Micro mobility and internet access performance for TCP connections in Ad Hoc networks

Nilsson Plymoth, Anders; Hamidian, Ali; Körner, Ulf

2004

Link to publication

Citation for published version (APA):

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Micro Mobility and Internet Access Performance for TCP Connections in Ad hoc Networks
Anders Nilsson, Ali Hamidian, Ulf Körner
Department of Communication Systems
Lund University, Sweden
Box 118, 221 00 Lund
Email: andersn, alexh, ulfk@telecom.lth.se

Abstract: In ad hoc mobile networks nodes typically communicate over wireless channels and are capable of movement. These are networks that support multihop communication and can be formed on a temporary basis. This paper evaluates a solution that allows mobile nodes to access the wired Internet and roam from base station to base station. The solution is based on the extension of Mobile IP capabilities to the ad hoc network while a micro-mobility protocol is adapted to support local migration. We evaluate the performance of this solution with regard to reliable transport layer connections. It is shown that a high throughput is possible to achieve for high mobility speeds. It is also observed that, as the number of hops between a mobile node and the base station increases, the throughput is decreased because of the characteristics of the wireless environment and the medium access layer protocol.

1 Introduction

Many portable computing devices such as laptops and PDAs now include wireless connectivity as a standard feature. More people are also carrying computers when they travel, and want access to the Internet anytime and anywhere.

The Internet Protocol (IP) as defined by The Internet Engineering Task Force (IETF) has become the most widely accepted standard for internetwork communication. Today, broadband wireless access networks based on IEEE 802.11 [1] are rapidly being deployed. In addition, other existing wireless technologies are moving towards an all IP infrastructure.

However, a big problem with IP is that it was never designed to support mobility management. Along with the mobility management issues, the new protocols that are currently under development, must also be radio independent. One of the most widely known mobility solutions for IP networks is the IP Mobility Support [2], commonly referred to as Mobile IP. Mobile IP do have some drawbacks and the concept of IP mobility has now been divided into two main categories: macro mobility and micro mobility. Macro mobility is the management of IP nodes at a larger global scale. Once a node enters a cellular or wireless network domain the Mobility management is local to that network; the node is allowed to move within the network and be controlled locally by the micro-mobility management protocol while the mobility management from a global scale remains unchanged.

Another emerging wireless architecture, namely mobile ad hoc networks (MANETs) [3], are networks that can be flexibly deployed in most environments without the need for infras-
Figure 1: The simulated scenario. A mobile multihop ad hoc network is connected to an access network that supports Mobile IP and micro mobility. Nodes in the wireless network are communicating with correspondent hosts on the Internet.

A MANET is a network consisting of a set of mobile nodes, which may communicate with one another and roam around at will. The routing path may consist of a sequence of wireless links without the need to pass base stations (i.e., in a multi-hop manner). This requires each mobile node to serve as both a host and a router. In most cases today, MANETs use IEEE 802.11 network interface cards. MANET applications and scenarios include situations in which a network infrastructure is not available but immediate deployment of a network is required, such as outdoor assembly or emergency rescue.

Integrating and combining these wireless and mobile architectures will facilitate the current trend of moving towards an all-IP wireless environment. This paper evaluates a solution that extends the typical wireless access points to multiple MANETs, each as a subnet of the Internet, to create an integrated environment that supports both macro and micro IP mobility, see Figure 1. From the mobile IP perspective, foreign agents service ranges are no longer limited to hosts within a single wireless hop; the use of MANETs lets mobile hosts immediately utilize available Internet services without concern about disconnection.

The rest of this paper is organized as follows: Section 2 briefly describes the involved protocols. Section 3 presents the evaluated solution and Section 4 presents performance results obtained through simulation. Section 5 discusses related work and Section 6 concludes the paper.
2 Protocol Descriptions

2.1 Mobile IP

Mobile IP is a proposed standard for location independent routing. It makes mobility transparent to applications and higher level protocols like TCP and UDP. Mobile IP allows mobile nodes to have seamless access to the Internet while roaming between different networks. In order to maintain existing transport layer connections while roaming, every mobile node is assigned a home address. The home address enables the mobile node to always be able to receive data as if it was on its home network, i.e., the network to which its home address belongs.

When the mobile node is attached to a network other than its home network, it uses a care-of address. The care-of address is an IP address valid on the foreign network that the mobile node is visiting.

In Mobile IP, the basic mobility management procedure is composed of two parts: the movement detection performed by the mobile node and the registration to the Home Agent (HA). The home agent is a dedicated router on the mobile node’s home network that forwards packets through tunneling to the foreign network. Every time the mobile changes its IP Point Of Attachment (IPPOA), these two steps must be accomplished in order to allow the mobile node to receive packets. However, it is the mobile node that initiates the process by sending a registration request once it has detected that it has moved from one network to another and has obtained a new care-of address. This introduces two causes of latency:

- Movement detection latency: this is the time required by the mobile node to detect that it has changed its IPPOA.
- Registration latency: as the home agent can be located anywhere on the Internet, this process can take a long time and sometimes be impossible to complete. This is obviously, by far, the main expected part of the total handover latency.

In the case of a quickly moving mobile node that changes its IPPOA rapidly, the registration process will become totally inefficient. Moreover, this mechanism produces a lot of control traffic inside the local domain and across the Internet.

2.2 Micro Mobility and HAWAII

To minimize the movement latencies discussed above, the concept of micro-mobility protocols have been introduced. A micro-mobility protocol operates as follows. The mobile node obtains a local care-of address when it first connects to a domain. This care-of address remains valid while it stays in the same domain and the mobile will thus make only one home registration (registration with the home agent) at the time it connects to the domain. The users movements inside the domain are managed by a micro-mobility protocol. This mobility is transparent to the home agent and the rest of the Internet. Latency and control traffic across the whole network are thus extremely reduced.

HAWAII [4], Handoff-Aware Wireless Access Internet Infrastructure, is a natural extension to Mobile IP to efficiently support micro mobility in wireless networks. After the first connection of a mobile node to a domain and its home registration, the mobile node will perform local registrations only. A common approach for allowing mobility to be transparent to
correspondent hosts is to divide the network into hierarchies. HAWAII uses a similar strategy, segregating the network into a hierarchy of domains, loosely modeled on the autonomous system hierarchy used in the Internet. The gateway to each domain is called the domain root router. Each mobile node is assumed to have an IP address and to have a home domain to where it belongs. While moving in its home domain, the mobile node retains its IP address. Packets destined to the mobile node reach the domain root router based on the subnet address of the domain and are then forwarded over special dynamically established paths to the mobile node. When the mobile node moves into a foreign domain, HAWAII reverts to traditional Mobile IP mechanisms. If the foreign domain is also based on HAWAII, the mobile node is provided with a care-of address from the foreign domain. While moving within the foreign domain, the mobile host retains its care-of address unchanged, and connectivity is maintained using dynamically established paths.

A mobile host that first powers up and attaches to a domain sends a Mobile IP registration request to the nearest base station. The base station is sometimes also called the access router, as it also has routing capabilities in addition to providing fixed network access. The base station is responsible for exchanging Mobile IP messages with the mobile host’s home agent, in order to register the current location of the mobile host. The base station also sends a path setup message to the domain root router, which is the gateway between the micro-mobility access network and the Internet. This has the effect of establishing a host specific route for the mobile host in the domain root router. Each intermediate router on the path between the base station and the domain root router also adds a forwarding entry for the mobile node, when forwarding the path setup message. Thus, the connectivity from the domain root router to the mobile hosts connected through it forms a virtual tree overlay.

The mobile node infrequently sends periodic registration renewal messages to the base station to which it is currently attached in order to maintain the registration and the host-based entries, failing which they will be removed by the base station. The base station and the intermediate routers, in turn, send periodic aggregate hop-by-hop refresh messages towards the domain root router.

2.3 AODV

The Ad hoc On-Demand Distance Vector (AODV) routing protocol [5] is a reactive protocol designed for use in ad hoc mobile networks. AODV initiates route discovery whenever a source needs a route, and maintains this route as long as it is needed by the source. Each node also maintains a monotonically increasing sequence number that is incremented whenever there is a change in the local connectivity information for the node. These sequence numbers ensure that the routes are loop-free.

2.3.1 Route Discovery

Route discovery follows a Route Request (RREQ)/Route Reply (RREP) query mechanism. In order to obtain a route to another node, the source node broadcasts a RREQ packet across the network, and then sets a timer to wait for the reception of a reply. The RREQ packet contains the IP address of the destination node, the sequence number of the source node as well as the last known sequence number of the destination. Nodes receiving the RREQ can respond
if they are either the destination, or if they have an unexpired route to the destination whose corresponding sequence number is at least as great as that contained in the RREQ. If these conditions are met, a node responds by unicasting a RREP back to the source node. If not, the node rebroadcasts the RREQ. In order to create a reverse route from the destination back to the source node, each node forwarding a RREQ also create a reverse route entry for the source route in its routing table.

As intermediate nodes forward the RREP towards the source node, they create a forward route entry for the destination in their routing tables, before transmitting the RREP to the next hop. Once the source node receives a RREP, it can begin using the route to send data packets.

If the source node does not receive a RREP before the timer expires, it rebroadcasts the RREQ with a higher time-to-live (TTL) value. It attempts this discovery up to some maximum number of attempts, after which the session is aborted.

2.3.2 Route Maintenance

Nodes monitor the link status to the next hops along active routes. When a link break is detected along an active route, the node issues a Route Error (RERR) packet. An active route is a route that has recently been used to send data packets. The RERR message contains a list of each destination that has become unreachable due to the link break. It also contains the last known sequence number for each listed destination, incremented by one.

When a neighboring node receives the message, it expires any routes to the listed destinations that use the source of the RERR message as the next hop. Then, if the node has a record of one or more nodes that route through it to reach the destination, it rebroadcasts the message.

3 Mobile Ad hoc Internet Access Solution

In this solution, base stations that are also acting as Home and Foreign agents advertise their services by periodically sending out Agent Advertisement messages. These messages are broadcasted to the wireless ad hoc network, and its dissemination is limited by specifying the TTL to an appropriate value that depends on the size of the network. When a currently unregistered mobile node receives an advertisement, the mobile node unicasts a Registration Request to the sending base station. The base station will reply to this message by sending a Registration Reply back to the mobile node. If this is the first registration sent by the mobile node inside this domain, HAWAII, the micro-mobility protocol, will send path setup power-up messages in order to establish a routing path within the domain hierarchy towards the mobile node. The mobile node now also attains its care-of address, which the base station registers with the home agent. Note that the mobile node will retain this care-of address throughout its stay in the current domain.

Packets between the home agent and the mobile node are routed towards the wireless network based on the network id part of the care-of address. The domain root router of the HAWAII domain is the root of the access network. It is also the gateway router between the local domain and the Internet, owning the network id. As the mobile node moves within the ad hoc network, from base station to base station, it will continue to be accessible from the Internet; only the local path within the lower hierarchy of the domain will be updated.
3.1 Internet Host Determination

When an on-demand routing protocol, such as AODV, is used within an ad hoc network, a node cannot expect to have routes to all hosts reachable within the network. This is because routes are only set up when they are needed. The fact that we do not have a host route to a host does not necessarily mean that it is not reachable within the ad hoc network. Thus, the route discovery mechanism of the routing protocol has to search for the destination within the ad hoc network, before it can decide whether the destination node is located in the network or not. Because the route discovery process of AODV repeatedly searches for the destination within an increasing radius, the time it takes for AODV to determine that the destination is unreachable is quite significant. This problem has been solved in our solution by letting the base stations send proxy route replies. When a base station receives a route request from one of its registered nodes, it searches its registration list (also called visitor list within the Mobile IP terminology), for a match with the requested destination. If a match is found, a normal route reply is generated. If a match is not found, a special proxy route reply indicated by an 'I'-flag, will be generated. This proxy route reply will also establish a route path between the requesting node and the requested destination. Note, however, that the ad hoc network may span several base stations, and therefore include nodes registered with other base stations. In order to let a direct route prevail, the proxy route reply will indicate a high hop count.

3.2 Handover

When a mobile node receives an agent advertisement from a base station that is closer than the one it is currently registered with, or if the old registration timed out, it initiates a handover. This is done sending a registration request to the new base station, that also includes information about the previous base station. When the new base station receives this message it replies by sending a registration reply as normal. The HAWAII part of the new base station also sends path update messages to the local micro-mobility domain and a handover notification is sent to the old base station. The old base station thus removes the mobile node from its registration list and updates its routing table accordingly.

The mobile node will now be reachable from the Internet through the registration in the new base station. The route within the ad hoc network, will however, point towards the old base station. A possible solution to this problem could be to let AODV send a route error message that deletes the route to the old base station, and then send a new route request as described above. The route error is needed because an intermediate node on the old route might otherwise reply to the request. This could, however, disrupt an active transport layer connection, something that we would like to avoid. In this solution, each mobile node instead has a list of destinations located on the Internet that it is currently communicating with, i.e., destinations learned of through the reception of route replies with an 'I'-flag. The mobile node parses this list of destinations, and sends a route request with the 'D'-flag specified, for each of these destinations. The 'D'-flag specifies that only the destination node may reply to the request, assuring that the message will propagate all the way to the new base station without any intermediate nodes replying to the request.
4 Performance Simulations

This paper aims to investigate the performance of micro-mobility movement in a hybrid ad hoc network as described above.

The presented solution has been evaluated in the popular network simulator, ns-2 [6]. The ns-2 simulator is a discrete event simulator widely used in the networking research community. It was developed at the University of California at Berkeley and extended at Carnegie Mellon University to simulate wireless networks. These extensions provide a detailed model of the physical and link layer behavior of a wireless network and allow arbitrary movement of nodes within the network. In our experiments, the MAC layer is implemented using the default characteristics of the distributed coordination function (DCF) of IEEE 802.11 [1]. The data rate for the simulations is 2 Mbps.

The simulations conducted aim to analyze the performance of TCP flows during an Internet connection and during handover. The current Internet host implementations contain a variety of TCP flavors. In order to investigate the differences between these, various TCP versions have been selected and analyzed. These are TAHOE, RENO, and VEGAS. The impact of mobility and ad hoc routing protocols and the relation between the two during handover have a big impact on the performance. It is also so that the different versions of TCP behave differently in this mobile environment.

The simulated scenario consists mainly of two parts, an access network consisting of base stations connected by wired links, and a wireless ad hoc network, see Figure 1. The access network is also the micro-mobility domain and consists of four base stations connected in a three-level network hierarchy of in total six routers, excluding the base stations. Connected to the top domain router is a correspondent wired host that will be communicating with nodes in the wireless ad hoc network.

Figure 2 shows the throughput of a TCP Vegas connection between the correspondent host in the wired domain and a mobile node in the ad hoc network. The mobile node moves in parallel with the base stations at different mobility speeds, and as it learns of new and closer base stations, it performs handovers. The hop distance between the mobile node and its associated base station varies between two and three, depending on the connectivity of the network. The hop distance is never shorter than two, because the distance between mobile node and the base station is such that, at least one intermediate node is needed for connectivity. The periodicity of base station advertisements is one second.

In Figure 2a we see the throughput when the node is moving at 20 m/s. The node is able to sustain a fairly high throughput, but it drops for a short duration during handovers. One reason for this can be found in the way mobile nodes decide when it is time to perform a handover. When a node learns of a new and closer base station, it switches to it. The mobile node also switches to a new base station if the registration of the old one timed out, and there is a certain latency before a new connection can be established. Another reason is that the next hop link towards the base station in the ad hoc network breaks, and the route repair mechanism of AODV is invoked. This is typically detected through a packet drop. If it was the next hop towards the base station that was broken, a check is performed to see if a handover is needed. The link break and packet drop in combination with TCP’s window behaviour may cause the throughput to momentarily drop to a lower level. Figure 2b, 2c and 2d show the same scenario for speeds at 30, 40, and 50 m/s respectively. We can see here that as the mobility speed increases, the dips frequently become wider and deeper. At 50 m/s it takes about one second.
for the connection to regain its throughput, but only for a short time before a new handover takes place. It should be noted that 50 m/s is a very high speed; it corresponds to 180 km/h.

The fact that the wireless environment itself is unreliable has a big impact on the performance. This can be observed in Figure 3, that illustrates a static scenario between the mobile node and the correspondent host. The distance between the mobile node and its base station is five wireless hops. As can be seen from the figures, the throughput is fairly poor, and is significantly lower than the ones seen in Figure 2, even though those figures refer to a mobile scenario. One of the reasons for this is the exposed node problem [7], which 802.11 does not address. This problem basically means that a node is prevented from transmitting when it is either within the range of a sender but not the receiver, or within the range of the receiver but not the sender. The result here is that the throughput is lowered for each additional hop. Another problem is the window behaviour of TCP. The different figures in Figure 3 all show the behaviour of TCP Tahoe, a fairly aggressive flavor of TCP. When Tahoe is used with a large maximum window size, TCP will start transmitting packets with an exponential increase in the window size for each received acknowledgement. This means that the sender transmits packets in fast order, causing collisions and interference to intermediate wireless nodes, with the result of packets being dropped. TCP will therefore timeout, the window lowered and the packets retransmitted, again in fast order. The same thing will happen again, causing the throughput to oscillate, as seen in Figure 3c.

We can also see a distinct difference in performance between the download and upload scenario in Figure 3a vs. 3c. The is because our network is heterogeneous, and the wired sender has a different sending behaviour than the wireless one. The wired sender will rapidly

Figure 2: TCP throughput (kbps) for different mobility speeds (a) 20 m/s, (b) 30 m/s, (c) 40 m/s, and (d) 50 m/s.
start sending packets, the wired link have a higher bandwidth, and when they reach the base station buffering will take place. Because of the lower bandwidth and the unreliable nature of the wireless channel, packets will be dropped. This can observed in Figure 4a that show TCP segment numbers and ACKs. The throughput of this scenario is the one seen in Figure 3a. Because of the lower bandwidth at the wireless sender in the upload scenario, the same amount of packet drops does not take place, see the corresponding throughput in Figure 3c.

One way to cope with this problem, and to make the transmission process less aggressive, is to lower the maximum congestion window size. When the windows size is lowered from 32 to 4 segments, the difference between the upload and download throughput disappears, see Figure 3b and 3d.

Yet another factor that impact the performance is the size of the packets. Figure 4c show the increase of the segments number for a downlink TCP Reno connection. The fastest segments number increase in this figure is those with a packet size of 512 bytes. The slowest increase is achieved when the packet size is 1460 bytes. The reason for this is quite simple; the longer the transmission of packet takes on the wireless channel, the higher the probability for interference to cause an unsuccessful reception.

Table 1 shows the corresponding throughput for various flavors during upload and download. We can see here that Tahoe achieves the highest throughput during upload for both 20 m/s and 30 m/s mobility. This is because Tahoe accesses the wireless channel more aggressively than Vegas, as was explained above. However, this aggressiveness is less advantageous in the download case where Vegas achieves the highest throughput. It should be noted that no congestion in the normal sense occurs in this scenario, which is the reason why Reno performs

Figure 3: TCP throughput (kbps) for a static upload and download scenario (a) download, (b) download, (c) upload, and (d) upload.
worse than Tahoe.

Another issue with TCP flows in ad hoc networks is unfairness. This can be observed in Figure 5. Here we can see that the ongoing TCP download connection is completely shut down by a short lived local connection. When the local flow terminates, the previous connection can be resumed, but only until another local flow starts. The reason for this is a complicated interaction between the 802.11 MAC layer and TCP that forces the MAC layer into exponential backoff. This is a problem that has been described before [8], and a few different solutions have been proposed. This problem needs to be solved before 802.11-based ad hoc networks can have any real success.

During the course of our investigation, we also observed that the throughput and delay clearly depend on the distance between a mobile node and its corresponding base station. As the number of hops increases, the throughput decreases while the delay increases, see Figure 6. We can see here that the throughput during upload is around 475 kbps when the distance is two hops, but only around 150 kbps for ten hops. The decrease is faster in the beginning and seems to be exponentially declining. The main reason for this is probably the exposed node problem, as nodes are prevented from transmitting because the next hop node is transmitting. The upload throughput is also always higher than during download, as discussed above. The increase in delay seems to be almost linear with the number of hops, at least during upload. For two hops,

<table>
<thead>
<tr>
<th>Throughput (kbps)</th>
<th>Upload</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegas 20 m/s</td>
<td>429.3</td>
<td>391.4</td>
</tr>
<tr>
<td>Vegas 30 m/s</td>
<td>424.0</td>
<td>386.0</td>
</tr>
<tr>
<td>Tahoe 20 m/s</td>
<td>442.0</td>
<td>382.8</td>
</tr>
<tr>
<td>Tahoe 30 m/s</td>
<td>439.8</td>
<td>374.8</td>
</tr>
<tr>
<td>Reno 20 m/s</td>
<td>421.9</td>
<td>370.1</td>
</tr>
<tr>
<td>Reno 30 m/s</td>
<td>423.5</td>
<td>346.6</td>
</tr>
</tbody>
</table>
the delay is around 25 ms, but for ten hops it has been doubled to around 50-60 ms. As each additional hop introduces additional processing time, this makes perfect sense.

5 Related Work

In [9], the authors present a solution that interconnects ad hoc networks with infrastructured networks. For micro mobility, the Cellular IPv6 [10] protocol is utilized on the edge of the Internet. AODV is used as the routing protocol within the ad hoc network. Performance is measured mainly with regards to control overhead and delivery ratio, when the mobility speed is varied.

In MIPMANET [11], the authors integrate AODV with Mobile IP. Their solution utilizes IP tunneling for separating the ad hoc network from Mobile IP. Nodes in the ad hoc network send their packets to a correspondent node in the Internet by encapsulating the packet into another IP packet, which is destined to the Mobile IP foreign agent. A Mobile IP care-of address is used to provide appropriate routing from the Internet to the mobile node.
6 Conclusion

We have in this paper presented a solution for Internet access and micro mobility for ad hoc networks. The solution relies on the AODV routing protocol for establishing multihop paths between a mobile node and a base station. For micro mobility, the solution is based on HAWAII, a domain-based micro-mobility scheme.

The transport layer performance of the proposed solution has been evaluated using simulations. The simulations indicate that a fairly high throughput can be achieved, even during very high mobility speeds. However, the characteristics of the wireless environment itself, as well as inefficiencies of the 802.11 MAC layer protocol, lowers the performance when the number of hops increases. By using a less aggressive version of TCP such as Vegas, or lowering the maximum window size, the throughput can be somewhat increased. The problem with unfairness needs to be solved before multiple TCP flows can be supported.

References


