

HANDOVER LATENCIES IN BCMP NETWORKS

In this article, we give a short overview on wireless access networks and the BRAIN Candidate Mobility Protocol (BCMP). The two different handover mechanisms of BCMP are introduced. With the help of standard network tools and an own measurement program, latencies caused by both kinds of handovers were examined. We show that latency values are largely hardware-dependent and that by an adequate hardware configuration, BCMP provides smooth and fast handovers. This makes BCMP suitable for using with real-time applications, such as packet-switched voice or video.

1. Introduction

Nowadays, wireless local area networks are of increasing importance. They provide comfortable network connectivity in such places, where cables are not affordable or disturbing. In the past few years, the evolution of radio technology made high bandwidth links available so that they are also suitable for audio and video applications.

Driven by the increasing demand for mobile and cordless broadband services, the BRAIN (Broadband Radio Access for IP based Networks) project [1] was to provide a true broadband multimedia IP-based radio technology. In addition, BRAIN offers the integration of end-to-end services over IP and evolves IP towards mobility. The project MIND [2, 3] (Mobile IP-based Network Developments) has been formed to research the extension of IP-based radio access networks (ANs) to include ad-hoc and wireless elements both within and attached to the fixed infrastructure. MIND is a follow up to the successful BRAIN project. The project will take an IP core as a starting point, accessed by a variety of technologies. They put the focus on the access network, where different IP QoS protocols could run and where IP micro-mobility management would be introduced.

One of the most important requirements in wireless access networks is to provide fast and smooth handovers because of real-time applications. There are several ongoing research projects to lessen handover latency, for example, BETH [4] is an extension to the Fast Mobile IPv6 handover protocol, which reduces layer 3 latency to zero. There is also a need to lower the handover latency caused by the link layer of IEEE 802.11 [5], to make it VoIP-enabled.

As a large umbrella project, BRAIN/MIND includes the development of a number of IP-based technologies. One of them is BCMP (BRAIN Candidate Mobility Protocol) an IP micro-mobility solution. As it was mentioned earlier, on the way to voice and other real-time applications, latency is a key point. That is why we

conducted an experiment-based analysis of latency caused by different kinds of handovers (HO) in a BCMP-enabled access network.

2 A Short Overview on Access Networks and BCMP

For a better understanding of the measurement results, a brief description of the participating entities in an AN is needed [6]. These are the following:

- *Mobile Hosts (MHs)* are computers which can move from a location to another one and from a network to another one. They have to be supported by IP-mobility enabled routers.
- *BCMP Access Routers (ARs)* are situated at the border of the AN and offer IP-connectivity to the MHs. They act as the default router to the served MHs.
- *Anchor Points (ANPs)* are located inside the AN at selected positions. They own and allocate IP addresses, authenticate users, maintain user records and tunnel packets towards MHs.
- *Access Network Gateways (ANGs)* form an explicit boundary of the AN. They work as standard border routers, so they do not need to implement mobility functions.
- *Internal Routers (IRs)* are supporting routers between an ANP and ARs.

Mobility management in the access network is considered as a single problem in most cases, but the BRAIN concept treats it as several separate issues. According to this idea there are three main responsibilities of an IP-mobility solution.

Packet forwarding and Path Updates: this refers to the mechanism for installing information in the interior of the AN so that packets can be successfully delivered to the MH at its new AR. It has to be well scalable (to support large networks), robust (to have a quick recovery process from failures) and it should allow multiple gateways (for improved scalability and reliability).

Handover Management: this refers to the impact of handovers on the MH. It deals with the local signalling involving the MH and

* Gergely Biczók, Kristóf Fodor, Balázs Kovács

High Speed Networks Laboratory, Department of Telecommunications and Mediainformatics, Budapest University of Technology and Economics, E-mail: biczok@tmit.bme.hu

the ARs to facilitate re-attachment to a new AR. The primary goal here is to minimize packet loss and delay during a handover.

Support for Idle Mobile Hosts: an idle mobile host (i.e. one not actively involved in data transmission) can lessen its signalling messages over the air and save on its terminal power. Its location is tracked through a combination of paging and location updates, which reduces router states in the AN.

From the article's point, the most interesting one among these functions is handover management. BCMP supports two basic kinds of handoffs [2].

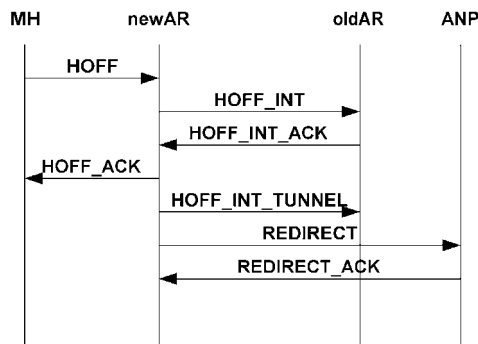


Fig. 1 Non-planned handover

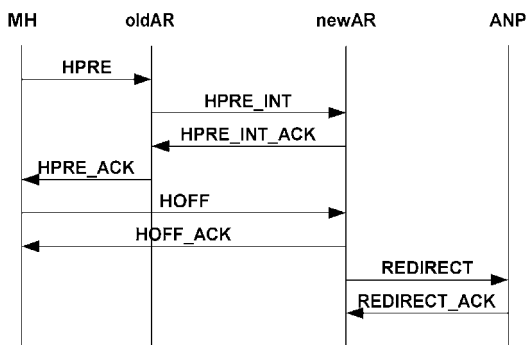


Fig. 2 Planned handover

Regular handoff (non-planned): if the MH wants to hand off to a new AR, but it has no contact to its current AR, then the procedure that can be seen in Fig. 1 takes place.

Prepared handoff (planned): this handoff occurs if the MH knows in advance which AR it is moving to, and it is still in contact with its current AR. In this case, first, the new AR processes the context data sent by the old AR, and if the new AR has sufficient resources, then a temporary tunnel is constructed by the old AR towards the new AR, which stores the incoming packets in its buffer. After that, the process is quite similar to the one in the previous case (Fig. 2).

As it can be expected, the prepared handoff has better performance in the field of packet loss, but latency – which is a key

factor in real-time applications – is lower at non-planned handovers.

3. Design of the Experiment

The experiment was done in two phases at the BCMP testbed of the High Speed Networks Laboratory (HSNLab) at the Department of Telecommunication and Telematics, Budapest University of Technology and Economics. At the first stage, the network consisted of an ANG, an ANP, an IR and two ARs. The ARs' configuration was: Pentium 133Mhz processor, 32 MB of RAM, but AR2 had 64 MB. The ANG was also built upon P133 with 32 MB RAM. The IR and the ANP had an AMD Duron 700Mhz processor with 128MB RAM. The AN was linked with 100Mbit/s Ethernet connections. The wireless interface of the ARs and the MH was type 802.11b, up to 11Mbit/s maximum throughput. For the MH we used a notebook with Pentium III 800 MHz processor and 128MB RAM. All of the computers were equipped with the same Linux distribution and kernel.

The ARs were quite close to each other so a program initiated the different types of handovers. The test bed used MAC filtering which disables packets arriving from the old AR of a MH. The experiment was logged on the MH.

During our experiment we measured packet inter-arrival times to analyze handover outage effects. We used two methods: `tcpdump` and an own test program. These programs let us show both the low and the user level handover outage impacts.

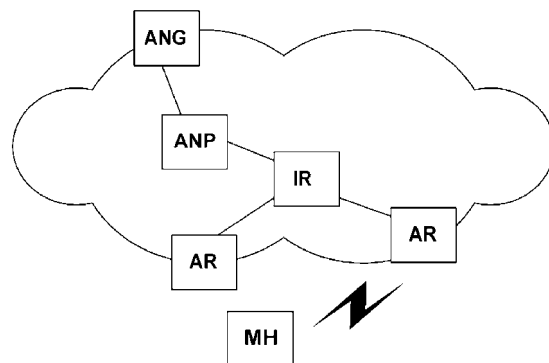


Fig. 3 Topology of the test bed

Our program consists of a packet sender and a packet receiver. The packet sender emits 65-byte-long UDP packets at a defined rate. We used the UDP transport protocol to avoid TCP's traffic control mechanisms. The receiver logs the sequence number of the packet arrived, the inter-arrival times, packet loss and if there was exchange in the packet sequence.

At the second stage, we had an opportunity to use a test bed having more powerful machines. All the fixed computers (ANG, ANP, ARs) had AMD Athlon 1800+ and 512MB of system memory,

the MH was a notebook equipped with a 1800 MHz Pentium 4-M processor and 256 MB RAM.

4. Stage 1: Results and Analysis

As it was mentioned earlier, our measurements were based on two kinds of handovers. Fig. 4 shows the planned handover latencies. We can see two mean values changing (60ms and 100ms). This is due to the different AR hardware configuration. If we do not take this into consideration we can say that the outage time caused by handovers is quite deterministic. In Fig. 5, it can be seen that the packet-sending interval is 10 ms (as we set it), the handover outage time - which includes the creation of the tunnel as well - hits nearly 60ms, but after the preparation phase the queued packets are re-sent towards the MH through the tunnel at maximum performance. After the "equalization" of the impact of the delayed packet, arriving intervals are set back to normal. The results of the measurement show that packet loss is close to zero in the case of planned handovers, except for very high sending rates. Exchanges in the packet sequence are also more common at smaller inter-departure times.

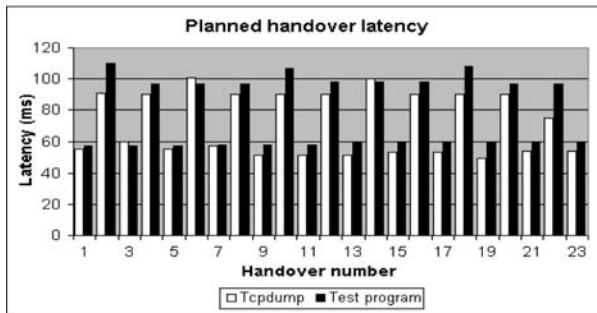


Fig. 4 Planned HO latency

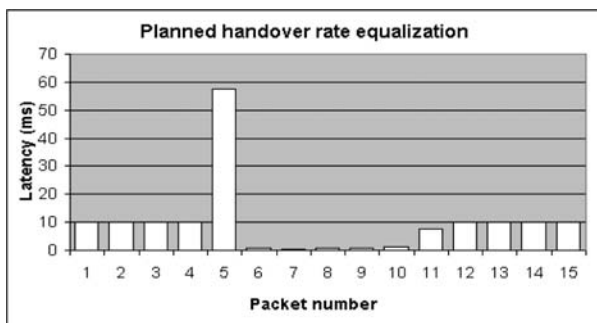


Fig. 5 Planned HO rate equalization

In the case of non-planned handovers, outage times are a little bit lower due to a lack of the preparation phase and the construction of a temporary tunnel, but accordingly, it has disadvantages in the quality of transmission. We can see handover latency values, oscillating between 60 and 75 ms (Fig. 6), instead of the 80 ms average of a planned handover. However, upon inter-arrival times

caused by non-planned HOs, we cannot draw the interfere that for how long exactly the HOs last. In order to do that, checking packet types in `tcpdump` output are necessary. Fig. 7 shows the time values elapsed from the HOFF message to the HOFF_ACK. (In the planned case, latency and real HO duration are fairly equal.) In the non-planned case the presence of packet loss is quite stationary and changes in the packet sequence are more frequent than in the planned case.

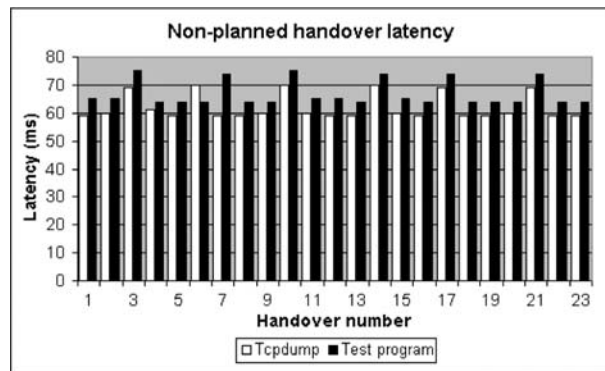


Fig. 6 Non-planned HO latency

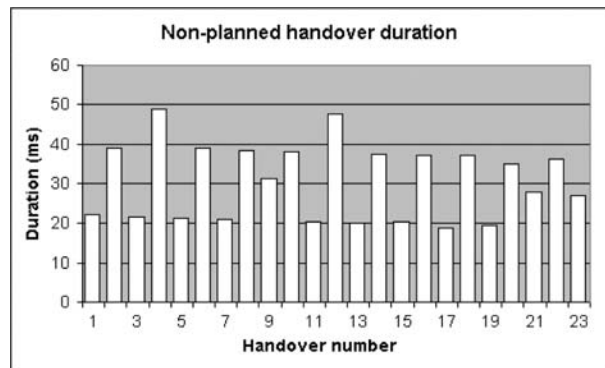


Fig. 7 Exact duration of non-planned HO

5. Stage 2: Results and Analysis

The second batch of measurements was conducted in order to observe if serious hardware upgrades could result in a significant drop of handover latency values. Meanwhile, a new measurement technique was designed to achieve better efficiency. As we knew the frequency of handovers and the packet-sending interval in advance, we could well localize the first arriving data packet in the data flow after a handover. In this way, we do not take "pseudo-handovers" into consideration (when packets are delayed due to some other reason). This method enhances the quality of the experiment. Since the aim of the second stage was to show the difference in latencies caused by new hardware, only handover outage times were logged with `tcpdump`. A packet is sent every 10 ms, and handovers are performed with a period of 2 seconds (as in the first stage).

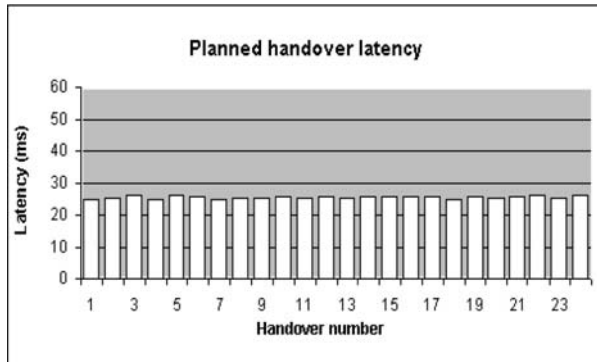


Fig. 8 Planned HO latency (Stage 2)

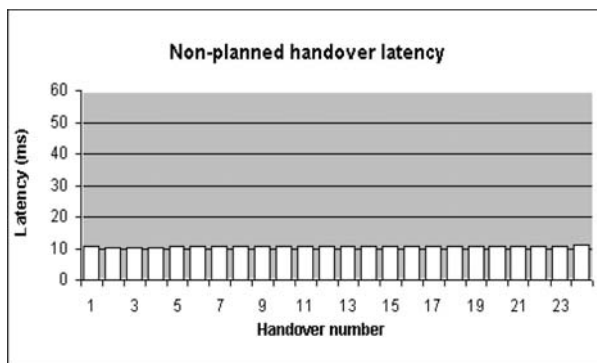


Fig. 9 Non-planned HO latency (Stage 2)

Fig. 8 shows planned handover latency values. As it follows from the upgraded hardware resources, values are around 25 ms: a significant reduction occurred (see Fig. 4 for comparison). As we expected, BCMP performed even better at the second stage.

There is some improvement in the case of non-planned handovers, as well (Fig. 9). The average latency is around 10.7 ms, which is only 0.7 ms more than the normal packet inter-arrival time (i.e. without switching Access Routers). Although non-planned handover latencies are smaller than that of planned handover, packet loss may occur. For this very reason, planned handover is preferred, since its performance is more than enough for real-life applications, such as packet-switched voice data or streaming video.

6. Conclusion and Future Work

It can be recognized that more powerful hardware has an advantageous effect on BCMP handover duration and data outage times. This explains the variation of values in the diagrams. Fast and smooth handovers are crucial for ensuring micro-mobility, while transcoding e.g. packet-switched voice data (a maximum delay of 50 ms is acceptable). According to the results, it seems that with the support of adequate hardware infrastructure BCMP is able to make the mobile hosts capable of sending and receiving real-time data. In the future, we plan to examine more complex access networks, which are equipped with more access routers, furthermore, to investigate the effects of simultaneous handovers caused by the presence of multiple mobile hosts.

References

- [1] BRAIN HOME PAGE: www.ist-brain.org
- [2] IST MIND: *Core Report*, November 2002
- [3] MIND HOME PAGE: www.ist-mind.org
- [4] KEMPF, J. et al.: *Bidirectional Edge Tunnel Handover for IPv6*, Internet Draft, September 2001.
- [5] MISHRA, A., SHIN, M., ARBAUGH, W.: *An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process*, ACM CCR, 2002.
- [6] KESZEL, CS., GEORGANOPOULOS, N., TURÁNYI, Z., VALKÓ, A.: *Evaluation of the BRAIN Candidate Mobility Management Protocol*, IST Global Summit, Barcelona, 2001.