transport and higher layers are known so that appropriate action can be initiated.

Based on these observations, Cellular IP handles migration very well because of its simple design and fast handoff approach. Out of all the four protocols, Cellular IP is the most efficient in terms of supporting hosts that are frequently moving.

3.3 Network Breakage

A mobile host protocol must be able to adapt to severe packet losses and must remain operational in case of network breakage, so that continued service can be satisfied to both the correspondent and the mobile hosts.

The Mobile IP, I-TCP, and Cellular IP approaches use and depend on intermediate agents or routers to establish a connection, thus exposing single points of failure. In Mobile IP, if a home agent goes down and correspondent hosts do not have a current binding cache, then they will not be able to communicate with the mobile host. The problem is amplified considering that a home agent can potentially serve a large number of mobile users. This case applies if a foreign agent breaks as well. In I-TCP, the situation is worse since MSRs don't support any caching schemes similar to Mobile IP's binding cache, meaning that any connection states between the correspondent and mobile hosts are forever lost if a MSR breaks. In Cellular IP, a base station failure is not a big concern since packets are simply rerouted to different base stations with a valid routing path. However, since Cellular IP is essentially a protocol extension for Mobile IP, it inherits Mobile IP's problems if a gateway router suddenly malfunctions.

In order to achieve protocol robustness, some kind of failure detection and handling mechanism is required. A simple solution is to use redundant resources for these intermediate routers. For example, multiple home agents (or foreign agents) will eliminate any routing dependencies from a single router. For I-TCP, MSR redundancy can potentially solve this problem as well. However, this redundancy approach will add extra overhead due to maintaining the states of each redundant resource, as well as added deployment cost.

In Mobile IP, there exists a recovery mechanism in the case of foreign agent breakage. In this situation, the mobile host would listen for signals sent from the foreign agent and after a timeout without hearing any responses, it finds another agent to represent as its new care-of-address

As mentioned above, breakage of the intermediate base stations in Cellular IP will simply cause packets to be rerouted to different paths within the access network. Like normal IP routing, some delay will be experienced in order to reconfigure the routing caches of the new path; however, this can be solved using the same methods that are employed in normal IP routing.

A gateway router crash in a Cellular IP network can also be solved using the approach Mobile IP takes when a foreign-agent crashes. However, there will be additional overhead in establishing this care-of-address because intermediate base stations need to reconfigure their routing caches to create a path from the host to the new gateway router.

The End-to-End approach has great advantages over the other protocols in terms of dealing with network breakage. It does not depend on any intermediate routers or stations to establish communication between two hosts except for those common routers that handle packet forwarding. If these common routers fail, packets are simply forwarded to different routes in the network, much like what is done today. The current infrastructure has the capability of supporting these rerouting of packets. Therefore, in terms of dealing with network breakage or fault tolerance, the End-to-End host mobility is the best approach.

3.4 Routing

Good routing performance is always desirable whether hosts are fixed or not. Poor routing can result in low bandwidth and resource utilization. This property is important in host mobility simply because of its dynamic nature. As hosts relocate, host mobility protocols must efficiently manage any available resources in the network in order to achieve good overall performance.

Mobile-IP, I-TCP, and Cellular IP all use a *triangle* approach to routing. This presents a level of indirection that can affect how packets are efficiently being routed. Such problems include added latency in packet transmission between the correspondent and mobile hosts. Also, packet transmission between two hosts in close proximity can potentially be routed to their common, "pivot" router than is far away.

Mobile IP presents some route optimization solutions in both the IPv4 and IPv6. The idea in both versions is to make the home agent send binding updates containing the care-of-address of the mobile host and the address of the correspondent host. The correspondent host would then store this update in its binding cache. Any subsequent packet transmission from this host would simply check its binding cache if it contains the mobile host's care-of-address. If it does, it uses this address to directly transmit its packets to the mobile host, thus avoiding any communication with the home agent. In situations where the home agent receives a packet from the correspondent host, the home agent assumes that this host does not have a valid binding cache entry for the mobile host. As a result, the home agent sends a binding update to the correspondent host. In full deployment this optimization allows routes for mobile hosts to be similar to those of fixed hosts. However, at the start of transmission, triangle routes may still occur when binding caches are not present at the correspondent host.

[2] does not address any routing optimizations for I-TCP, especially in the transport layer where this protocol is implemented. Therefore, we can only assume that most of the routing optimization, if any, depends on how the lower layers handle this indirect routing.

In Cellular IP, base stations use routing and paging caches to keep track of the path where packets transmitted to the mobile host are forwarded. A property of these caches is that only the recently transmitted cache-update packets are kept. This results in fast lookup time in these caches, which is especially important for access networks that serve a large number of subscribers. In addition, Cellular IP base stations use hop-by-hop routing much like normal IP-forwarding routers. Any routing optimizations that can be applied to the normal IP routers can easily be applied to Cellular IP packet routing. However, because of the presence of gateway routers acting as the host's care-of-address, there is actually *two* levels of indirection between the correspondent host and the mobile host. Any delay due to routing within the access network will be added on with the delay in the global Mobile-IP network.

The E2E approach is entirely different from the three protocols mentioned above. The correspondent host and the mobile host communicate directly without any dependency to triangle routing. It uses no special or atypical routing to implement its solution. The correspondent host simply uses a DNS lookup for host locations much like how normal, fixed hosts use today. As a result, there is no added delay in packet transmission, except for what is already present in the current network. A problem with this approach is the potential for race conditions when the correspondent host queries a DNS lookup when a mobile host moves. However, this is not really a routing optimization issue and it can be solved simply using support from the higher-layers.

The E2E approach to host mobility is the favorable protocol in achieving good routing performance because it is the only protocol that routes packets without any level of indirection.

3.5 Bandwidth Utilization

Host mobility protocols must allow efficient use of bandwidth to avoid performance degradation in the network infrastructure. This is especially important since there are not enough resources available to serve every single host in the Internet.

Mobile IP's triangle routing approach affects how packets are efficiently processed before they reach their destination. Without any routing optimizations, packets must traverse through any intermediate routers in the indirect path, where they will be individually queued and processed before they are sent. This extra queuing and processing step adds unnecessary overhead for packet transmission. This problem can be eliminated if routing optimizations are applied, since these extra routers in the indirect path will no longer queue the packets. The only queuing involved now will be by those routers that are in between the direct path of the correspondent and mobile hosts.

Meanwhile, I-TCP, like Mobile IP, uses the bandwidth efficiently in some instances and poorly in others. For example, like Mobile IP without routing optimization, I-TCP requires the connection to traverse through an indirect path in the network in order to establish mobile connectivity. Also, there is considerable copying overhead within the MSR when packets are being transmitted [2]. However, some amount of bandwidth is utilized during migration since I-TCP buffers packets in flight, preventing retransmission for these particular packets.

Frequent handoffs can also affect how bandwidth is efficiently utilized in the current network infrastructure. In the Mobile IP approach, the foreign agent is consistently forwarding registration requests made by the mobile host to the home agent. This is undesirable in a global environment that supports hundreds of mobile hosts that repeatedly send registration requests to the foreign agent. These requests will eventually load the network and consume too much bandwidth. A possible solution to this problem is to create a cluster, or subnetwork, of foreign agents, where host registration is only necessary if the mobile host roams out of this cluster. In essence, a mobile host can freely move within this subnetwork without having to send frequent registration requests.

Frequent handoffs in I-TCP will result in poor overall performance because of the significant overhead involved in transferring the states of the old MSR to the new MSR. Unfortunately, [2] does not present any solutions to this problem.

Cellular IP adopts a similar approach as the proposed "clustering" solution in Mobile-IP, where the access network and its base stations constitute this "cluster". Cellular IP supports location independent addressing so that mobile hosts need not register its new care-of-address to its home agent every time it moves. Mobile hosts instead keep the gateway router as its temporary care-of-address as it moves within the access network and uses base stations as units that manage host migration. This eliminates the overhead required in managing the mobile host's location if Mobile IP is solely used, especially if the mobile hosts are moving frequently.

However, during semisoft handoff in Cellular IP, the "crossover" station (the station that is shared by both the old and new base station during mobile host handoff) will transmit essentially the same packets to these two stations thus doubling the load in the network for a single host. This does not scale well if a large number of hosts are doing semisoft handoff as well. As a result, there needs to be a balance number of base stations so that they are not too dense or too sparse. Too dense would result in increased network load; too sparse would cause numerous packet drops during handoff, thus lowering performance.

Fortunately, Cellular IP supports paging caches to minimize the load in the network. Paging caches essentially avoids sending broadcast packets to its neighbors when routing cache access results in a miss. When a station sees an invalid routing cache entry, it uses the paging cache (assuming the page cache entries are valid) to route any packets to its destination. Invalid routing and paging caches however results in broadcasting the packets to its entire adjacent base station except to the station it originally came from. In addition, Cellular IP supports passive connectivity so those hosts that are idle or active can be easily tracked down. Passive connectivity minimizes air traffic in the wireless network since idle hosts require less signaling and at the same time allow these hosts to remain connected as long as it wishes.

The E2E host mobility solution is the most efficient in terms of bandwidth utilization because it is not exposed to the inefficiencies inherent in transparent solutions, since E2E is not transparent. As a result, this protocol has a greater control on how bandwidth can be used since it knows the behavior of the transport layers and above.

This form of communication allows E2E to avoid routing and handoff problems that plague the other host mobility protocols. With the ability to migrate a live connection, routing is optimized because the E2E approach does not require any level of indirection. In addition, there is less overhead during handoff. Data communicated between the correspondent and mobile hosts that contain tokens denoting previous connections are contained within the already transmitted SYN packets. Also, packet losses are minimized during handoffs because the connection can be stalled until the handoff is completed. Also, the protocol limits the amount of retransmissions required for a reliable transport such as TCP when losses occur.

4. Hybrid Approach

The major distinction between these protocols, and any other host mobility protocols not surveyed here, is whether they provide mobility service transparently to the existing infrastructure or not. Protocols that are not transparent are designed around the E2E philosophy that better protocol performance can be obtained if greater functionality is given to the upper layers because they know more about how the data are being transmitted. While this may be true, these protocols suffer because they require major modifications to the existing infrastructure, which is not easily accepted by the networking community. On the other hand, the transparent protocols can work without any major modifications but suffer from inefficient routing due to the inflexible nature that this transparency feature brings. This infrastructure transparency and performance tradeoff result in conflicting viewpoints for a best solution to host mobility.

Due to the problems we found through examination of the proposed solutions, we understand that providing a well-rounded approach is not an easy task. However, we can propose a foundation of ideas that may lead to a better overall solution for host mobility, by grouping together smart approaches from existing protocols.

Our first approach was to propose ideas that together provide *transparency* within the underlying infrastructure of the current network because we felt this approach would be more widely accepted. However, we were not satisfied with the performance due to the indirect routing that this approach caused. This can be attributed to lack of network support because the current network infrastructure was designed and optimized for fixed hosts.

Therefore, we have concluded that in order to achieve good performance under host mobility in the current IPv4 network infrastructure, modifications must be made to the existing networking components. We realize that requiring modifications immediately warrants a red flag and often causes the proposed solution to be written off. However, we feel a grouping of non-transparent ideas can provide performance similar to the performance of traditional fixed hosts. Further, we argue that the modifications we propose are far less dramatic than what is needed to support E2E host mobility.

Our methodology incorporates ideas from E2E, Cellular IP, and Mobile IP host mobility protocols. More specifically, these ideas are DNS lookup for host mobility location from E2E, paging and registration from Cellular IP, and Mobile IP with optimized routing. We have excluded I-TCP because we did not find any ideas within the protocol that could contribute to our hybrid approach and as we saw in [1], it suffers from poor performance. Formulations of the ideas mentioned above are discussed below in hopes of providing new ideas to encourage future research.

DNS lookup allows us to take advantage of the fact that DHCP servers update DNS entries upon serving out IP addresses whether hosts are mobile or not. Therefore, as the mobile host changes network subnets, and acquires new IP address, DHCP will update the DNS server giving correspondent hosts the ability to locate the mobile host when establishing a connection. Furthermore, in order to minimize the overhead of the home and foreign agents, it is a good idea for mobile hosts to register periodically rather than every time it enters a new subnet. A paging technique can be deployed using multicast in order to locate a mobile host over a region of several foreign agents. Upon registration, we envision a technique where surrounding foreign agents are notified by the current foreign agent to join a multicast group. The multicast tree must span a large enough region that geographically covers the range that the mobile host may move between registration periods. This gives rise to a tradeoff between the time interval of a registration by a mobile host and the size of a multicast tree. The larger the interval between registration periods, the larger the multicast tree must be in order to cover the region of the wandering host. The multicast techniques as well as the tradeoff amongst registration periods remain an open issue that needs to be researched to evaluate the performance.

Another important issue occurs when the mobile host moves during a live connection. In this case, the connection can be migrated as in Mobile IP. Once a mobile host obtains its new IP address, DHCP will update the DNS entry for further connections by correspondent hosts. This leaves a small time period where the DNS entry is inconsistent during the migration of the mobile host. We argue that this is not a major problem since DNS will handle it by retransmissions. The triangulation caused by Mobile IP after migration will not cause severe performance problems since it will be employed only once due to binding updates being cached immediately by the correspondent hosts. After the correspondent host gets the binding update, packets will then be forwarded optimally. This requires a modification at the IP level for implementing Mobile IP migration.

We would like to mention that the hybrid approach would be insignificant if IPv6 is fully deployed into the current infrastructure because the integration of Mobile IP into IPv6 can provide the scalable and efficient solution that has eluded host mobility protocols up to this point [7].

5. Conclusion

We have presented four different host mobility protocols (Mobile IP, I-TCP, E2E, Cellular IP) and described both their advantages and disadvantages. Based on our analysis there is no ideal protocol that can satisfy all the requirements of host mobility. Therefore, we proposed a hybrid approach that brings together the benefits of existing protocols but requires slight modifications to the existing infrastructure. Hopefully, this approach will encourage future research.

6. Reference

[1] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. Katz, *A Comparison of Mechanisms for Improving TCP Performance over Wireless Links.* Proc. ACM SIGCOMM Conference, Stanford, CA, USA, Aug 1996.

[2] A. Bakre and B. Badrinath. *I-TCP: Indirect TCP for mobile hosts*. Technical Report DCS-TR-314, Rutgers University, Oct. 1994.

[3] A. Campbell, et.al. *Design, Implementation, and Evaluation of Cellular IP*. IEEE Personal Communications, August 2000.

[4] A. Campell and J. Gomez. *An Overview of Cellular IP*. IEEE Wireless Communications and Networks Conferance 1999 (WCNC'99).

[5] D. Johnson, *Scalable Support for Transparent Mobile Host Internetworking*. Wireless Network, Vol. 1, October 1995.

[6] A. Myles and D. Skellern. *Comparison of mobile host protocols for IP*. Journal of Internetworking Research and Experience, 4(4), December 1993.

[7] C.E. Perkins and D.B. Johnson. *Mobility support in IPv*6. In Proceedings of the Second Annual International Conference on Mobile Computing and Networking (MobiCom'96), Rye, New York, USA, November 1996. ACM.

[8] A. C. Snoeren and H. Balakrishnan. *An End-to-End Approach to Host Mobility*. 6th ACM MOBICOM, August 2000

[9] A. Valko, *Cellular IP: A New Approach to Internet Host Mobility*. ACM Computer Communication Review, January 1999.