ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Journal of Network and Computer Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Journal of NETWORK and COMPUTER APPLICATIONS

www.elsevier.com/locate/jnca

SIP-based proactive and adaptive mobility management framework for heterogeneous networks

Y.C. Yee[a], K.N. Choong[b,*], Andy L.Y. Low[a], S.W. Tan[a]

[a]Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
[b]British Telecommunications plc., Asian Research Centre, 1B-17, Block 1B, Plaza Sentral, Jalan Stesen Sentral 5, 50470 Kuala Lumpur, Malaysia

Received 3 April 2007; received in revised form 5 September 2007; accepted 5 September 2007

Abstract

In this paper, we present and evaluate the performance of a mobility management system called the Proactive and Adaptive Handover (PAHO) system. PAHO is an application-level approach that uses SIP to manage client-initiated connection handoff across heterogeneous networks based on the IEEE 802.21 framework with designated user/configuration policy. Unlike conventional systems which make sub-optimal decision when managing connection handoff due to limited awareness of the relevant context for the application/service being delivered, PAHO defines proper interface to interact with the application as to determine when and to where the handoff and/or codec switching should take place in the event of network performance degradation. The results showed that using the PAHO approach on an audio/video conferencing session helps reducing the overall handover delay from 10.766 s (on non-PAHO system) down to at least 288 ms, and slowing down the degradation of MOS value throughout the entire experiment in the event of signal degradation as well as network congestion. It is also shown that load balancing among the access points (AP) could be achieved with an improved Information Server (IS).

Keywords: Proactive and Adaptive Handover system; Seamless handover; SIP-based client-initiated handover; Network assisted handover

*Corresponding author.
E-mail addresses: yee.yoke.chek06@mmu.edu.my (Y.C. Yee), khongneng.choong@bt.com, choongkn@hotmail.com (K.N. Choong), andy@mmu.edu.my (A.L.Y. Low), swtan@mmu.edu.my (S.W. Tan).

1084-8045/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2007.09.002

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

1. Introduction

Wireless networks are evolving into wireless IP networks to not only overcome the limitations of conventional circuit-switched wireless networks, but also bringing the successful Internet service creation and offering paradigm to wireless networks. Fueled by the omnipresence and overlay of various networks such as GPRS, 3G, WiFi and later WiMAX, vast array of IP or Internet-based services such as VoIP, instant messaging, video conferencing, and web browsing have been and will continue to be made possible for users on the move. As networks are gradually converged, support for seamless connection handoff is becoming an essential requirement for enabling the next generation mobility services, so as to allow users to continue accessing the different services regardless of location, terminal and most likely independent of network access type. It is therefore necessary to have an efficient support for switching between different wireless or wired networks, between various devices, and between different codec rates. The switching must be smartly triggered as to ensure the application-specific QoS is not significantly affected. Generating triggers solely based on the Received Signal Strength Indicator (RSSI) monitoring as in Jayaram and Sreenivasulu (2006), Cho and Kim (2005) and Wang and Bao (2005) would fail and render bad QoS impact to the current application session, if what has really happened was network congestion instead of signal degradation. Also, always hunting for the next best network to handoff to for better performance would be expensive if the same QoS level could be maintained by performing some application-level adjustment, such as codec switching (Ng et al., 2005; Perkins and Gharai, 2006). Moreover, unmanaged individual handover that is asynchronous could cause all mobile devices to switch to the same access points (AP) upon detecting performance degradation, resulting worst overall performance (Brickley et al., 2005; Balachandran et al., 2002).

In this paper, we proposed a framework called the Proactive and Adaptive Handover (PAHO) that aims to proactively generate trigger to the real-time application such as VoIP or video conferencing system as to perform adaptive switching, based on predicted RSSI degradation and/or network congestion. We incorporated the Fast Fourier Transform-based (FFT) signal decay detection scheme, application-layer RTCP monitoring routine as well as the network-hosted information element as to facilitate possible switching or handover. Three types of switching are supported in PAHO. The first switching is about changing the network point-of-attachment for vertical handoff or more commonly known as terminal mobility. The second refers to switching from one device to another while maintaining the contact and ongoing services, or commonly regarded as personal mobility. The third switching is about rate/service adaptation whereby upon experiencing network congestion, an ongoing VoIP session will switch from higher bitrate codec (e.g. PCMU at 128 kbp) to lower bitrate codec (e.g. GSM at 12.2 kbp) as to maintain the ongoing session.

The need of supporting three different adaptive switching schemes within a single system could be exemplified with a usage scenario that consists of three different networking environments, which are the home, outdoor and office network as shown in Fig. 1. Imagine you are late to the office so have to join in the video conferencing session from home, and upon leaving your house, the WiFi connection starts losing signal causing a switch to WiMAX that operates in the outdoor space. The WiFi interface is subsequently shut down to conserve power usage. Given WiMAX connection is more expensive, the video stream is automatically switched off leaving only the audio stream active according to your user preference configuration. When you are nearby your office where WiFi signal can be

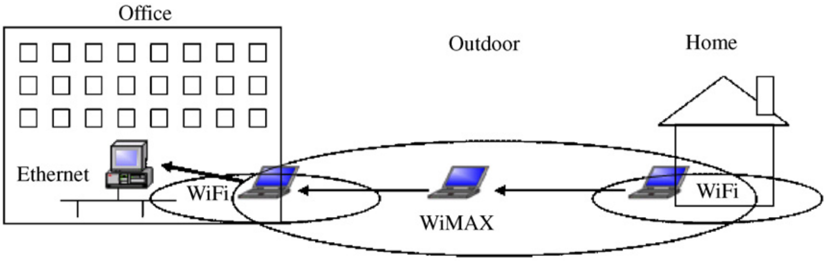

Fig. 1. The usage scenario for multiple switching in heterogeneous network environment.

picked up, the connection is switched back to WiFi as it is free. However, before reaching your desk, your office wireless network experiences congestion, causing an audio codec switching from PCMU to GSM as to retain the QoS expectation. Upon reaching your desk, the session is transferred (upon a button press) to your desktop that is connected to the company Ethernet, and this causes the video to be automatically turned on again followed by a switch of codec from GSM back to PCMU.

There have been several works focusing on vertical handoff support (Dutta et al., 2004, 2005; Puttonen et al., 2005; Guo et al., 2004, 2005) and codec switching (Ng et al., 2005; Perkins and Gharai, 2006). Our unique contributions were on integrating different level of switching schemes within a mobility management system that performs switching selectively according to the network conditions (including RSSI analysis and network traffic status), application QoS requirements and user preference, while considering the connection cost and battery power, as highlighted in the above usage scenario. We reported our connection switching experiences in networks consisting of both WiMAX and WiFi. We borrowed and extended the concept of always-best-connected as described in Gustafsson and Jonsson (2003) with the additional capability of performing adaptive codec switching as to not always hunting for the best network. This would be an important feature to operators as to avoid calls from being handover unnecessarily to other operators' networks due to network congestion. Our prototype implementation of PAHO is based on the IEEE 802.21 Media Independent Handover (MIH) seamless handover framework, whereby we utilized the concept of event which triggers possible handover, the command that executes handover and the IS that keeps both static and dynamic network-related information. The use of IS in our prototype helps removing the need to implement extra network monitoring as in Puttonen et al. (2005) and Guo et al. (2004, 2005) on the mobile device which utilizes more power; and also resolving the unbalanced load problems among APs as the result of uncontrolled simultaneous connection handover to the same AP.

We have chosen SIP as the mobility management protocol because it operates at the application-level, offering better portability as opposed to Mobile IP (MIP) given that protocol at the higher layer reduces further dependence on the access networks (Banerjee et al., 2003). SIP supports host mobility from end-to-end, eliminating some of the shortcomings associated with MIP and its route optimization variants (Guo et al., 2004, 2005). Moreover, SIP supports both personal and terminal mobility well (Akhtar et al., 2003), and it is identified as the IP Multimedia Sub-system (IMS) standard protocol giving both the compatibility and scalability advantage of our prototype to future communication systems development.

The remainder of the paper is organized as follows. Section 2 reviews some related works from the literature. Section 3 presents the proposed PAHO system describing its architecture, the switching sequence and algorithm. The descriptions of the prototype implementation, network testbed and IS are provided in Section 4. Section 5 discusses the performance measurement of system using PAHO and non-PAHO approach, and the impact of using an improved IS on load balancing. Section 6 summarizes the paper and highlights some future works.

2. Related works

To date, many mobility management schemes have been proposed to address seamless handover focusing on intelligent selection of APs and maintaining session continuity, where each scheme employs a different handover strategy (Jayaram and Sreenivasulu, 2006; Cho and Kim, 2005; Wang and Bao, 2005; Dutta et al., 2004, 2005; Puttonen et al., 2005; Guo et al., 2004, 2005; Banerjee et al., 2003; Akhtar et al., 2003; Sharma et al., 2004; Pahlavan et al., 2000; Bi et al., 2004; Sundaresan and Papagiannaki, 2006; Leggio et al., 2005). In general, these schemes could be categorized into the traditional and the advanced mobility management schemes based on a few factors. These factors include the network type (either homogeneous or heterogeneous network) to be supported, handover metrics to be considered for handover, and the terminal capability. A detailed discussion of the different factors and techniques used for connection handovers can be found in Siddiqui and Zeadally (2006).

Traditional systems are usually designed for homogeneous network, e.g. the GSM. As opposed to the advanced scheme, the limitation here is that no user selection of networks is allowed because there is only one choice of access technology. Moreover, connection handover here is mainly network-initiated and network-managed. The traditional systems (Jayaram and Sreenivasulu, 2006; Cho and Kim, 2005; Wang and Bao, 2005) rely only on the channel quality indicated by the RSSI as to determine possible handover, whereas in the advanced systems other metrics such as packet lost, bandwidth availability, connection price, user preferences, and security are considered for a handover. In Cho and Kim (2005), the mobile terminal is designed to only recognize changes of RSSI as to perform seamless handover between the 3G and WLAN networks. OmniCon (Sharma et al., 2004) suffers from the same limitation and ignores any application requirement and user preferences for switching network. Although more intelligent signal processing techniques could be employed into both the traditional and advanced systems, such as applying the exponential Moving Average technique on RSSI as in Wang and Bao (2005), or even introducing the neural network technique as in Pahlavan et al. (2000) to analyze the RSSI, ignoring other metrics will lead to inefficient handover decision, especially in the heterogeneous network environment because there is no comparable signal strength at the physical-layer due to different physical/radio techniques (Guo et al., 2004).

Leveraging on more powerful and intelligent mobile terminal, it is now possible to design advanced system that studies more handover metrics (on the mobile device) as to make more optimal handover decision. The handover scheme in Bi et al. (2004) uses both the static and dynamic parameters to justify for possible handover. The static parameters here include the handover policies, user profile, application settings and network maximum capabilities, while the dynamic parameters include the network-proved information, perceived QoS and application requirements. It is shown in their simulation results that more efficient seamless handoff is possible, but technical details on how to implement the scheme were not provided. In Sundaresan and Papagiannaki (2006), a new metric called the expected throughput is proposed as to improve handover decision. However, this work requires low-level modification to the firmware and thus not scalable or portable to other platforms. While Guo et al. (2004, 2005) have exploited the Network Allocation Vector (NAV) through the WLAN Media Access Control (MAC) layer as to define two new metrics which are the available bandwidth and access delay, for making more accurate handoff decision. A more generic solution is found in Puttonen et al. (2005), whereby a cross-layer framework called the Link Information Provider (LIP) that provides underlying network events and parameters to upper layer is introduced. Using LIP, it took 3 s to complete a homogeneous connection handover, 0.18 s to switch from WiFi to Bluetooth and nearly 30 s to activate an Ethernet connection. However, these are only low-level connection time, without considering the presence of any application session. Our PAHO system uses similar higher-level metrics as others, but collected at a higher protocol layer. Instead of modifying firmware or performing MAC layer sensing, we collect network status from multiple layers. The RSSI is collected from layer 2 through the network driver while other metrics (e.g. the packet lost and jitter) are extracted from the Real Time Control Protocol (RTCP) packets. By doing this, PAHO could directly or closely monitor the performance of the real-time application and take appropriate switching decision.

After making optimal handoff decision (i.e. to which AP to handoff to), the next challenge is to conduct the handoff execution at minimal time as to eliminate or at least minimize possible disruption to the existing application session. From the literature, faster handover is made possible by delivering both data and contextual information to the new point of attachment prior actual handover execution. Seamoby Working Group is focusing on context transfer (Leggio et al., 2005) as to reduce the handoff time. This is done by transferring the information related to the mobile node from the current access router to the next access router over the wired network, avoiding using the limited wireless bandwidth resources. In the absence of wired network, mobile device with multiple radios could utilize multiple network interfaces concurrently as to disguise the handover delay, by performing the soft handover or adopting the make-before-break model. The limitation of context transfer solution is that both the new and current access routers must support the same context-transfer candidate services; otherwise it could lead to service disruption (Siddiqui and Zeadally, 2006). Hence, we decided to incorporate the make-before-break model in PAHO for achieving seamless handover as to avoid possible infrastructure limitations.

Mobility management framework should not only be focusing on fast handoff, but also on addressing the effect of unmanaged individual handover which causes unbalanced load among the APs as clients are always looking for the AP with the best RSSI or bandwidth rate. This unbalanced load results in unfair bandwidth allocation among users as stated in Bejerano et al. (2004) and generates unnecessary traffic on performing handover. Some literature and standard development efforts exist on the subject of load balancing for WLAN. In Bejerano et al. (2004), this problem is addressed by intelligently associating clients to the APs, instead of allowing clients to greedily associating with the APs by themselves. Brickley et al. (2005) have proposed a cell-breathing technique whereby an AP can reconfigure its cell boundaries by changing the power at which it transmits based on the traffic rate at each AP. By doing this, clients at the edge of overloaded network are indirectly forced to connect to the neighboring AP. However, there are cases where such technique could cause worst load distribution as reported in García et al. (2006).

The 802.11k (IEEE 802.11k Radio Resource Measurement, 2007) is an IEEE standard development effort that works on the standardization of radio measurements. Under this effort, the AP could ask clients to report their network or radio related information, and the clients in turn could request data (i.e. the site report) from the APs. The concept of site report is similar to the information service defined in IEEE 802.21 framework, except that it holds only both layer 1 and 2 information of the 802.11 network and is accessible only through the 802.11 radio. While, the IS could further hold other information such as network operator information, cost related information and services available at a specific AP. Moreover, the information service could be accessed using any radio. In this paper, we address the unbalanced load issue by incorporating some dynamic network information into the IS which helps on the preparation of a filtered network list that is to be sent to the clients. Clients are expected to always start choosing the first network from the list as to gain the optimal network performance. Detailed description of our improved IS can be found in Section 4.3.

3. The PAHO system

3.1. Architecture

Fig. 2 shows the multi-layered architecture of PAHO. It is generally divided into three layers, which are the SIP-based application, PAHO and the underlying network drivers.

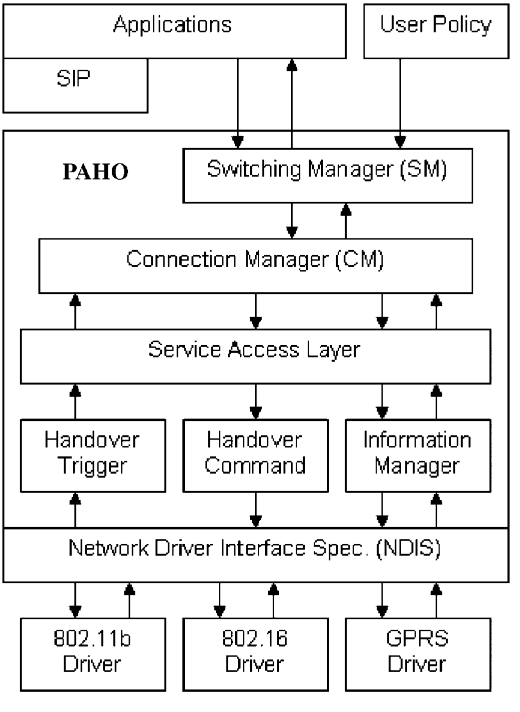

Fig. 2. Architecture of PAHO.

There are several important modules that work together as to support seamless handover/switching in PAHO. The fundamental functionality of Connection Manager (CM) is to detect, connect to, and disconnect from the network, and also query various properties of the network such as the type of networks (802.11, 802.16 or others), signal strength, bandwidth rate, etc. These functionalities are exposed to higher level modules through clearly defined APIs. The CM works closely with the Switching Manager (SM), which is responsible for performing connection switching that includes network switching, device switching and codec switching. SM is responsible for deciding which type of switching to perform and by when to perform it after collecting information or event via the CM. SM is placed in between the CM and application as to abstract the underlying network changes and connectivity from the application. For the application, it only needs to specify which network interface to use and the relevant IP address will be returned as to establish the application-level connectivity. For handling connection switching, application is required to register a callback function to the SM, which SM will later invoke with the relevant parameters as to enable the application to manage the application-level switching adjustments. To deal with ping-pong switches, SM employs timer limits on handovers from one network to another, as to avoid frequent handover leading to service degradation.

The user policy is a text file as shown in Fig. 3a and b which the user could configure his/her connectivity preferences. At the current stage of implementation, we have catered only for some initial settings. First is the priority setting of which network to connect to first, e.g. the user would prefer connecting first to Ethernet (if available), then only to other networks such as WiFi and WiMAX, in the order of connectivity cost or available bandwidth rate. Another setting is the on/off setting for different service based on the type of network connected to, e.g. given that Ethernet is fast and free, the video session is set to be on, whereas video is off for connecting to WiMAX.

Further down the architecture is the Service Access Layer (SAL) which is equivalent to the Service Access Point (SAP) in the IEEE 802.21 framework. SAL is a set of APIs through which the different MIH Function (MIHF) can communicate with the upper layer and lower layer entities using the 802.21 proposed primitives. MIHF defines three different services which are the event service, command service and information service. These are mapped into three separate modules underneath our SAL, namely the handover trigger, handover command and information manager. The handover trigger module is responsible

a

```
[Network]
Always connect to Ethernet: Yes/No.
Factor: Cost, Highest/Lowest.
        Bandwidth, Highest/Lowest.
        Auto,

[Application]
[Video]
Ethernet: on/off.
WiFi: on/off.
WiMAX: on/off.
```

b

```
[Network]
Always connect to Ethernet: Yes.
Factor: Bandwidth, Highest.

[Application]
[Video]
Ethernet: on.
WiFi: on.
WiMAX: off.
```

Fig. 3. (a) User policy template. (b) Sample user policy.

for monitoring the connectivity of the mobile device to the connected network as to generate handover triggering event to upper layer module of predicted link-going-down (or actual link-down event) by applying the FFT-based decay detection algorithm on the RSSI as in Puttonen et al. (2005) and Guo et al. (2004). The RSSI samples are taken at 100 ms interval over a dynamic sampling window size. The handover command is a set of commands which allows the SM (through the CM) to initiate and coordinate the connection switching. The information manager provides a query/response mechanism for the SM to query for network-related information from the IS, such as the list of available networks nearby and their types, the operators, and the connection cost.

3.2. Switching interaction

In this section, we describe the detailed message flow for performing both connection switching and codec switching as supported by PAHO. We omit the discussion on device switching (personal mobility) because it is a straight-forward process and references to such works can be found in Thai et al. (2003) and Zeadally and Siddiqui (2004).

3.2.1. Connection switching/handover interaction

Fig. 4 shows the message flow/interaction between various components for performing a connection handover using PAHO. Initially, the mobile device which is a Multi Radio Terminal (MRT) is connected to the WiFi network, with the WiMAX radio turned off. When the mobile device is gradually moving away from WiFi, a link-going-down event is generated to the CM, where the SM shall then perform a connection switching or

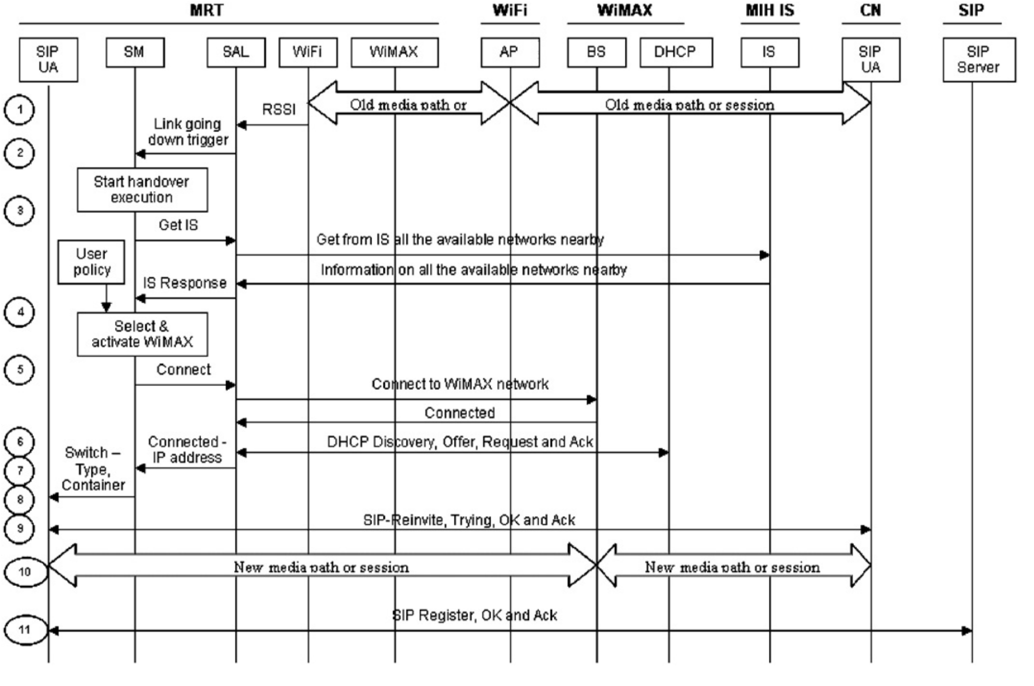

Fig. 4. Message flow for connection handover.

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

handover. The make-before-break model is adopted as to maintain the application session with zero packet lost. The handover steps are as follows:

1. The SIP UA on the MRT has established an active media session with the SIP UA in the corresponding node (CN).
2. When the MN is moving away the WiFi coverage, a link-going-down event is triggered by SAL to the SM due to the drop of signal quality, i.e. the RSSI.
3. Upon receiving the event, the SM starts the handover process by first contacting the MIH IS. Through a single query and response message flow, the SM may get the relevant network information such as the list of available network around and their network types.
4. In our case, the WiMAX is selected given that it is the only wireless WAN available.
5. The SM (through CM) proceeds to try connecting to the WiMAX network through the underlying APIs provided by SAL.
6. Once connected to the WiMAX network, the SAL starts the DHCP address acquisition process.
7. The new IP address which was provided by SAL is then passed to the SM.
8. The SM then invokes the callback function registered by the SIP UA, supplementing with parameters such as the network type of the newly connected network and the new IP address.
9. Once the new IP address is obtained, the SIP UA creates a Re-invite message with the new IP address and sends it to the CN (instead of via the SIP server for performance benefits) maintaining the same call-ID.
10. The new media session is then established.
11. As there is a change of IP address, the SIP UA is required to update the SIP server.

3.2.2. Codec switching interaction

Fig. 5 illustrates another set of SIP signaling messages that is used for performing codec switching, i.e. modifying the audio session parameters using the Session Description Protocol (SDP) between two SIP UAs. The RTCP which is the control protocol of RTP is used to monitor the performance of the RTP session as used in Ng et al. (2005). The collected RTCP Receiver Report (RR) which provides periodic control information of the receivers (such as the total packets received, total packet lost and jitters at certain time interval) is passed to the SM for possible switching decision. Possible codec switching is determined by comparing the Packet Lost Rate (PLR) against a predetermined threshold. To remove variations in the PLR, we have adopted the Moving Average technique (or the Low-pass filter) as used in Hong et al. (2002) to provide a smoothing effect as to determine when switching should happen.

There are two types of codec switching in our current implementation. Upon exceeding the packet lost threshold, the VoIP session will be downgraded from the PCMU at 128 kbp to GSM at 12.2 kbp. When the packet lost is later recovered below the threshold, an upgrading of codec is performed.

3.3. Switching algorithm

Fig. 6 shows the flowchart of the switching algorithm. The algorithm starts by first checking the RSSI value. If it is lower than the pre-defined threshold, the network

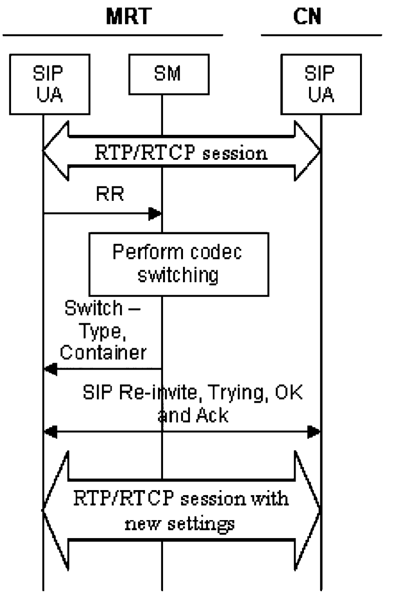

Fig. 5. Message flow for codec switching.

switching process as shown in Fig. 4 will be performed. Otherwise, the flow continues to check whether any PAHO-based RTP application is currently executing. If yes, the respective RTCP packet will be examined to check whether the PLR is higher than the pre-defined threshold. If it is higher, the switching codec process as shown in Fig. 5 will be executed.

4. Prototype implementation and testbed

4.1. System implementation

We have implemented a prototype system on the Windows platform, based on the architecture described in previous section, as to measure the handover and switching performance. We modified and customized The GNU oSIP library (2007) to meet our requirements. The oSIP is selected as it comes with the SIP/SDP parser, state machines for client/server transactions, and other related functions to SIP/SDP transactions. The VoIP library we used is JVOIPLib (2007). Both the mobile device and the CN run on the Windows laptops with 1.68 GHz processor and 1 GB memory capacity. These notebooks are equipped with both the Orinoco 802.11b and Symmetry 802.16 interfaces.

As explained, the conventional system relies only on the OS (Windows XP in our case) to detect on possible link-down event. We make use of the Windows IP Helper API (IP Helper API Programming, 2007) to detect the presence and absence of network connection and information.

We have pre-configured several threshold values for connection switching and codec switching. For connection switching, we have defined four threshold settings according to

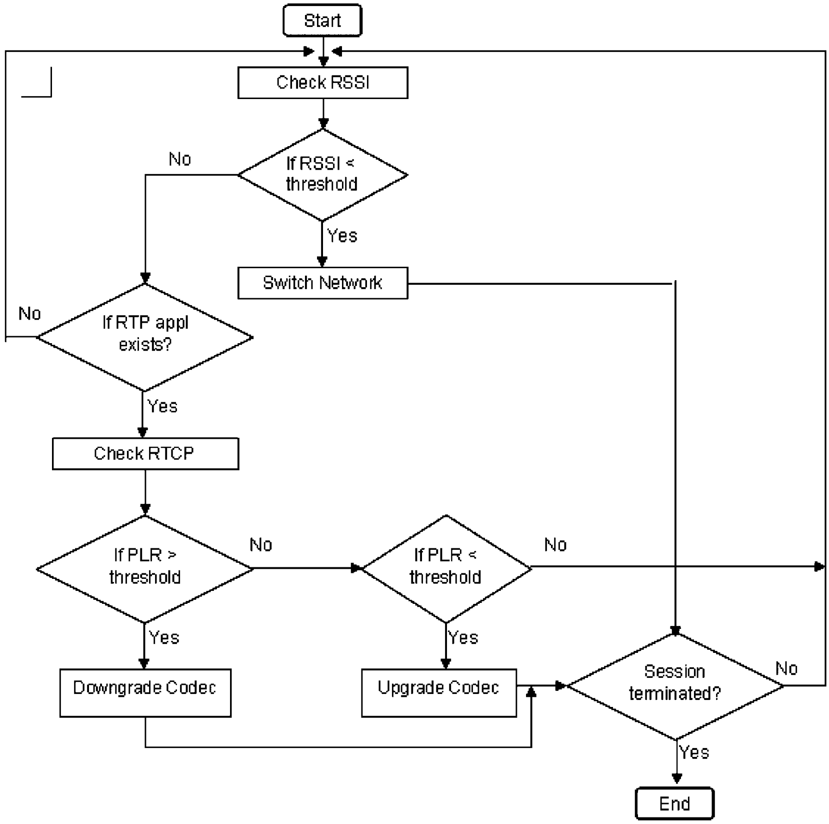

Fig. 6. Switching algorithm.

the 802.21 framework, while for codec switching, only one threshold is defined at this stage of implementation, as shown in Table 1.

4.2. Network testbed

The network testbed we have used for experiment is shown in Fig. 7. The arrow depicts the usage scenario that we have described earlier. There are a total of four zones, i.e. two WiFi zones (A and C), one WiMAX zone (B) and the Ethernet zone (D) which is under zone C coverage. The WiFi access points used are the Netgear (Netgear ProSafe 802.11G Wireless Access Point, 2007) and LinkSys, in zone A and C, respectively. Netgear is specially chosen as it supports real-time change of power level, which we could configure to simulate RSSI fading (i.e. client mobility). The mobile device and the CN are placed over different subset across the router. We have also attached the traffic generator (i.e. the Distributed Internet Traffic Generator (D-ITG), 2007) in zone C to generate artificial traffic as to flood the WiFi network. The IS is contacted twice when the mobile device is on the move, i.e. whenever the mobile device is located in the overlapping zones, which are zone A–B and zone B–C. The first IS query is sent by the mobile device upon receiving the link-going-down event. The second IS query however, is manually triggered at the moment

Table 1
Threshold values for experimental testing

Threshold type	Value
Link down	−60 dB
Link-going-down	−50 dB
Link-coming-up	−40 dB
Link up	−35 dB
PLR	3% (VoIP service quality and principles and guarantees, 2007)

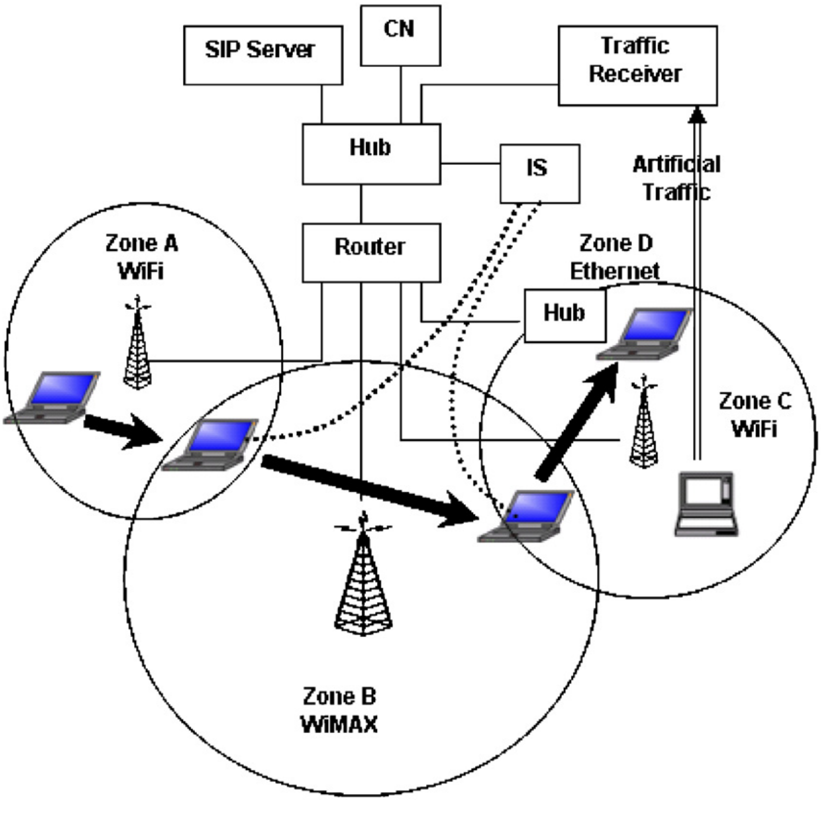

Fig. 7. Experiment network testbed.

because the mobile device is not equipped with GPS, hence unable to track its location as to when to trigger a handover. Anyhow, we are able to collect the performance metrics for the performing handover.

4.3. Information Server (IS)

In our current IS implementation, we have included only limited number of information elements sufficient to demonstrate the seamless handover scenario. We store a list of

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

networks from different locations, where each network holds the information such as the network type, network identifier, operator's name, connection cost and the available bandwidth rate. Among these information, only the bandwidth rate is dynamically updated at fixed interval by an agent that monitors each AP through the use of Simple Network Management Protocol (SNMP). In our current experiment, both the overlapping zones are assigned as location 1 and 2, respectively. This location number is submitted to the IS as part of the IS query packet and the IS shall return the list of available networks. For example, the WiMAX network elements are returned upon submitting IS query with location 1, and the WiFi in zone C is returned for submitting location 2, at each overlapping zone, respectively.

As explained earlier, we proposed to use the dynamic information (i.e. the available bandwidth rate) as the key to prepare a list of sorted networks in descending order. This pre-arranged network list is dynamically generated upon receiving an IS query, and clients are expected to always select the network with the highest bandwidth rate, i.e. the first network from the list generated. Hence, upon sending the IS response to the client, the IS will make a record assuming that this client will join the first network within the next period of time (say 30 s), occupying (or reserving) a specific amount of bandwidth. If there is a second IS query at this time, the network list may be different given that certain bandwidth in the first network has already been reserved for the first client. Whenever the client has successfully connected to the selected network, it will acknowledge the IS of which network it has connected to. If no acknowledgement is received within the next 30 s, the reserved bandwidth will be released. The acknowledgment is important for the IS to update its records and to generate the next network list. Fig. 8 shows the proposed interaction between the client and the improved IS. Regardless of whether the second query (from client 2) came in at Time 1 or 2 (the arrow), the bandwidth reservation made upon serving the first query (from client 1) will affect the sorted network list provided to client 2.

The proposed sorted list is still feasible if cost is the factor instead of the bandwidth rate on selecting the target network. The IS will first sort the network list based on the cost in ascending order, followed by a second sort using the bandwidth rate as the key in descending order. Hence, the network list will always have the network with the highest bandwidth rate on the top at the cheapest cost. Client is asked to traverse through the list from the top as to find a network that matches its requirements. In short, we have included the capability of sorting the networks as well as performing bandwidth reservation for two reasons. First is on helping client to locate the network faster. Second, reserving bandwidth allows the IS to perform load balancing as clients are asked to connect to the suggested network. It is interesting to study the effect of applying the proposed IS mechanism into the call admission scenario of our earlier work in Yee et al. (2007).

4.4. Seamless handover

While soft handover generally means the mobile device can communicate with more than one APs during handoff, we further define two new terms: semi-soft and full-soft handover. Given that the handover process is broken into a few phases, which are the address acquisition (obtaining IP address), binding update (address association, session registration, exchange of new session information) and media redirection (actual switching

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

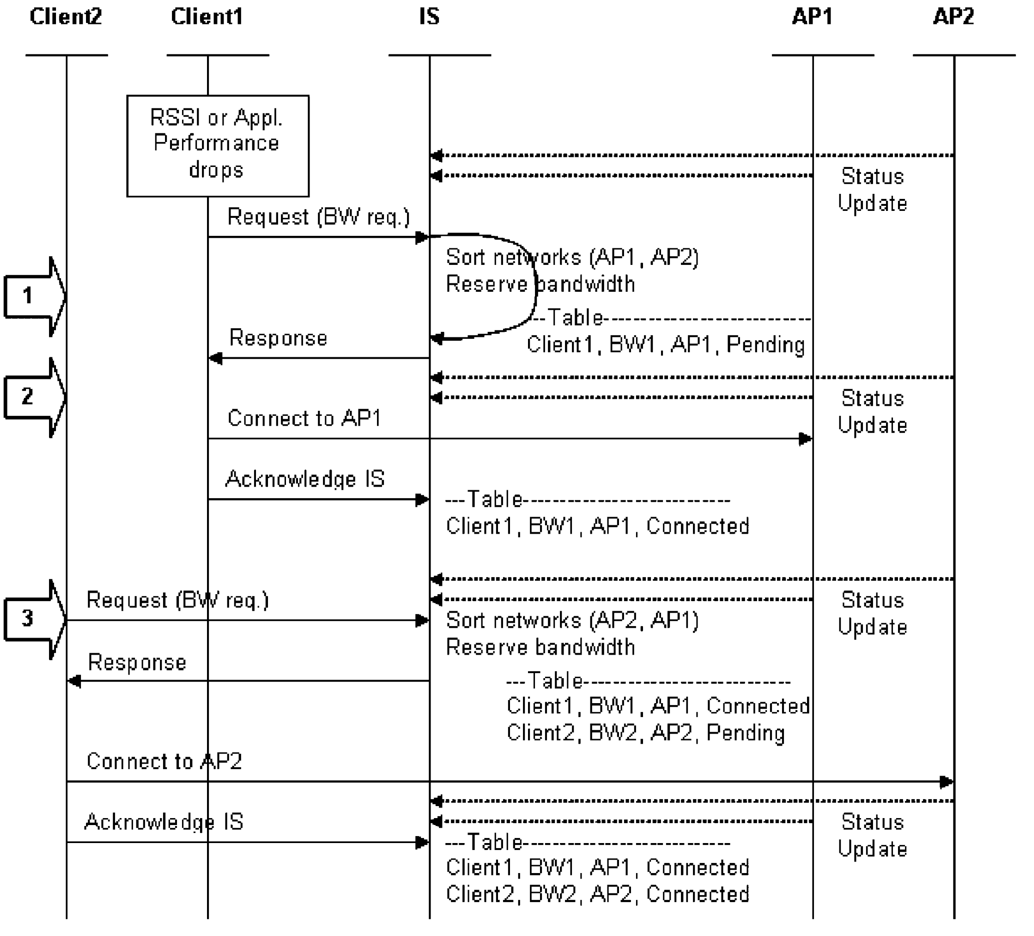

Fig. 8. Message flow between the clients and the improved IS.

of media flow between interfaces), our two new terms are different in terms of the number of phases it supported during a handover as shown in Fig. 9.

Full-soft handover means all three phases are executed over the alternative (second) network interface, before the currently used (first) interface is brought down. This means the ongoing application (RTP) session is created (duplicated) and delivered over the second interface before the first interface is turned off. This type of handover technique will generate duplicated packets temporarily and hence the target client must be designed to ignore these duplicated packets. In the semi-soft handover technique, only the first two phases are performed over the second interface. The first interface is turned off right before the media redirection starts. As expected, there will be slight switching delay during the media redirection, at the time where the application session is terminated on the first interface and then restarted on the second interface. The benefit of this type of handover is that there will be no duplicated packets generated to both the network and target receiver(s). However, the practical usage of the semi-soft handover lies on the ability to

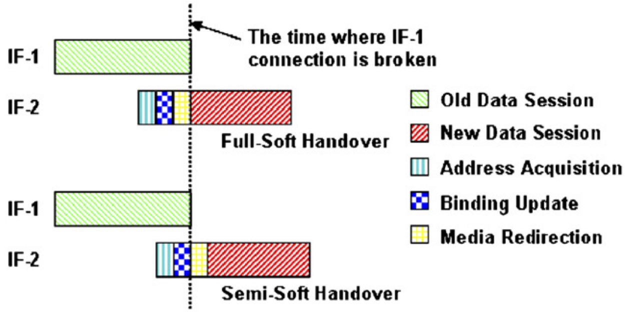

Fig. 9. Differences between full-soft and semi-soft handover.

minimize the application stop-and-start delay. We have measured and reported the performance of both soft handover techniques in the next section.

5. Performance evaluation of PAHO

We have performed and analyzed three sets of performance evaluations from the usage scenario described in Section 1. The first set is the handover performance of our PAHO system against the conventional system that relies only on the OS to report network disconnection. The evaluation carried out here includes the verification of seamless handover from WLAN to WiMAX, i.e. the changes of sequence numbers during handover, followed by the breakdown of handover time. The second set is the study of PAHO against changes to the network condition, i.e. its response to network congestion. For this study, we will flood the network as to create congestion and expect PAHO to trigger codec switching. We reported the MOS value when using PAHO against the conventional system which relies only on the RSSI as the switching factor. The third set is on the network load distribution with and without using the improved IS.

We have used both the Ethereal (2007) and OmniPeek (2007) to collect and infer the required performance metrics, which are the handover disruption delay, MOS and PLR. The disruption delay ($T_B - T_A$) refers to the time difference from the first packet received from network interface B (T_B) to the last packet received from network interface A (T_A) (Guo et al., 2004).

5.1. Handover performance

To verify that the seamless handover has taken place, we have collected the changes of RTP sequence number for handover from WiFi to WiMAX. The performance of the two soft handover techniques as described earlier is measured as shown in Fig. 10. As indicated, there were some overlapping packets for the full-soft technique for about 288 ms, whereas a gap of similar duration is observed for the semi-soft technique. This delay is slightly longer than the limit for acceptable quality voice connection delay which is at 250 ms (Understanding Delay in Packet Voice Networks, 2007). We believe the existing VoIP library could be further optimized as to reduce the delay.

Fig. 11 shows the performance comparison between the PAHO (uses the full-soft technique) and the conventional handover system (non-PAHO). As expected, the full-soft

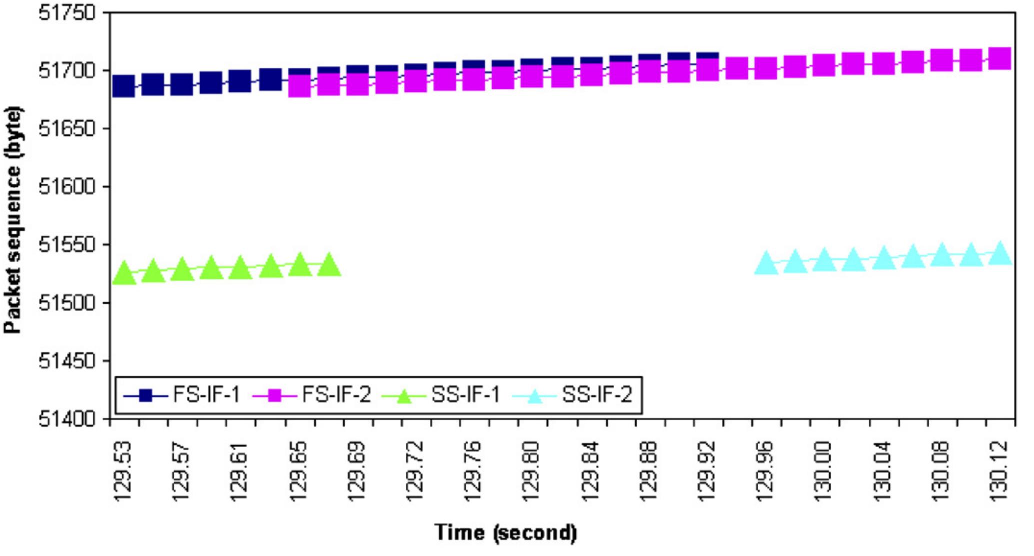

Fig. 10. Handover delay using semi-soft and full-soft techniques.

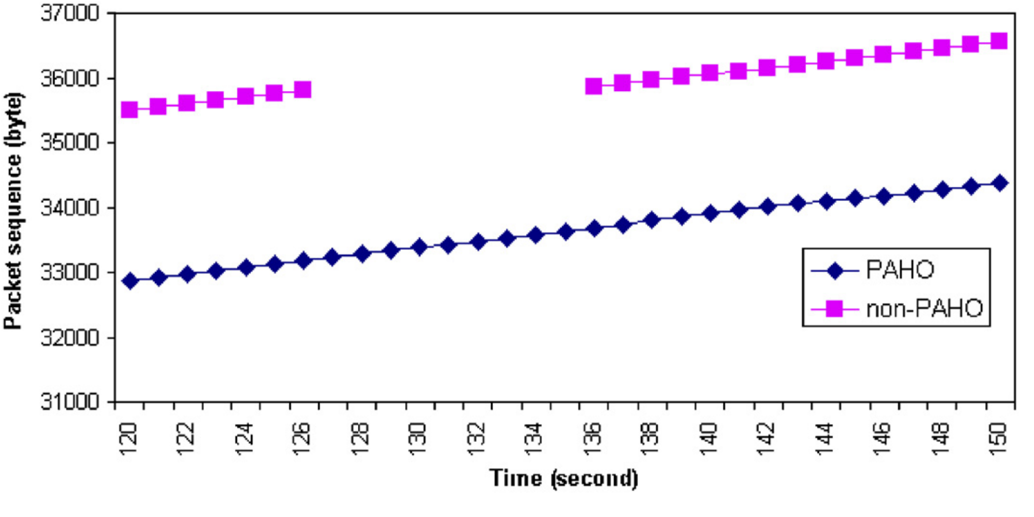

Fig. 11. Handover delay using PAHO and non-PAHO systems.

technique has outperformed the conventional system that suffered from a disruption delay of 10.766 s.

Table 2 gives a breakdown of the handover time for switching from WiFi to WiMAX network, with and without using the PAHO system.

As shown in Fig. 11, handover without PAHO has failed performing a seamless handover because the second connection is established after the first connection (WiFi) is broken for 10.766 s, causing a significant break in the audio session. This delay is mainly due to the dependency on the OS to detect the link-down event, occupying 74% of the total

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

Table 2
Time spent on the handover process

Handover stage	Without PAHO (s)	With PAHO (s)
Link down/going-down detection to network selection	7.969 (Signal lost → link-down detection)	2.327 (Link-going-down → network selection)
IS query/response	N/A	0.563
Address acquisition	1.000	1.000
Binding update	1.141	1.150
Media redirection	0.656	0.653
Total handover time	10.766	5.693

handover time, before the system can take further handover action. The time spent on address acquisition, binding update and media redirection was quite consistent on both the systems.

5.2. Effect of adaptive codec switching

By creating congestion through flooding the network using the D-ITG tool, we have studied the effect of performing codec switching. Fig. 12 shows that as more packets are generated into the network, the bandwidth becomes limited and the PLR increases, resulting lower MOS value. As indicated, the PCMU codec is less resilient than the GSM when the network is experiencing congestion.

Fig. 13 shows the MOS value for system that runs with and without PAHO. Without PAHO, it is clearly seen that the MOS drops as more traffic is flooded into the network. This is because the non-PAHO system relies only on the RSSI to trigger switching, which failed to react to network congestion. However, in the case with PAHO, when the PLR reaches the pre-defined threshold of 3%, the application is triggered to switch to a lower bitrate codec (i.e. the GSM codec) as it can work at lower bandwidth. The switching helps deferring the degradation of the audio quality at a slower rate than before. The codec switching took only 0.523 s, which is an insignificant value as compared to the handover time discussed above. As a result, it is advisable to first exploit the possibility of switching codec before attempting to perform connection handover, in the event of performance degradation. Moreover, prioritizing to first perform codec switching also helps preventing mobile devices from handover unnecessarily to alternative network that is administered by other operators.

5.3. AP load balancing

We have modified both the testbed and the user policy to conduct the experiment on load balancing between APs. We used four WiFi APs and 10 clients. The factor option under the network column in each user policy (see Fig. 3) is set to Auto, i.e. always selecting the first network from the list. All 10 clients were initially connected to AP-4 (Netgear) which we would later reduce its power to force handover to the rest of the APs. We have also removed the antenna(s) of AP-1 and AP-3 as to reduce their signal strengths. OmniPeek was used to monitor the network traffic at each AP throughout the experiment and the various readings are collectively shown in Figs. 14 and 15.

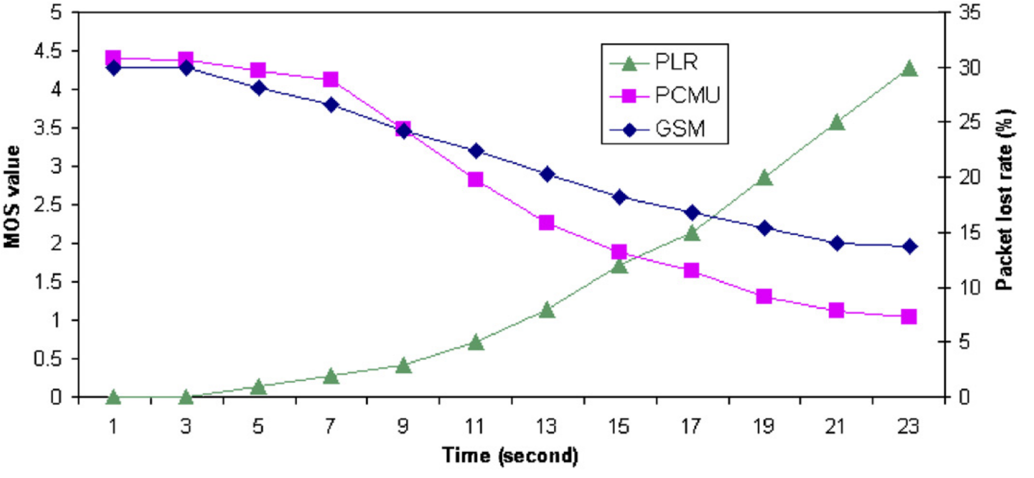

Fig. 12. MOS for using PCMU and GSM codecs.

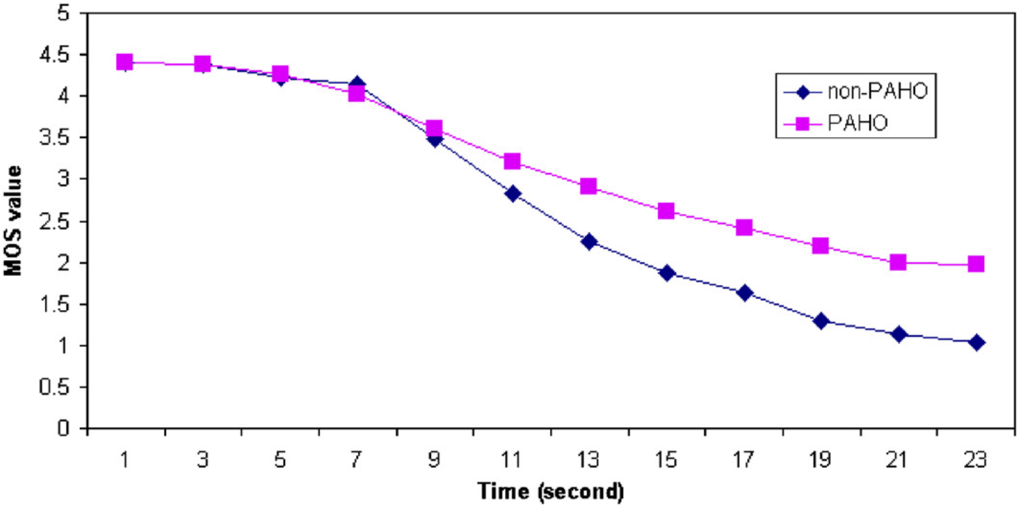

Fig. 13. MOS using PAHO and non-PAHO systems.

It can be seen from Fig. 14 that the traffic load in the network was unevenly distributed with AP-2 carrying most of the traffic while AP-3 is left under-loaded throughout the experiment. The reason AP-2 gained so much traffic was mainly because it has the highest signal strength (antennas are attached), causing all non-PAHO clients which rely only on RSSI as the indicator to connect to it.

With the use of PAHO and the modified IS, the AP traffic loads distribution were significantly improved as shown in Fig. 15. It demonstrated that the proposed sorted network list and reservation scheme can indeed improve the overall network performance.

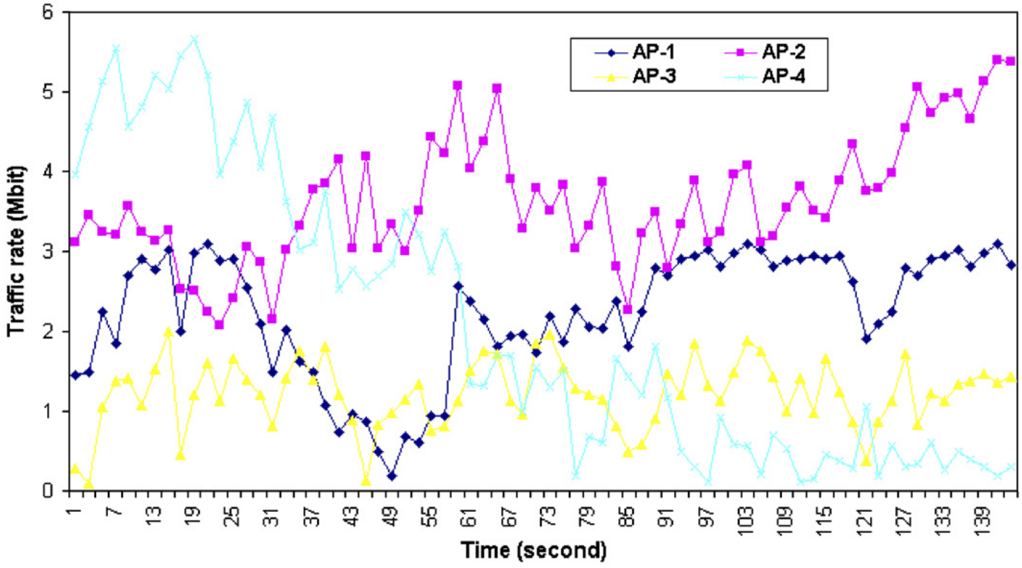

Fig. 14. Traffic loads at APs using non-PAHO systems.

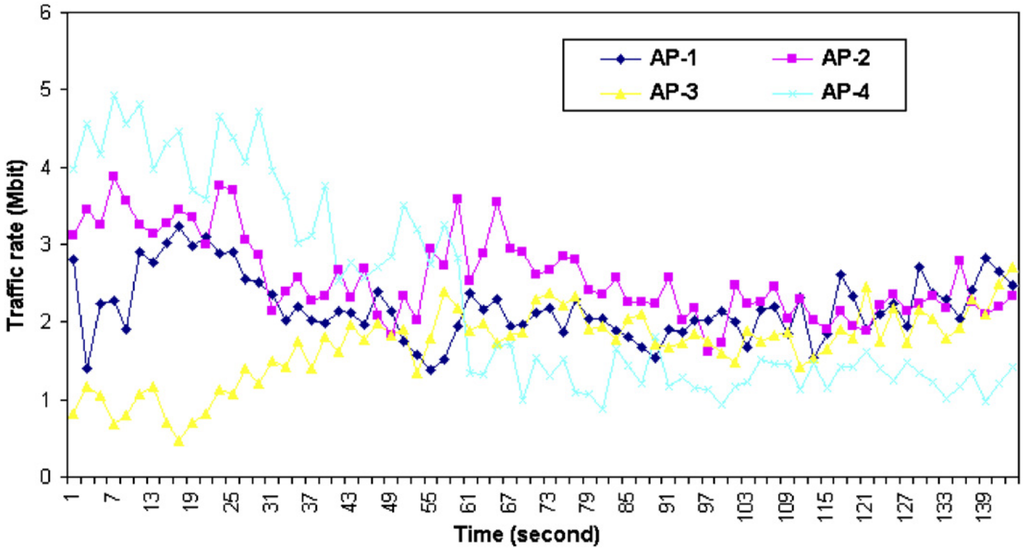

Fig. 15. Traffic loads at APs using PAHO and modified IS.

6. Summary and future works

This paper has presented a SIP-based client-initiated handover approach called PAHO that performs proactive and adaptive switching according to the changes of signal strength and/or network condition. We have described the PAHO architecture, the switching sequence and algorithm, as well as the prototype implementation details and testbed of our works. The results showed that using the PAHO approach helps improving the overall system and network performance, particularly on achieving seamless handover and load balancing among APs.

Some future works include optimizing the existing JVOIPLib as to shorten the stop-and-start delay as described in Section 4.4. Next would be exploiting other metrics that was proposed in the 802.11k to possibly enhance the handover capability as to address different handover usage scenarios. For practical purpose, it is also planned to equip the mobile device and the IS with actual location information instead of the current hard-coded location number. It is also interesting to compare the handover performance of both the client-initiated and network-initiated approaches in our next stage of research.

References

Akhtar N, Georgiades M, Politis C, Tafazolli R. SIP-based end system mobility solution for all-IP infrastructures. In: Proceedings of the IST mobile & wireless communications summit; 2003.

Balachandran A, Bahl P, Voelker GM. Hot-spot congestion relief in public-area wireless networks. In: Proceedings of 4th IEEE workshop on mobile computing systems and applications; June 2002. p. 70–80.

Banerjee N, Wu W, Das SK, Dawkins S, Pathak J. Mobility support in wireless Internet. IEEE Wireless Commun 2003;10(5):51–61.

Bejerano Y, Han S, Li L. Fairness and load balancing in wireless LANs using association control. In: Proceedings of the 10th annual international conference on mobile computing and networking; 2004. p. 315–29.

Bi Y, Huang J, Iver P, Song M, Song J. An integrated IP-layer handover solution for next generation IP-based wireless network. In: Proceedings of the IEEE 60th vehicular technology conference, vol. 6; 2004. p. 3950–4.

Brickley O, Rea S, Pesch D. Load balancing for QoS enhancement in IEEE802.11e WLANs using cell breathing techniques. In: Proceedings of the 7th IFIP international conference on mobile and wireless communications networks; 2005.

Cho H, Kim W. A seamless handover between cdma2000 and WLAN for 3G-WLAN interworking service continuity. In: Proceedings of the advanced industrial conference on telecommunications/service assurance with partial and intermittent resources conference/e-learning on telecommunications workshop (AICT/SAPIR/ELETE'05); 2005. p. 261–6.

D-ITG, Distributed Internet Traffic Generator. ⟨http://www.grid.unina.it/software/ITG/⟩; 2007.

Dutta A, Agrawal P, Das S, Elaoud M, Famolari D, Madhani S, et al. Realizing mobile wireless Internet telephony and streaming multimedia testbed. Elsevier Comput Commun 2004;27(8):725–38.

Dutta A, Oshba Y, Schulzrinne H. Seamless handover across heterogeneous networks—an IEEE 802.21 centric approach. In: Proceedings of the international wireless submit—wireless personal multimedia communications (WPMC); 2005.

Ethereal: A Network Protocol Analyzer. ⟨http://www.ethereal.com/⟩; 2007.

García E, Vidal R, Paradells J. Load balancing in WLANs through IEEE 802.11k mechanisms. In: Proceedings of the IEEE symposium on computers and communications (ISCC); 2006. p. 844–50.

Guo C, Guo Z, Zhang Q, Zhu W. A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks. IEEE J Select Areas Commun 2004;22(5):834–48.

Guo C, Wu H, Zhang Q, Song J, Zhou J, Huitema C, et al. End-system-based mobility support in IPv6. IEEE J Select Areas Commun 2005;23(11):2104–17.

Gustafsson E, Jonsson A. Always best connected. IEEE Wireless Commun 2003;10(1):49–55.

Hong GY, Fong ACM, Fong B. Adaptive QoS control of multimedia transmission over band-limited networks. IEEE Trans Consumer Electron 2002;48(3):644–9.

IEEE 802.11k Radio Resource Measurement. ⟨http://grouper.ieee.org/groups/802/11/Reports/tgk_update.htm⟩; 2007.

IEEE 802.21 Media Independent Handoff WG. ⟨http://www.ieee802.org/21/⟩; 2007.

IP Helper API Programming. ⟨http://english.geekpage.jp/programming/iphlpapi/⟩; 2007.

IP Multimedia Subsystem (IMS). ⟨http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem⟩; 2007.

Jayaram D, Sreenivasulu N. Seamless call transfer between cellular (GSM) and enterprise (vowlan) networks. In: Proceedings of the IASTED international conference on networks and communication systems; 2006.

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002

JVOIPLib. ⟨http://research.edm.uhasselt.be/~jori/page/index.php?n=CS.Jvoiplib⟩; 2007.
Leggio S, Manner J, Raatikainen K. Achieving seamless mobility in IP-based radio access networks. IEEE Wireless Commun 2005;12(1):49–54.
Netgear ProSafe 802.11G Wireless Access Point. ⟨http://www.netgear.com/Products/WirelessAccessPoints/WirelessAccessPoints/WG302.aspx?detail=Specifications⟩; 2007.
Ng SL, Hoh S, Singh D. Effectiveness of adaptive codec switching VoIP application over heterogeneous networks. In: Proceedings of the 2nd international conference on mobile technology, applications and systems; 2005. p. 1–7.
OmniPeek. ⟨http://www.wildpackets.com/products/omni/overview⟩; 2007.
Pahlavan K, Krishnamurthy P, Hatami A, Ylianttila M, Makela JP, Pichna R, et al. Handoff in hybrid mobile data networks. IEEE Commun Mag 2000;7(4):34–47.
Perkins C, Gharai L. RTP and the Datagram Congestion Control Protocol. In: Proceedings of the IEEE international conference on multimedia & expo (ICME); 2006. p. 1521–4.
Puttonen J, Fekete G, Mkel J, Hmlinen T, Narikka J. Using link layer information for improving vertical handover. In: Proceedings of the 16th annual IEEE international symposium on personal indoor and mobile radio communications (PIMRC), vol. 3; 2005. p. 1747–52.
Sharma S, Inho B, Dodia Y, Tzi-cker C. OmniCon: a mobile IP-based vertical handoff system for wireless LAN and GPRS links. In: Proceedings of the 2nd IEEE annual conference on pervasive computing and communications; 2004. p. 155–64.
Siddiqui F, Zeadally S. Mobility management across hybrid wireless networks: trends and challenges. Elsevier Comput Commun 2006;29(9):1363–85.
Sundaresan K, Papagiannaki K. The need for cross-layer information in access point selection algorithms. In: Proceedings of the internet measurement conference; 2006. p. 257–62.
Thai B, Wan R, Seneviratne A, Rakotoarivelo T. Integrated personal mobility architecture: a complete personal mobility solution. J Mobile Netw Appl 2003;8(1):27–36.
The GNU oSIP Library. ⟨http://www.gnu.org/software/osip/osip.html⟩; 2007.
Understanding Delay in Packet Voice Networks. ⟨http://www.cisco.com/warp/public/788/voip/delay-details.html⟩; 2007.
VoIP Service Quality, Principles and Guarantees. ⟨http://www.accellent-group.com/media/VoIP_service_quality_Principles_and_Guarantees.pdf⟩; 2007.
Wang J, Bao L. Layer-2 mobility management in hybrid wired/wireless systems. In: Proceedings of the second international conference on quality of service in heterogeneous wired/wireless networks (QSHINE); 2005.
Yee YC, Choong KN, Low ALY, Tan SW, Chien SF. A conservative approach to adaptive call admission control for QoS provisioning in multimedia wireless networks. Elsevier Comput Commun 2007;30(2):249–60.
Zeadally S, Siddiqui F. Design and implementation of a SIP-based VoIP architecture. In: Proceedings of the 18th international conference on advanced information networking and applications, vol. 2; 2004. p. 187–90.

Yoke Chek Yee received her Master of Science from the Faculty of Computer Science, University of Putra Malaysia (UPM), Malaysia in 2001, majoring in Distributed Computing. She has 5 years experience working as senior development/quality engineer in Motorola Malaysia. She is currently a Ph.D. student in the Multimedia University Malaysia (MMU). Her research interests include mobile computing, handover algorithms, and distributed systems design.

Khong Neng Choong received his Master of Science and Ph.D. from the Department of Computer and Communication Systems Engineering, Faculty of Engineering, University of Putra Malaysia (UPM), Malaysia in 1999 and 2003, respectively, majoring on IT and Multimedia Systems. He worked as a senior software engineer in Intel Malaysia for nearly 3 years, and currently is a principal researcher in British Telecom (BT) Malaysia. His research interests include seamless handover in heterogeneous network, Internet content delivery methods, QoS provisioning issues in the field of fixed/mobile convergence and distributed systems design.

Andy L. Y. Low received the B. Eng. (Hons.) degree in Electronic and Communication Engineering with 1st Class honours and Ph.D. degree from the University of Birmingham, United Kingdom, in 1998 and 2002, respectively. He was previously a senior lecturer and also the Associate Dean in the Faculty of Engineering and Technology, Multimedia University, Malaysia. In 2005, he joined British Telecom Asian Research Centre in Malaysia as Technical Group Leader in Fixed-Mobile Convergence (FMC). Dr. Low is a senior member of the Institute of Electrical and Electronics Engineers, USA (IEEE) and Institution of Electrical Engineers, UK (IEE) since 1998. His research interests are in high-speed communication networks, and QoS issues related to fixed-mobile convergence.

Su Wei Tan is currently a senior lecturer at the Faculty of Engineering, Multimedia University, Malaysia. His research interests include application-layer multicast, peer-to-peer and overlay networks, and quality of service in wireless networks.

Please cite this article as: Yee YC, et al. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. J Network Comput Appl (2007), doi:10.1016/j.jnca.2007.09.002