CHAPTER 5

ANALYTICAL MODEL

As mentioned earlier in this thesis that though the literature reviews on mobility management is handful, the session setup in mobile environment has received very low attention. However all the existing mobility management techniques are directed towards post session setup data transfer. This thesis discusses the mobility management when a session is being setup. This section uses the basic delay of M/M/1 model and some fixed delays to compute end-to-end delay to setup sessions.

A source node or a destination node or both may be mobile while they are participating in a session establishment process. Usually if a terminal is mobile (UE) in IMS session setup, the SIP redirect server which is an application server, informs the originating Corresponding Host (CH) to initiate a new INVITE message destined to a new location of the UE. Also, there are some other issues like channel handoffs while a UE changes location. This chapter compares the existing session set up scenario with PBQMS model and analyzes the pros and cons of them in the 4G networks.

5.1 SCENARIO DESCRIPTION

When a UE moves under a new base station (BS), the delay due to the foreign network identification, by listening to the beacons transmitted from the new BS and the following DHCP address acquisition, can be minimized using link layer mechanisms. Also packets are sent to the new UE location as soon as the INVITE message reaches the CH and gets processed.
So effectively the hand-off delay is the one-way transmission time of the INVITE message to the CH and its subsequent processing time at the CH. The association with the older BS is removed from the UE only when the new association is acknowledged by the CH. However, since SIP is an application layer protocol, the SIP based messages may not be served with highest priority in the associated components and this may introduce additional delay. Thus, when an UE sends an INVITE message to the CH, the CH’s operating system may be busy with its own time critical operating system functionality and queue the INVITE request for deferred processing. Assume that the UE is in a wireless access network, which introduces its own delay due to the additional error recovery protocol layers to circumvent the unreliable nature of the wireless links. This thesis uses the queuing model and a wireless link delay model to estimate the total delay involved with the hand-off.

The UE generates SIP message to update its current location and the messages are sent to the nearest base stations for transmission to the CH or the communicating server through the Internet. Once the CH receives the INVITE message from the UE and if UE moves, the CH gets the updated location information and henceforth sends data to the right location. So the hand-off delay is typically the time required for the INVITE message to reach the CH from the UE. Major delays in this hand-off procedure occur at (i) the UE, (ii) the wireless radio link between the UE and the BS, (iii) the IMS and (iv) CH or the server.

Consider the use case specified in section 4.2. Sam is currently enjoying a coffee seating in a cafe. As an HSDPA subscriber, Sam is registered to his 4G-network provider that hosts an IMS domain for multimedia services. Sam’s HSDPA PCMCIA-based data card is used as a modem for the notebook, which is a multimode device that also includes a WLAN capability. Since, Sam has set preferences for secure connection, the
PBQMS module in the notebook detects all available networks (HSDPA and WLAN hot-spot) and automatically selects the HSDPA access. After registration, Sam starts a video-conferencing application with a friend Rama. After a while, Sam gets a notice from his notebook of a pending appointment at home. While on the way to home, Sam continues to maintain the video-conferencing session. Since Sam has set the preferences of using WLAN at home, the moment Sam reaches home, the PBQMS module detects the availability of the WLAN and conducts the handover to the WLAN. Because IMS manages the handover process to SIP session transparently and PBQMS governs the process optimally, Sam doesn't notice any service interruption. In this case the PBQMS take care of doing all the hand-over process while the Sam is on HSPDA network. In this case, the hand-over delay (DPBQMS) could be only the Internet delay.

5.2 EXISTING SESSION SETUP SCENARIO

This thesis modeled each of these delay components as a queue with the exception of the Internet transmission delay. The delay introduced by the Internet depends on the number of routers and the type of links in the path of datagram transmission. It is rather difficult to standardize such heterogeneous transmission paths and compute the transmission delay. For this reason, assume that the Internet delay is a constant and is equal to the typical worst-case delay for SIP messages as reported (Eyers et al 2000).

This thesis assumes an M/M/1 queuing model for the UE and the base station, and a priority based M/G/1 model for the CH or the server. The queuing model for the delay incurred in hand-off is shown in Figure 5.1. Table 5.1 presents the list of parameters used in the analysis and their corresponding symbols.
Assuming that there are many UEs under a single base station, so \( \lambda_M \ll \lambda \), and \( \lambda_M \) is a fraction of \( \lambda \). The handoff delay \( D_{HO,\text{EXIST}} \) in transmitting a SIP message can be computed as:

\[
D_{HO,\text{EXIST}} = D_1 + D_2 + D_3 + D_4 + D_5
\]  

where, \( D_1 \) is the delay at the UE, \( D_2 \) is the delay incurred in transmitting the SIP message over the wireless link, \( D_3 \) is the queuing delay in the IMS, \( D_4 \) is the constant Internet transmission delay and \( D_5 \) is the queuing delay in the CH.

![Diagram showing the delay components and queuing model involved in handoff delay]

**Figure 5.1** Delay components and queuing model involved in handoff delay
### Table 5.1 List of system parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>Delay at UE (in millisecond)</td>
</tr>
<tr>
<td>$D_2$</td>
<td>Delay in transmitting the SIP message over the wireless link (in millisecond)</td>
</tr>
<tr>
<td>$D_3$</td>
<td>Queuing delay in the IMS (in millisecond)</td>
</tr>
<tr>
<td>$D_4$</td>
<td>Constant Internet transmission delay (in millisecond)</td>
</tr>
<tr>
<td>$D_5$</td>
<td>Queuing delay in the CH (in millisecond)</td>
</tr>
<tr>
<td>$\lambda_M$</td>
<td>SIP message arrival rate at the UE (in millisecond)</td>
</tr>
<tr>
<td>$\lambda_s$</td>
<td>SIP message arrival rate at the CH (in millisecond)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>SIP message arrival rate at the IMS (in millisecond)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Processing delay for each SIP message in the UE and IMS (in millisecond)</td>
</tr>
<tr>
<td>$\mu_s$</td>
<td>Processing delay for each SIP message in the CH (in millisecond)</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>Processing delay for at CH for messages other than SIP (in millisecond)</td>
</tr>
<tr>
<td>$\rho_s$</td>
<td>Load at CH for SIP messages</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>Load at CH for messages other than SIP</td>
</tr>
<tr>
<td>$\Delta_l$</td>
<td>Fixed Internet delay in transmission of SIP messages</td>
</tr>
<tr>
<td>$N_m$</td>
<td>Number of TCP transmissions before a successful transmission</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Inter frame time.</td>
</tr>
<tr>
<td>$R$</td>
<td>Items reside in the queuing system</td>
</tr>
</tbody>
</table>
Using the results from queuing theory (Kleinrock 1975) the delay parameters are estimated as follows:

\[
D_1 = \frac{1}{\mu - \lambda_M} \tag{5.2}
\]

\[
D_3 = \frac{1}{\mu - \lambda} \tag{5.3}
\]

\[
D_4 = \Delta_1 \tag{5.4}
\]

\[
D_5 = \frac{\frac{1}{\mu_s} (1 - \rho_s - \rho_s) + R}{(1 - \rho_s)(1 - \rho_s - \rho_s)} \tag{5.5}
\]

Assuming that the SIP message arrival rate and the processing rate at the base station and the CH are the same (i.e. \( \mu_s = \mu \)) and \( \lambda_M = 0.1 \lambda \). The derivation of \( D_5 \) is a little bit tricky, since it involves the second moment of the processing delay at the CH. The second moment can be derived using mean and the variance. For analysis, this thesis assumes that the standard deviation of the processing rates at the CH is 5% of the mean. Now \( X_1^2 = E[X_1^2] \) and \( X_2^2 = E[X_2^2] \). Also \( E[X_1^2] = \sigma_1^2 + (E[X_1])^2 \) and \( E[X_2^2] = \sigma_2^2 + (E[X_2])^2 \), where \( \sigma_1^2 \) and \( \sigma_2^2 \) are the respective variances. Substituting \( \mu_s \) and \( \mu_1 \) for \( E[X_1] \) and \( E[X_s] \) and the values for the variances, where \( R = 0.501[\mu_s^2 + \mu_1^2] \) The expression for \( D_5 \) is obtained by using the result of a non-preemptive priority-based M/G/1 queue. Since the objective is to estimate the SIP message processing delay, the CH considers only the messages having higher priority than SIP messages and neglected other lower priority messages.

Derivation of the term \( D_2 \) requires adopting a delay model over wireless links. Since SIP messages can be sent using both TCP and UDP and TCP being a reliable protocol and dealing with wireless links, assuming that
the SIP messages are sent using TCP. So there is a need for a delay model for TCP transmission over wireless links. However, wireless links being unreliable in nature it requires an underlying recovery mechanism for reliable transmission. Though CP has its own recovery mechanism, TCP being an end-to-end protocol, its error recovery mechanism is not appropriate for real-time transmission. This is because end-to-end retransmission is not recommended for real-time applications to avoid delay variance. In this case, semi-reliable link layer retransmission mechanism like Radio Link Protocol (RLP) is used to reduce the air link FER and thus increase reliability. RLP works on the basis of a NAK based acknowledgment scheme.

According to the model used reported in Das et al (2002), the delay to transmit a TCP segment consisting of k frames over a radio link without RLP operating on it, is given by

\[
D_2 = \sum_{i=1}^{N_m} \left[ (k-1)\tau + \frac{D}{(1-q^{N_m})(1-2q)} \right]
\]

\[
+ \frac{1-q}{1-q^{N_m}} D \left[ \frac{q^{N_m}}{1-q} - \frac{2^{N_m}}{1-2q} \right]
\]

where, \(N_m\) is the number of TCP retransmissions before a successful transmission, \(\tau\) is the inter-frame time, \(D\) is the end-to-end frame propagation delay over the radio channel, \(q = 1 - (1-p)^k\) is the packet loss rate, and \(p\) is the probability of a frame being in error in the air link. Following Das et al (2002), typical values are \(D = 100\) msec and \(\tau = 20\) msec.

Equation (5.6) expresses the delay for the case without considering the RLP support in the wireless links. When RLP is used to reduce the TCP
retransmission overhead, the delay component $D_2$ will no longer remain the same. Instead it is given as:

$$D_2 = D + (k - 1)\tau + \frac{k(P_i - (1 - p))}{p_i^2}$$

$$\times \left(\sum_{j=1}^{n} \sum_{i=1}^{j} P(C_{ij})(2jD + \left(\frac{j(j+1)}{2} + i\tau\right)) + \frac{2D_q(1 - q)}{1 - q^{N_q}}\right)$$

$$\left[1 + \frac{4q(1 - (2q)^{N_q-2})}{1 - 2q} - \frac{q(1 - q^{N_q-2})}{1 - q}\right]$$

where, $n=3$ is the maximum number of RLP retransmission trials;

$$P_i = 1 - p + \sum_{j=1}^{n} \sum_{i=1}^{j} P(C_{ij}) = 1 - p(p(2 - p))^\frac{n(n+1)}{2}$$

and $C_{i,j}$ the first frame received correctly at the destination, is the $i$th retransmission frame at the $j$th retransmission trial.

To derive the value of $k$, assume that a TCP segment is carried in one packet and assume that the air link frame duration is 20msec. Therefore, a 9.6 Kbps radio channel contains $9.6 \times 10^3 \times 20 \times 10^{-3} \times 1/8 = 24$ bytes in each frame. Also assume that the size of a SIP message is 500 bytes. Therefore, number of air link frames in a SIP message is $500/24 = 21$. For 19.2 and 128 Kbps channels the number of frames are $k=11$ and $k=2$ respectively.

5.3 **PBQMS SESSION SETUP SCENARIO**

Consider the use case specified in section 4.2. IMS manages the handover process (to SIP session) transparently and PBQMS governs the process optimally, Sam doesn't notice any service interruption. In this case
the PBQMS takes care of doing all the hand over process while Sam is on HSPDA network. In this case the worst cast handover delay (\(D_{PBQMS}\)) will be equal to that of the Internet delay.

\[
D_{PBQMS} = D_4
\]  \hspace{1cm} (5.9)

### 5.4 NUMERICAL RESULTS

Due to the varying nature of the Internet delay and the computing power of the intermediate servers, it is difficult to compute the exact the end-to-end handoff delay. With proper traffic engineering, the Internet delay can be made to suit the application requirements. Hence this thesis focuses on the component of the handoff delay introduced due to the wireless access networks to get an estimate of the minimum handoff delay. Subsequently this thesis estimates the end-to-end handoff delay by assuming a constant value for the Internet delay and some representative values for the computing capabilities of the servers as shown in Table 5.2.

#### Table 5.2 System parameters values

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>(4 \times 10^{-4}) sec</td>
</tr>
<tr>
<td>(\rho_s)</td>
<td>(\lambda/\mu) ((\lambda &lt; \mu))</td>
</tr>
<tr>
<td>(\rho_1)</td>
<td>0.7</td>
</tr>
<tr>
<td>(\Delta_t)</td>
<td>200 msec</td>
</tr>
<tr>
<td>(N_m)</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 5.2 shows the increase in the handoff delay component due to the wireless access, when the FER is increased for channel bandwidth of
2Mbps and 11 Mbps. Here the UE moves from a UMTS network to a WLAN network. Table 5.3 shows the corresponding end-to-end handoff delay including the queuing delay at different servers and the transmission delay over the Internet.

Figure 5.3 shows the increase in the handoff delay component due to the wireless access, when the FER is increased for channel bandwidth of 9.6, 19.2 and 128 kbps. Here the UE moves between UMTS networks. Table 5.4 shows the corresponding end-to-end handoff delay including the queuing delay at different servers and the transmission delay over the Internet.

It is observed from Figure 5.2 and Figure 5.3 that the end-to-end handoff delay using PBQMS for a 128 kbps GPRS radio access of a UMTS network is 1.404818 s, when the channel FER is 0.05 and the SIP-based multimedia session arrival rate is 50 messages/second. Whereas for the 11 Mbps WLAN, the handoff delay is only 0.267 ms. The end-to-end handoff delay got reduced to that of Internet delay, because of PBQMS which uses ‘make before break connection’ technique. As mentioned earlier, to ensure QoS for streaming multimedia the maximum handoff delay should be ideally less than 200 ms. This requirement cannot be satisfied for a UMTS network even with a channel data rate of 128 kbps.

**Table 5.3**  Handoff delay components when UE Moves from a UMTS network to WLAN

<table>
<thead>
<tr>
<th>Channel FER</th>
<th>Processing Delay</th>
<th>Wireless Delay 2Mbps</th>
<th>Wireless Delay 11Mbps</th>
<th>Total Delay in 2Mbps</th>
<th>Total Delay in 11Mbps</th>
<th>Total delay using PBQMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0100</td>
<td>0.2019</td>
<td>0.0014</td>
<td>0.0002</td>
<td>0.2033</td>
<td>0.2021</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0200</td>
<td>0.2019</td>
<td>0.0014</td>
<td>0.0003</td>
<td>0.2033</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>0.0300</td>
<td>0.2019</td>
<td>0.0014</td>
<td>0.0003</td>
<td>0.2033</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0400</td>
<td>0.2019</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.2034</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0500</td>
<td>0.2019</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.2034</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0600</td>
<td>0.2019</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.2034</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0700</td>
<td>0.2019</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.2034</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0800</td>
<td>0.2019</td>
<td>0.0016</td>
<td>0.0003</td>
<td>0.2035</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0900</td>
<td>0.2019</td>
<td>0.0016</td>
<td>0.0003</td>
<td>0.2035</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.1000</td>
<td>0.2019</td>
<td>0.0017</td>
<td>0.0003</td>
<td>0.2036</td>
<td>0.2022</td>
<td>0.1000</td>
</tr>
</tbody>
</table>
Table 5.4  Handoff delay components when UE moves between UMTS networks

<table>
<thead>
<tr>
<th>Channel FER</th>
<th>Processing Delay</th>
<th>Total Delay in 9.6 kbps</th>
<th>Wireless Delay in 19.2 Kbps</th>
<th>Wireless Delay 128 in Kbps</th>
<th>Wireless Delay in 9.6 Kbps</th>
<th>Total Delay in 19.2kbps</th>
<th>Total Delay in 128kbps</th>
<th>Total Delay using PBQMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0100</td>
<td>0.2144</td>
<td>3.8344</td>
<td>2.3800</td>
<td>1.4000</td>
<td>3.6200</td>
<td>2.5944</td>
<td>1.6144</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0200</td>
<td>0.2144</td>
<td>3.8354</td>
<td>2.3805</td>
<td>1.4001</td>
<td>3.6210</td>
<td>2.5949</td>
<td>1.6145</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0300</td>
<td>0.2144</td>
<td>3.8392</td>
<td>2.3825</td>
<td>1.4006</td>
<td>3.6248</td>
<td>2.5969</td>
<td>1.6150</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0400</td>
<td>0.2144</td>
<td>3.8491</td>
<td>2.3876</td>
<td>1.4019</td>
<td>3.6347</td>
<td>2.6020</td>
<td>1.6163</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0500</td>
<td>0.2144</td>
<td>3.8699</td>
<td>2.3984</td>
<td>1.4048</td>
<td>3.6555</td>
<td>2.6128</td>
<td>1.6192</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0600</td>
<td>0.2144</td>
<td>3.9072</td>
<td>2.4178</td>
<td>1.4099</td>
<td>3.6928</td>
<td>2.6322</td>
<td>1.6243</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0700</td>
<td>0.2144</td>
<td>3.9690</td>
<td>2.4499</td>
<td>1.4184</td>
<td>3.7546</td>
<td>2.6643</td>
<td>1.6328</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0800</td>
<td>0.2144</td>
<td>4.0642</td>
<td>2.4993</td>
<td>1.4315</td>
<td>3.8498</td>
<td>2.7137</td>
<td>1.6459</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.0900</td>
<td>0.2144</td>
<td>4.2036</td>
<td>2.5717</td>
<td>1.4507</td>
<td>3.9892</td>
<td>2.7861</td>
<td>1.6651</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.1000</td>
<td>0.2144</td>
<td>4.3997</td>
<td>2.6366</td>
<td>1.4777</td>
<td>4.1853</td>
<td>2.8510</td>
<td>1.6921</td>
<td>0.1000</td>
</tr>
</tbody>
</table>

Figure 5.2  Total Handoff delay Vs Channel FER when UE moves from UMTS network to WLAN network
Figure 5.3 Total Handoff delay Vs Channel FER when UE moves between UMTS networks

5.5 SUMMARY

This chapter presents an analysis of delay incurred in the PBQMS based mobility management for a mobile host in a wireless network. From the results, the following observations are made:

- SIP message processing delay due to PBQMS reaches nil because of using the technique called ‘make before break connection’. In other words, end-to-end handoff delay reaches to that of Internet delay
- Message arrival rate does not affect the session set up delay
- Increase in channel bandwidth decreases the delay for session setup and handover

Media streams can function normally with a maximum interruption of 50 msec, while an interruption of 200 msec is generally acceptable. High speed
data access (in the order of Mbps) is promised in emerging 4G networks using the proposed PBQMS. Hence, it is observed that the PBQMS based session setup and mobility management is suitable for media streaming in 4G wireless networks.