行政院國家科學委員會專題研究計畫　成果報告

針對 VHE 使用者描述所帶出的服務品質增益的研究

計畫類別：個別型計畫
計畫編號：NSC93-2213-E-009-107-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：國立交通大學電信工程學系(所)

計畫主持人：廖維國

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 10 月 31 日

行政院國家科學委員會補助專題研究計畫 ■成果報告 □期中進度報告

針對 VHE 使用者描述所帶出的服務品質增益的研究

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC　93-2213-E-009-107-
執行期間：2004 年　8 月　1 日至　2005 年　7 月　31 日

計畫主持人：廖維國
共同主持人：
計畫參與人員：陳怡中、劉政澤、廖怡翔、林國瑋、林于彰、黃健智、陳憲良

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
　　　　□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

執行單位：
中　華　民　國　94 年　10 月　31 日

可供推廣之研發成果資料表

□ 可申請專利　　□ 可技術移轉　　　　　　　　　　日期：__年__月__日

國科會補助計畫	計畫名稱： 計畫主持人： 計畫編號：　　　　　　　　學門領域：	
技術/創作名稱		
發明人/創作人		
技術說明	中文：	
	英文：	
可利用之產業 及 可開發之產品		
技術特點		
推廣及運用的價值		

行政院國家科學委員會專題研究計畫成果報告

計畫編號：NSC 93-2213-E-009-107-

執行期限： 2004 年 8 月 1 日至 2005 年 7 月 31 日

主持人：廖維國　國立交通大學電信工程系

計畫參與人員：陳怡中、劉政澤、廖怡翔、林國瑋、林于彰、黃健智、陳憲良

國立交通大學電信工程系

Abstract — **In provisioning the seamless handovers under UMTS/WLAN loose-interworking for real-time services, issues related to the support by the network-layer technologies are how to execute such a vertical handover in a fast and efficient manner as well as how to achieve it in a viable way. We propose a novel scheme called "Designated Crossover Point" (DCP) to resolve these issues within the IP end-to-end QoS architecture. The DCP is an agent at the edge router in the customer's premise to join the old RSVP signaling path and new RSVP signaling path. With the DCP in place, three major advantages are uncovered: First, the vertical handover can be greatly accelerated. Second, by enabling the DCP to handover to the edge router in the customer's network currently visited by the mobile, lower cost can be achieved. Third, there is no modification required for the devices in the provider's network and thus the viability is much improved. We then end up with a topology for resource management in IP backbone for UMTS/WLAN loose-interworking to support the idea of DCP. To determine when to perform the rerouting to eliminate the possible derouting caused by DCPs, we suggest using an adaptive rule. Simulation results show in general the adaptive rule is effective.**

Keywords: IP end-to-end QoS architecture, mobile IP, RSVP, alternating renewal theory, Brownian motion, UMTS/WLAN.

1 Introduction

The dual-mode hand-held devices to connect Wireless LAN (WLAN) and Universal Mobile Telecommunications System (UMTS) now become commercially available. These devices may be capable of connecting to UMTS and WLAN either simultaneously or interchangeably. With this facility in place, there is a widespread customers' expectation that the experienced QoS level of received services could be automatically sustained or leveraged while the charging rate could be lowered.

To make it a reality, the handover between UMTS and WLAN should be enabled so that the mobile at different time instants builds the associations with the stations in the chosen wireless networks. Such type of handover is named as the *vertical handover*. In general, the mobile attempts to connect to WLAN due to its lower charging rate and high-bandwidth provisioning. Consequently, the vertical handover is triggered when the received signal strength from the WLAN changes dramatically.

During the vertical handover, one of the key issues is to reduce the disruption to ongoing mobile's session. Therefore, the challenge with which the service providers are confronted is to find a solution with less cost subject to constrained handover latency, e.g., 100 milliseconds. Besides, the solution should be manageable, with minimum extension to the existing technology and standards, and flexible enough to enable various business models to run such interworking scenario.

In dealing with the aforementioned challenges, we propose a novel architecture which places agents, referred to as *designated crossover point* (DCP), only on the edge routers in the customers' premises to support vertical handovers. Basically for each mobile there is one DCP (in UMTS or WLAN) sitting in the path of the communication to deroute the signaling message and data traffic between the wireless networks for handling the mobility caused by vertical handovers. In essence, our proposed architecture bearing the following features:

1) Diverse business models are allowed;
2) our architecture works around the delay bottleneck of the existing IP end-to-end QoS reference architecture;
3) the rerouting capability, i.e., a process of moving DCP for a mobile to enhance the routing efficiency; the process is hereafter abbreviated as *DCP rerouting*;
4) except during DCP rerouting, the mobility due to the vertical handover is hidden by DCP from the remote end of any connection in progress.

To find a practical solution to minimize the cost introduced by the vertical handover, we derive a rule based on the *alternating renewal theory* for deciding the timing to perform the DCP rerouting.

2 Scope

UMTS is divided into three major parts: the air interface, the UMTS *Terrestrial Radio Access Network* (UTRAN), and the *core network* (CN). The *radio network controller* (RNC) can be considered roughly as *base station controller* (BSC) in *GSM*. The packet-switched portion of the core network in UMTS consists of *serving GPRS support node* (SGSN) and *gateway GPRS supporting node* (GGSN). The SGSN is responsible for session management, producing charging information, and routing packets to correct RNC. A GGSN is like an IP gateway. It implements the firewall and has methods of allocating IP addresses. On the other hand, WLAN is deployed to achieve hotspot coverage. As a result, we study the case where the WLAN coverage is mostly overlapped with the UMTS coverage. Consequently, when the mobile crosses the WLAN coverage, it may perform the vertical handovers between UMTS and WLAN back and forth to let its real-time services best connected.

The scale of our referenced heterogeneous wireless access services consisting of less than ten UMTS networks and thousands of WLANs, and all are connected to an IP backbone. In each UMTS network, there are thousands of base stations, tens of RNCs, and very few SGSNs and GGSNs.

There are two fundamental ways of the interworking in regarding to the handover management proposed in [1]; entitled *tight interworking* and *loose interworking*, and outlined in the followings.

A. Tight Interworking/UMTS mobility management

The WLAN1 and UMTS in Figure 1 show a scenario of the tight interworking. The services offered by a UMTS system can be accessed through WLAN and UMTS address scheme is applied. The vertical handovers are based on the usual UMTS mobility management. The main drawback of such a scenario is the routing inefficiency, i.e., all the traffic to the mobile accessing WLAN should be derouted to pass one of the GGSNs in the UMTS core network.

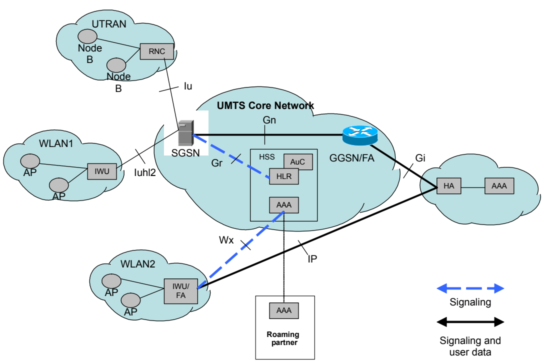

Figure 1. Two scenarios to interwork the UMTS with WLAN.

B. Loose Interworking/Macro-mobility management

The WLAN2 and UMTS in Figure 1 display a loose interworking scenario where SGSN and GGSN in WLAN are avoided. The Mobile IP can be used for the *macro mobility management*: The *interworking unit* (IWU) and GGSN act as the *foreign agents* (FA). Packets destined for the mobile will be transferred to the *home agent* (HA) first and then relayed to the FA. The Mobile IP standards make the *route optimization* possible by sending the address of FA, known as *care of address* (CoA), to the *corresponding node* (CN).

The major advantage of loose interworking is its potential routing efficiency and cheaper product in WLAN. Besides, it bears the flexibility such that the WLANs could be governed by the different administrations. The shortcoming of this scenario is the possible long vertical handover latency introduced by the Mobile IP. The loose interworking scenario is also facilitated with centralized charging and billing system by letting the WLAN communicate with *home location registrar* (HLR) in UMTS via an IWU that has an interface with the *AAA server*.

We are interested in improving the performance of the vertical handover based on Mobile IP between UMTS and WLAN in the loose interworking scenario for real-time services. In the followings, we review the IP end-to-end QoS architecture for the investigation of related issues.

Figure 2 shows an interworking scenario between the customer network, which could be an all-IP UMTS network or a WLAN infrastructure, and an IP backbone. As shown, the customer's network and IP backbone include *QoS servers*, namely QoS server in customer's network (QCS) and QoS server in provider's network (QNS), respectively. The main purpose of QoS server is to accommodate the new signaling protocols, to interwork between customer's network and IP backbone, and to enforce the policy [5]. The multi-protocol label switching (MPLS) technique is deployed in the IP backbone. Note that in the IP world, resource reservation is via *RSVP*. The complete survey on the IP end-to-end QoS architecture is beyond the our scope. For details and essentiality of QoS servers, please refer to [5].

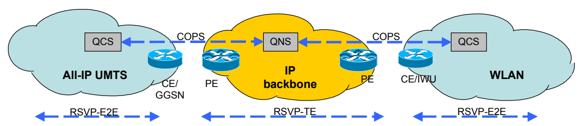

Figure 2. A reference IP end-to-end QoS architecture.

There are two noteworthy points here:

- Unlike in the customer's network, the resource reservation is through the RSVP-TE (*tunnel extension*), rather than the RSVP-E2E (*end-to-end flow*). As a result, reserving the resource over the IP backbone requires the

QCS-QNS negotiation, which may be via the protocol *COPS* and could still be at least few seconds in the near future.

- The micro-flow is recognizable only by *customer's edge router* (CE), those routers behind the CE in the customer's networks, and possibly the *provider's edge routers* (PEs). Those *core routers* (i.e., those not PEs in the IP backbone), indicated as *P*, enforce the *QoS policy* only on the aggregated flows under IETF Diffserv architecture.

To simplify our discussion, we let IWU be CE in WLAN and let GGSN be CE in UMTS. In what follows, we refer to the CE of the network where an entity, i.e., mobile, CN, or HA, is currently located as "*CE of entity*."

3 Fast Vertical Handover Via Designated Crossover Points

When a vertical handover happens, the mobile changes its association of the network and CoA. As a result the QCS-QNS negotiation for the resource reservation over IP backbone to the "CE of mobile" is mandatory and thus induces a high latency of vertical handover and serious interruption of real-time services.

To relieve the aforementioned problem, two approaches of latency-hiding techniques are possible: 1) delaying the vertical handover, i.e., enabling the vertical handover after the new path, which is requested on-demand, is established, and 2) multicasting the streams to the base stations where the mobile currently and possibly visits, e.g., *MRSVP* [10] or *mobility support based on IP multicast* [3]. Under the IP end-to-end QoS architecture, the efficiencies of both approaches depend on the location to join the old and new paths, as illustrated in the followings:

3.1 Endpoint

When joining the old and new paths in the endpoint (CN or HA), different drawbacks are induced with respective to the latency hiding technique:

- In delaying the vertical handover, the latency involves the part induced by QCS-QNS negotiation. As such, the real-time service is interrupted if the signal from the currently associated base station or AP is suddenly degraded to an unacceptable level so as to force a vertical handover.
- In multicasting, a reasonable scheme to avoid the above-mentioned service interrupt is that whenever mobile detects the beacon signal from a WLAN, the real-time stream will be multicast to the GGSN of the subscribed UMTS and "CE of WLAN." However, "multicasting from the endpoint" implies pre-allocating an extra pipes for the real-time services in-between the "CE of CN" (or HA) and "CE of mobile."

3.2 Crossover Router

To cut down the latency of the vertical handover, we can simply join the old and new paths at the crossover router, e.g., adding an agent, such as *second home agent* [2] and *RSVP mobile proxy* [6], to translate the *flow identifier.* The original proposals on such a solution are to join at the optimal crossover point [9] when dealing with macro-mobility or a fixed gateway [2, 6] when dealing with the micro-mobility problem. However, within the IP end-to-end QoS architecture more considerations are needed to be taken. Possible crossover points to perform the join operation under the reference IP end-to-end QoS architecture. As shown, the mobile moves from the network 10.1.0.0 to 10.2.0.0. The optimum crossover point is (a), which is a core router. Secondly, (b) and (c) are a PE and a CE, respectively. For the cases (b) and (c), the new path should be *derouted*, performed by PE attaching to and CE in network 10.2.0.0, respectively, to meet the *designated crossover point* (DCP). At the first glimpse, placing the crossover point at (a) induces best routing efficiency. However, abiding by the case (a) or (b) raises the following concerns:

- *Longer delay*: The reservation needs the QCS-QNS negotiation. For the case (c), there may be *pre-allocated pipe,* i.e., *label switched path (LSP) tunnel*, between the old CE/FA and new CE/FA (will be discussed shortly). Such a negotiation process is not needed to reserve the un-used resource over the pre-allocated pipe.
- *Scalability concern*: The core routers shall be enhanced to recognize the micro flow (e.g., [4]).
- *Multiple reservations*: The cases (a) and (b) bear the possibility that the crossover points are not unique when there exists several ongoing sessions. In case (a), the optimum point depends on the locations of CNs. In case (b), the old CE/FA may connect to two PEs (i.e., *multi-homing*) where two ongoing sessions flow, respectively.

With the above reasoning, we focus on placing the DCP in CE, as in case (c).

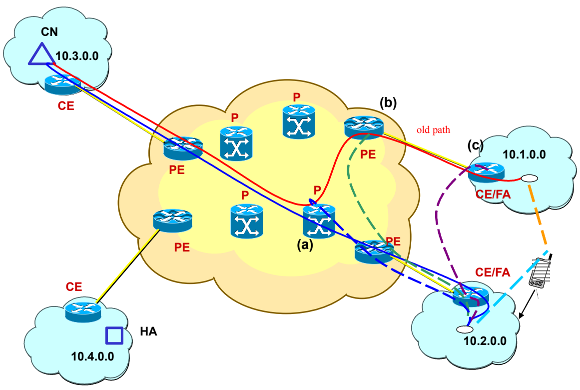

Figure 3. Three possible locations for placing the designated crossover point (DCP).

To overcome the resultant routing inefficiency when DCP is not on the "CE of mobile," we enable the process, referred to as DCP rerouting in Introduction, to move the DCP to the “CE of mobile.” When incorporating with Mobile IP, the address of CE with DCP in place is always the CoA registered and bound at the HA and CN, respectively. A DCP rerouting then includes the operation to use the address of “CE of mobile” as new CoA. Consequently, the derouting can be avoided. In the next section, we will address the timing issue to perform DCP rerouting for achieving lower cost.

In summary, with the reference IP end-to-end QoS architecture, a reasonable resource management scheme over IP backbone for UMTS/WLAN loose interworking may pre-allocate the pipes between CEs to avoid the lengthy QCS-QNS negotiation. While there are a couple of possible topologies, such as *fully connected* and *multi-staged* (see Chapter 7 in [7]), can be used to connect all the CEs. However, we advocate the *hub-and-spoke with redundancy* topology where the hubs are the GGSNs and spokes are the “CEs of WLANs.” As depicted in Figure 4, there exist two UMTS networks, UMTS_1 and UMTS_2, three WLANs, indicated by WLAN_i, and many wired LANs, indicated by LAN_i. The pipes across the IP backbone are only pre-allocated for each UMTS-WLAN pair to support the vertical handoff due to the followings:

- The WLAN is small-scaled and thus cannot afford such a large amount of pipes (consider thousands of WLANs and thousands of LANs). Most mobiles are supposed to subscribe to a single UMTS operator and a relatively large amount of traffic will be handovered between the GGSN of the subscribed UMTS and “CE of WLAN”. Therefore, it is valuable to pre-allocate the pipes to/from the GGSNs for the WLAN because of the potentially high utilization.
- In contrast, the pipes pre-allocated between LANs (WLAN or wired LAN) are expected to be rarely utilized whereby the LAN-LAN pipe is reserved on-demand for each individual real-time service.

Accordingly, when CN is in a wired LAN, the operation of DCP rerouting (rather than vertical handover) includes the QCS-QNS negotiation for setting up the new pipe from the “CE of CN” to the “CE of mobile” as the mobile handovers in-between WLAN and UMTS. We assume that the pre-allocated pipe can only be used for de-routing the handover traffic between UMTS and WLAN.

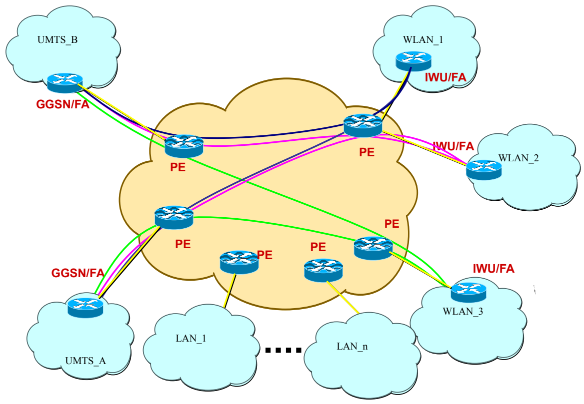

Figure 4. A reasonable pipe pre-allocation for handover management in WLAN/UMTS loose coupling.

4 Reducing Expected Cost By Delaying DCP Rerouting

We formulate the expected cost over infinite horizon with delaying the DCP rerouting. We assume that the pre-allocated pipes are engineered such that blocking a vertical handover due to limited bandwidth in the fixed wired network rarely happens.

The followings show the related parameters for delaying the DCP rerouting by applying *dwell timer*:

- τ: Duration to smooth the DCP rerouting.
- μ: Initial dwell timer value.
- L_1: Connection over pipe between CE of CN and CE of UMTS or WLAN.
- L_2: Connection over the pre-allocated pipe between GGSN of UMTS and “CE of WLAN.”

We depict the state chart for applying dwell timer to DCP rerouting in the Figure 5. The name of the state indicates the number of connections of certain types in the IP backbone being used for a mobile’s real-time session. Likewise, the state “$2L_1+L_2$” stands for the session currently using two L_1 connections and one L_2 connection. The label of transition stands for the event firing the transition. The initial state is in “L_1”, which means the DCP will be placed in the “CE of mobile” and thus only the pipe for the type “L_1” connection is used over the IP backbone. Upon a vertical handover, an extra connection of type “L_2” is added to deroute the traffic. One of the actions in this transition is to set the dwell timer to be μ. If there is another vertical handover prior to the expiration of dwell timer, which means the mobile switches back its association to the wireless network where DCP is located, the connection for derouting the traffic is removed immediately. As the dwell timer expires, the DCP rerouting is invoked. During the DCP rerouting, pipes over IP backbone between “CE of CN” and “CE of mobile” are being set up. To tolerate the setup delay, which is smaller than τ, the original connections will not be removed until τ time units later to smooth the DCP rerouting.

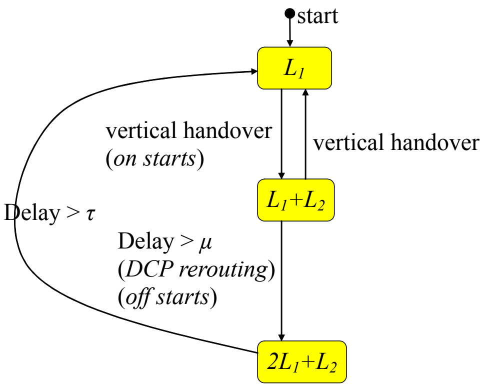

Figure 5. The state transition for DCP rerouting.

We apply the *alternating renewal process* by assuming that the holding time of a session is sufficiently long, i.e., the vertical handover is relatively frequent. We also assume that during a time interval of length τ, the probability to perform a vertical handover is sufficiently small. Therefore, we ignore its effect in our modeling. When applying the dwell timer, an "on" period starts immediately after entering into the state "L_1+L_2" upon a vertical handover. The "off" period starts immediately after the timer expiration or a vertical handover. Therefore, the length of "on" period does not depend on the history while the length of the "off" period depends on the length of previous "on" period, i.e., if the length of previous "on" period is equal to μ, then it is DCP rerouting with probability one, not vertical handover, resulting in the end of the period. Notice that the sequences of "on" period and "off" period can be modeled into an i.i.d. processes, respectively. Thus, our formulation fits nicely into an ordinary alternating renewal process (see [8]). Figure 6 shows an example to display the idea of our formulation.

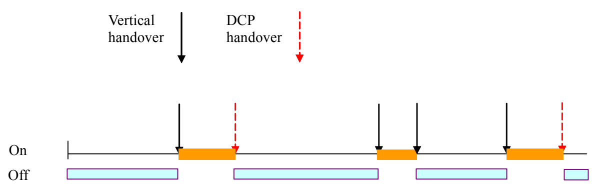

Figure 6. The alternating renewal process.

Let U be the random variable standing for the time difference between two consecutive vertical handovers. The "on" period is then the random variable $\min(U,\mu)$. The "off" period is the random variable of $I(U_1>\mu)\times(U_1-\mu) + I(U_1 \leq \mu)\times U_2$ where U_1 and U_2 are two mutually independent copies of U.

Theorem 1: The expected cost over infinite horizon is then yielded by the renewal theory as follows:

$$E(COST) = c(L_1) + \frac{c(L_2)E[on] + (c(L_1) + c(L_2))\tau P(U > \mu)}{E[on] + E[off]},$$

where $E[on]$ and $E[off]$ denote the expected lengths of an "on" period and that of "off" period, respectively, and $c(L_i)$ denotes the cost of a connection with the type L_i.

Proof: The result is obtained by directly applying the renewal theory.

□

Remark 1: The expected cost consists of three major parts: the constant cost $c(L_1)$, the extra cost during on-period, i.e., $c(L_2)$, and the extra cost due to the DCP rerouting.

Therefore, by taking out the constant $c(L_1)$, our criteria turns to minimizing the following *expected overhead*:

$$E(overhead) = \frac{c(L_2)E[on] + (c(L_1) + c(L_2))\tau P(U > \mu)}{E[on] + E[off]}.$$

Now we turn to obtaining $E[on]$:

$$\begin{aligned} E[on] &= E[\min(U,\mu)] \\ &= E[I(U \leq \mu)U + \mu I(U > \mu)] \\ &= E[I(U \leq \mu)U] + \mu P(U > \mu). \end{aligned}$$

For $E[off]$, we have the following derivation:

$$\begin{aligned} E[off] &= E[I(U_1 > \mu)\times(U_1 - \mu)] + P(U_1 \leq \mu)E[U_2] \\ &= E[I(U > \mu)U] - \mu P(U > \mu) + P(U \leq \mu)E[U] \\ &= E[U] - E[I(U \leq \mu)U] - \mu P(U > \mu) + P(U \leq \mu)E[U]. \end{aligned}$$

Therefore, the expected overhead due to handover turns to the followings:

$$\frac{c(L_2)\left[(E[I(U \leq \mu)U]) + \left(\mu + \left(\frac{c(L_1)}{c(L_2)} + 1\right)\tau\right)P(U > \mu)\right]}{E[U](1 + P(U \leq \mu))},$$

which converges to $c(L2)/2$ as $\mu \rightarrow \infty$ if $\mu P(U>\mu) \rightarrow 0$ as $\mu \rightarrow \infty$ and converges to the following as $\mu \rightarrow 0$:

$$\frac{(c(L_1) + c(L_2))\tau}{E[U]}.$$

Hence, under the assumption of $\mu P(U>\mu) \rightarrow 0$ as $\mu \rightarrow \infty$, a simplified rule to set the dwell timer to reduce the expected overhead is described as follows:

$$\mu = \begin{cases} 0 & \frac{c(L_2)}{c(L_1)} > \frac{2\tau}{E[U] - 2\tau}, \\ \infty & otherwise. \end{cases} \qquad (1)$$

Note that when μ=0, the DCP rerouting will be immediately taken once the vertical handover is performed and when $\mu=\infty$, the DCP rerouting will never be performed. The simplifed rule works well when the expected overhead is a function of μ with at most one extreme point which is a maximal point over the domain [0, ∞]. Hence the minimum will take place only when μ equals either zero or infinity. In general, as will be observed shortly, when $P(U>\mu)$ does not quickly vanish (or equivalently pdf of U is not large for small μ) helps the corresponding function to have such a shape.

By observing the above derivations, obtaining the distribution of $P(U \leqq \mu)$ is key to finding a more accurate dwell timer value μ to minimize the expected overhead. We believe that the corresponding distribution is highly personalized and should be maintained in the user profile. In the followings, we consider two extreme types of motions, called *random-walk motion* and *drifted motion*.

In the case of the drifted motion, the vertical handover occurs mainly because of crossing different boundaries. We suggest using the empirical distribution instead and finding a distribution with close-form CDF and finite mean to best match the empirical distribution. Suppose we use the exponential distribution with parameter λ. Thus we have the following expected overhead:

$$\frac{c(L_2)(1-e^{-\lambda\mu}+\lambda\mu e^{-\lambda\mu})+(c(L_1)+c(L_2))\lambda\tau e^{-\lambda\mu}}{2-e^{-\lambda\mu}}.$$

In the case of random-walk motion, the vertical handover takes place largely because the mobile moves across the same boundary back-and-forth. We assume that the vertical handover is based on the hysterics, i.e., the handover will be performed if the mobile is in the corresponding coverage area and with at least δ unit apart from the line of equal RSS to avoid the possible ping-pong effect when traveling along a line as shown in Figure 6. Also, we let the distance between two cell boundaries be a.

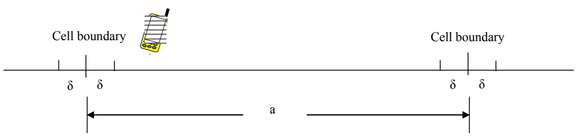

Figure 7. The one-dimensional motion.

Suppose mobile follows the *Brownian motion*, denoted by B_t. Then U can be restated as follows:

$$U = \inf\{t : B_t \notin (-2\delta, a)\}.$$

Therefore, the reflection principle [8] and the assumption of $a >> \delta$ yield the followings:

$$\begin{aligned}
&P(U<\mu) = P(B_U=-2\delta, U<\mu)+P(B_U=a, U<\mu)\\
&= P(U<\mu \mid B_U=-2\delta)P(B_U=-2\delta)\\
&+ P(U<\mu \mid B_U=a)P(B_U=a)\\
&\approx P(\inf\{t : B_t=-2\delta\}<\mu)P(B_U=-2\delta)\\
&+ P(U<\mu \mid B_U=a))P(B_U=a)\\
&= P(\inf\{t : B_t=-2\delta\}<\mu)\frac{a}{a+2\delta}+P(U<\mu \mid B_U=a))\frac{2\delta}{a+2\delta}\\
&\approx P(\inf\{t : B_t=-2\delta\}<\mu)\frac{a}{a+2\delta}\\
&= 2P(B_\mu \geq 2\delta)\frac{a}{a+2\delta} = 2\frac{a}{a+2\delta}\int_{2\delta}^{\infty}(2\pi\mu)^{-1/2}\exp^{-x^2/2\mu}\,dx.
\end{aligned}$$

It is well known that $E[U]$ is infinite if using this approximation (actually $E[U]=2\delta a$). Unfortunately, to obtain the optimal μ, numerical evaluation is inevitable.

Note-worthily, the minimum expected overhead is always less than c(L_2)/2, i.e., when $\mu \to \infty$. Thus, as long as c(L_1) > 2c(L_2), the expected overhead is always less than c(L_1) and thus our proposed scheme always has the cost smaller than that of multicasting from the endpoint.

Chances are that the mobile has no sufficient record to obtain the accurate $E[U]$ to apply the simplified rule (Equation (1)). We suggest using the adaptive rule which replaces $E[U]$ by the maximum of $2\tau+\varepsilon$ and sample mean of U in the simplified rule. More specifically, we sample each lifetime. Suppose there are n samples, denoted as $U_1,\ldots,U_n$. Let the time interval between the last vertical handover and now being t. The DCP rerouting will be initiated if the following inequality holds true:

$$\frac{c(L_2)}{c(L_1)} > \frac{2\tau}{\max\left\{\sum_{i=1}^{n}\frac{U_i}{n}, \sum_{i=1}^{n}\frac{U_i}{n+1}+\frac{t}{n+1}, 2\tau+\varepsilon\right\}-2\tau}.$$

Or equivalently we have the timer value as follows:

$$\mu = \begin{cases} 0 & if \sum_{i=1}^{n}\frac{U_i}{n} > 2\tau\frac{c(L_1)}{c(L_2)}+2\tau, \\ (n+1)\left(2\tau\frac{c(L_1)}{c(L_2)}+2\tau-\sum_{i=1}^{n}\frac{U_i}{n+1}\right) & otherwise. \end{cases}$$

(2)

It is trivial to see that the proposed adaptive rule will converge to the simplified rule with probability 1 as n goes to infinity.

5 Simulation

In this section, we conduct simulation for the different motions. The values of parameters in the simulation settings are listed in Table 1. The threshold for c(L_2)/c(L_1) to apply the proposed simplified rule is 2*1.67/(10-2*1.67) which is approximately 0.5. We also set the holding time of the real-time service as 10 minutes. For each setting, we conduct the simulation 100 times and obtain the average of them.

Parameter	Value
τ	1.67

	seconds
E[U]	10 seconds
δ	0.5
a	10

Table 1. Parameters to be used for determining timer value.

In the simulation, U is exponentially distributed in drifted motion and U is the stopping times of Brownian motion, as stated in the previous section. Figure 8 shows the corresponding pdfs. As displayed, the pdf of random-walk motion is larger than the drifted motion when U is in the range [0, 3] but exhibits heavier tail.

Figure 9 displays the simulation results for the drifted motion. As shown, the simplified rule well applies to this case because each curve exhibits at most one extreme point, which is a maximal point, over the domain $[0,\infty)$. As shown in Figure 10, the simplified rule is generally well applicable to the random-walk motion except when $c(L_2)/c(L_1) = 0.5$, the best μ is around 10 seconds.

It is noteworthy that our proposed DCP scheme outperforms the solution of multicasting from the endpoint in each simulated case, including $c(L_2)=c(L_1)$, i.e., the simulated average overhead is always less than one. Besides, the simplified rule significantly reduces the averaged overhead in two extremes of the ratios $c(L_2)/c(L_1)$.

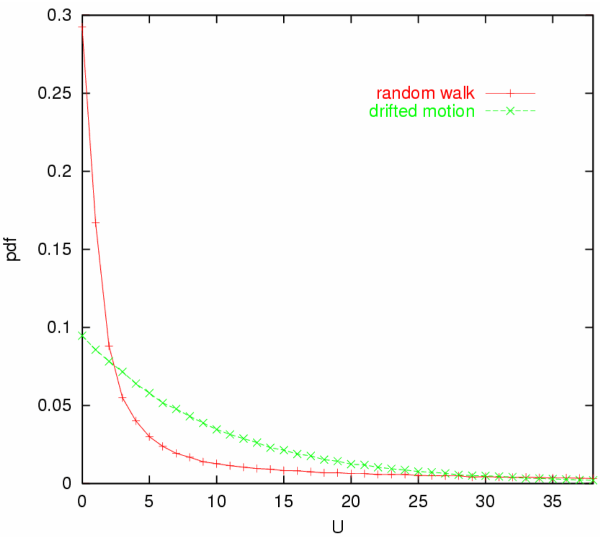

Figure 8. The probability distribution of two motions.

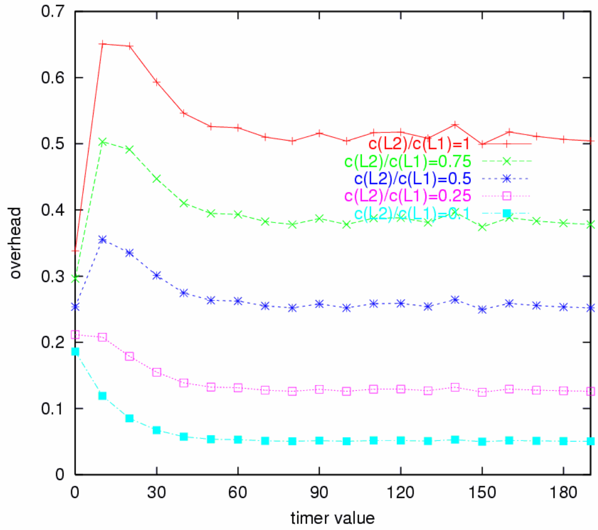

Figure 9. The simulated average overhead for drifted motion.

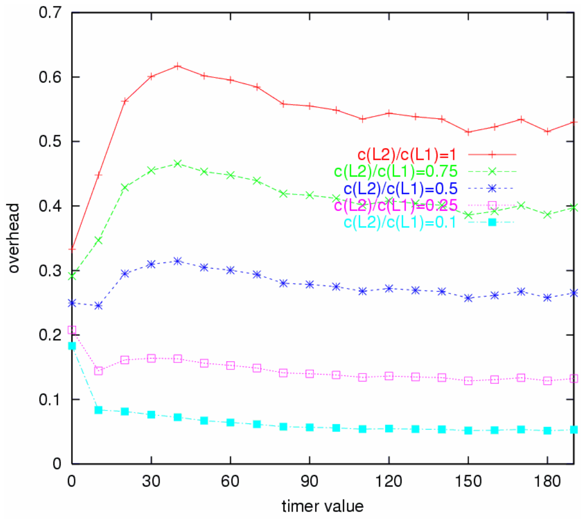

Figure 10. The simulated average overhead for random-walk motion.

The last simulation is to verify the effectiveness of the adaptive rule in Equation (2). As shown in Figure 11, the averaged overheads induced by the adaptive rule (labeled with random-walk and drifted) are slightly worse than the expected overhead by the simplified rule (labeled with simple rule) for both motions.

6 Self Assessment

This project reviews the issues of supporting the fast vertical handover for real-time service from the network architectural aspect. It identifies that the QCS-QNS negotiation could be the latency bottleneck of vertical handover. To workaround such a bottleneck, it introduces the concept of designated crossover points (DCP) and discusses its placement and the potential benefit to delay the DCP rerouting. We also derive an adaptive rule to set such a delay. In our simulation, the adaptive rule works well for the motion with exponential-distributed lifetime and in most cases for the Brownain motion.

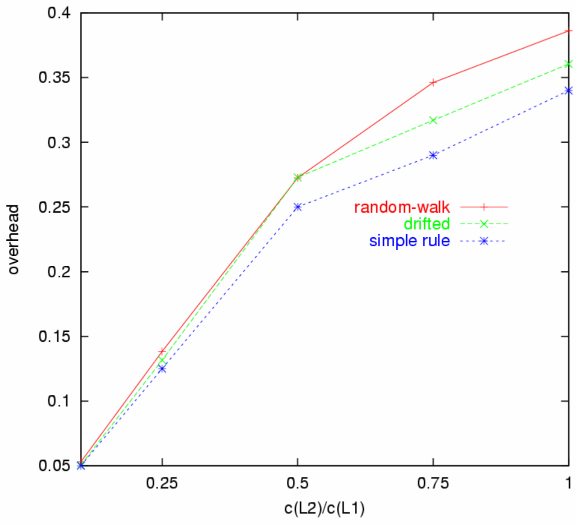

Figure 11. The simulated average overhead for verifiying the adaptive rule.

REFERENCES

[1] ETSI TR 101 957 v1.1.1 (2001-08) "Broadband Radio Access Networks (BRAN); HIPERLAN TYPE 2; Requirements and Architectures for Interworking between HIPERLAN/2 and 3rd Generation Cellular Systems."

[2] A. De Carolis et al, "QoS-Aware Handover: Secondary Home Agent", IETF draft-decarolis-qoshandover-02.txt, April 2000.

[3] W.T. Chen and L.C. Huang, "RSVP Mobility Support: A Signaling Protocol for Integrated Services Internet with Mobile Hosts," Infocom 2000.

[4] F. M. Chiussi, D. A. Khotimsky, and S. Krishnan, "Mobility Management in Third Generation All-IP Networks," IEEE Communication Magazine, Sep 2002, pp. 124-135.

[5] V. Fineberg, "A Practical Architecture for Implementing End-to-End QoS in an IP Network", IEEE Communications Magazine, January 2002, Page(s): 122-130.

[6] S. Paskalis, A. Kaloxylos, E. Zernas, and L. Merakos, "RSVP Mobility Proxy," Internet draft, Dec. 2001.

[7] I. Pepelnjak and J. Guichard, MPLS and VPN Architectures, Cisco Press, 2000.

[8] S. M. Ross, "Stochastic Process, 2nd ed.," 1996 John Wiley & Sons.

[9] Q. Shen et al., "On Providing Flow Transparency Mobility Support for IPv6-based Real-Time Services, " Proc. MoMuC 2000, Tokyo, Japan, Oct. 2000.

[10] A. K. Talukdar, B.R. Badrinath, and A. Achaya, "MRSVP: A Reservation Protocol for an Integrated Services Packet Network with Mobile Hosts," Dept. Comp. Sci. tech rep. TR-337, Rutgers Univ.