Supporting vertical handover between universal mobile telecommunications system and wireless LAN for real-time services

W.-K. Liao and Y.-C. Chen

Abstract: The provisioning of the seamless handovers for real-time services in-between UMTS/WLAN loose-interworking by the network-layer technologies has been studied. Issues 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. To resolve such issues, a novel scheme called 'designated crossover point' (DCP) within the existing IP end-to-end Quality of Servce (QoS) architecture has been proposed. The DCP is an agent at the edge router in the customer's premise to join the old RSVP signalling path and new RSVP signalling path. With the DCP in place, the vertical handover can be greatly accelerated. Besides, by enabling the DCP to handover to the edge router in the customer's network currently visited by the mobile, lower charging cost can be achieved. Finally, there is no modification required for the devices in the provider's network and thus the viability is much improved. It is 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 remove the possible derouting caused by DCPs, use of an adaptive rule has been suggested. Simulation results show in general that the adaptive rule is effective.

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 because because of 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, for example, 100 ms. 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 signalling 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, that is, 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 because of the vertical handover is hidden by DCP from the remote end of any connection in progress.

To find a practical solution to minimise 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.

The rest of the paper is organised as follows: In Section 2, we give the scope of the problem. In Section 3, we propose architecture with DCP. In Section 4, we analyse the profit of applying dwell timer concept to the DCP rerouting and propose an adaptive rule. We conduct simulation in Section 5. We draw the conclusion in Section 6.

2 Scope

UMTS is divided into three major parts: the air interface, the UMTS terrestrial radio access network and the core network (CN). The radio network controller (RNC) can be considered roughly as base station controller in GSM. The packet-switched portion of the CN 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 following.

2.1 Tight interworking/UMTS mobility management

The WLAN1 and UMTS in Fig. 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, that is, all the traffic to the mobile accessing WLAN should be derouted to pass one of the GGSNs in the UMTS CN.

2.2 Loose interworking/macro-mobility management

The WLAN2 and UMTS in Fig. 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
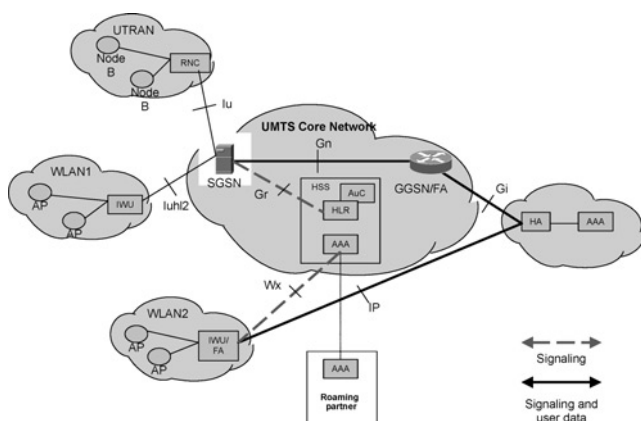
mobile will be transferred to the home agent (HA) first and then relayed to the FA. The Mobile IP standards make the route optimisation possible by sending the address of FA, known as care of address (CoA), to the corresponding node (CN).

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 centralised charging and billing system by letting the WLAN communicate with home location registrar 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 following, we review the IP end-to-end QoS architecture for the investigation of related issues.

Fig. 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 signalling protocols, to interwork between customer's network and IP backbone, and to enforce the policy [2]. The multiprotocol label switching (MPLS) technique is deployed in the IP backbone. Note that in the IP world, resource reservation is via resource reservation protocol (RSVP). The complete survey on the IP end-to-end QoS architecture is beyond the scope of the paper. For details and essentiality of QoS servers, please refer to [2].

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 common open policy service (COPS) protocol and could still be at least few seconds in the near future.
- The micro-flow is recognisable 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, that is, mobile, CN or HA, is currently located as 'CE of entity'.

3 Fast vertical handover via DCPs

When a vertical handover happens, the mobile changes its association of the network and CoA. As a result, the
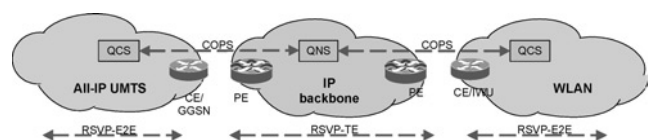

Fig. 1 *Two scenarios to interwork the UMTS with WLAN*

Fig. 2 *Reference IP end-to-end QoS architecture*

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, that is, 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, for example, *MRSVP* [3] or mobility support based on IP multicast [4]. 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 following

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 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, for example, adding an agent, such as second home agent [5] 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 [7] when dealing with macro-mobility or a fixed gateway [5, 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 in Fig. 3, 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 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 case (c), there may be pre-allocated pipe, that is, label switched path 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 recognise the micro flow (e.g. [8]).
- *Multiple reservations:* The cases (a) and (b) bear the possibility that the crossover points are not unique when
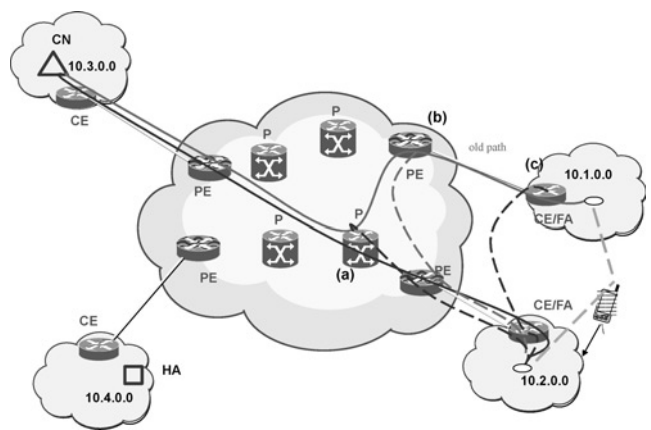
Fig. 3 *Three possible locations for placing the DCP*

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).

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 Section 1, 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 [9]), 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 Fig. 4, there exist two UMTS networks, UMTS_1 and UMTS_2, three WLANs, indicated by WLAN_i, and many wired
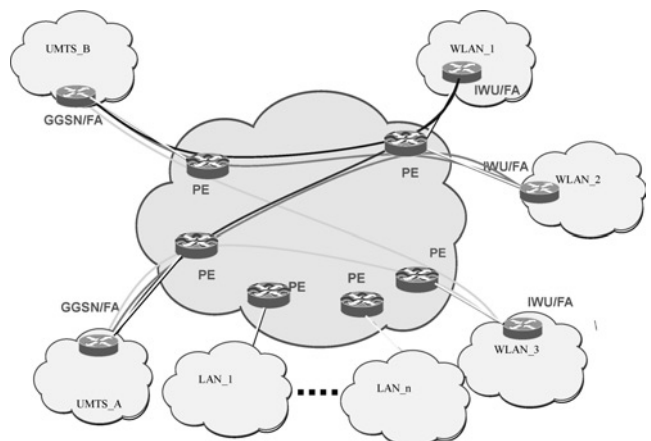

Fig. 4 *Reasonable pipe pre-allocation for handover management in WLAN/UMTS loose coupling*

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 because of the following:

- 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 utilisation.
- In contrast, the pipes pre-allocated between LANs (WLAN or wired LAN) are expected to be rarely utilised 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.

4 Reducing expected cost by delaying DCP rerouting

We formulate the expected charging cost over infinite horizon with delaying the DCP rerouting. We assume that the pre-allocated pipes are engineered such that blocking a vertical handover because of limited bandwidth in the fixed wired network rarely happens. In our following discussion, we focus on the handover involving one UMTS cell and WLAN cell first.

4.1 One UMTS cell and one WLAN cell

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 Fig. 5. The name of the state indicates the number of connections of certain types in the IP backbone
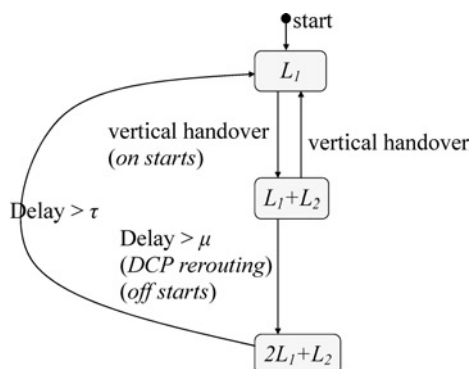
being used for a mobile's real-time session. Likewise, the state '$2L_1 + L_2$' stands for the session currently using two L_1 connections and one L_2 connection. The label of transition stands for the event firing the transition. The initial state is in 'L_1', which means the DCP will be placed in the '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.

We apply the alternating renewal process by assuming that the holding time of a session is sufficiently long, that is, the vertical handover is relatively frequent. We also assume that during a time interval of length τ, the probability of performing a vertical handover is sufficiently small. Therefore we ignore its effect in our modelling. 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 whereas the length of the 'off' period depends on the length of previous 'on' period, that is, if the length of previous 'on' period is equal to μ, then it is DCP rerouting with probability 1, not vertical handover, resulting in the end of the period. Notice that the sequences of 'on' period and 'off' period can be modelled into an i.i.d. processes, respectively. Thus, our formulation fits nicely into an ordinary alternating renewal process (see [10]). Fig. 6 shows an example to display the idea of our formulation.

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 $I(U_1 > \mu) \times (U_1 - \mu) + I(U_1 \le \mu) \times U_2$ 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

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

where $E[\text{on}]$ and $E[\text{off}]$ 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.
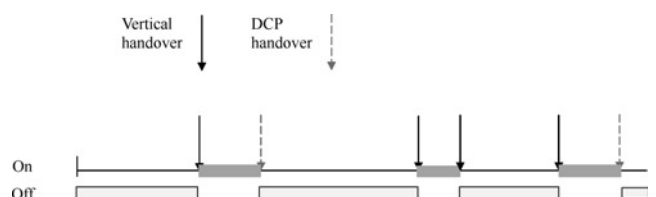

Fig. 5 *State transition for DCP rerouting*

Fig. 6 *Alternating renewal process*

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, that is, $c(L_2)$, and the extra cost because of the DCP rerouting.

Therefore by taking out the constant $c(L_1)$, our criteria turn to minimise the following expected overhead

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

Now we turn to obtain $E[\text{on}]$

$$E[\text{on}] = E[\min(U, \mu)] = E[I(U \leq \mu)U + \mu I(U > \mu)]$$
$$= E[I(U \leq \mu)U] + \mu P(U > \mu)$$

For $E[\text{off}]$, we have the following derivation

$$E[\text{off}] = E[I(U_1 > \mu) \times (U_1 - \mu)] + P(U_1 \leq \mu)E[U_2]$$
$$= E[I(U > \mu)U] - \mu P(U > \mu) + P(U \leq \mu)E[U]$$
$$= E[U] - E[I(U \leq \mu)U] - \mu P(U > \mu)$$
$$+ P(U \leq \mu)E[U]$$

Therefore the expected overhead because of handover turns to the following

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

which converges to $c(L_2)/2$ as $\mu \to \infty$ if $\mu P(U > \mu) \to 0$ as $\mu \to \infty$ and converges to the following as $\mu \to 0$

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

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

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

Note that when $\mu = 0$, 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 $[0, \infty]$. 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 $P(U \leq \mu)$ is key to find a more accurate dwell timer value μ to minimise the expected overhead. We believe that the corresponding distribution is highly personalised and should be maintained in the user profile. In the following, 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 cumulative distributed function (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, that is, 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 travelling along a line as shown in Fig. 7. Also, we let the distance between two cell boundaries be a.

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 [10] and the assumption of $a \gg \delta$ yield the following

$$P(U < \mu)$$
$$= P(B_U = -2\delta, U < \mu) + P(B_U = a, U < \mu)$$
$$= P(U < \mu | B_U = -2\delta)P(B_U = -2\delta)$$
$$\quad + P(U < \mu | B_U = a)P(B_U = a)$$
$$\simeq P(\inf\{t : B_t = -2\delta\} < \mu)P(B_U = -2\delta)$$
$$\quad + P(U < \mu | B_U = a))P(B_U = a)$$
$$= P(\inf\{t : B_t = -2\delta\} < \mu)\frac{a}{a + 2\delta}$$
$$\quad + P(U < \mu | B_U = a))\frac{2\delta}{a + 2\delta}$$
$$\simeq P(\inf\{t : B_t = -2\delta\} < \mu)\frac{a}{a + 2\delta}$$
$$= 2P(B_\mu \geq 2\delta)\frac{a}{a + 2\delta}$$
$$= 2\frac{a}{a + 2\delta}\int_{2\delta}^{\infty}(2\pi\mu)^{-1/2}e^{-x^2/2\mu}\,dx$$

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. Noteworthily, the minimum expected overhead is always less than $c(L_2)/2$, that is, 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 (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
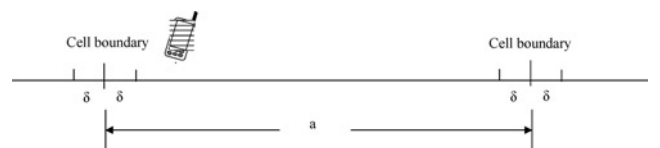

Fig. 7 *One-dimensional motion*

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

Parameter	Value
τ	1.67 s
$E[U]$	10 s
δ	0.5
a	10

inequality holds true

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

Or equivalently we have the timer value as follows

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

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

### 4.2 One UMTS cell and many WLAN cells

The scenario of more than one WLAN cells with overlapped coverage, which is all covered by a single one UMTS cell, has been studied in [11]. In such a scenario, existing pipes link not only the UMTS and WLAN but also WLANs. Basically, based on the state chart in Fig. 5 and by applying the theory of semi-Markov decision process, we extend our approach to find a near optimal decision rule. The simulation results show that DCP scheme also performs well in such a scenario. Because of the space consideration, we omit the discussion.

5 Simulation

We conduct simulation for the different motions for scenario of one UMTS cell and one WLAN cell. 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
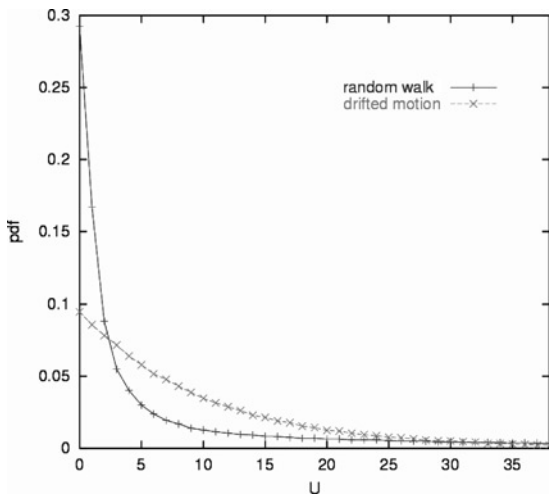
**Fig. 8** *Probability distribution of two motions*
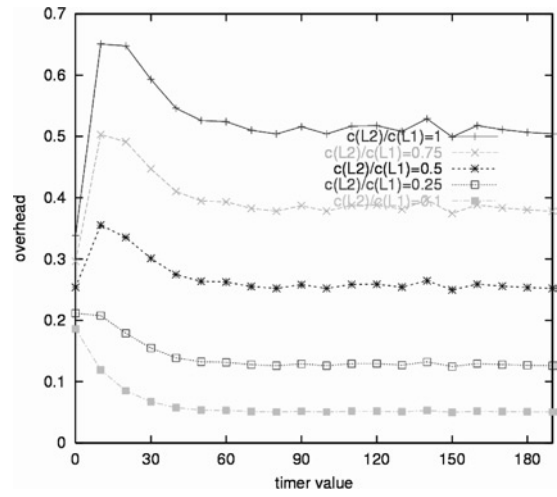
Fig. 9 *Simulated average overhead for drifted motion*

approximately 0.5. We also set the holding time of the real-time service as 10 min. For each setting, we conduct the simulation 100 times and obtain the average of them. 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. Fig. 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.

Fig. 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,∞). As shown in Fig. 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 s.

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

The last simulation is to verify the effectiveness of the adaptive rule in (2). As shown in Fig. 11, the averaged overheads induced by the adaptive rule (labelled with random-walk and drifted) are slightly worse than the expected overhead by the simplified rule (labelled with simple rule) for both motions.
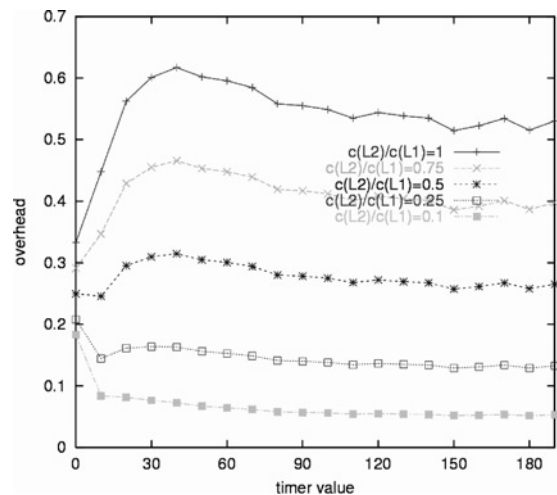
Fig. 10 *Simulated average overhead for random-walk motion*
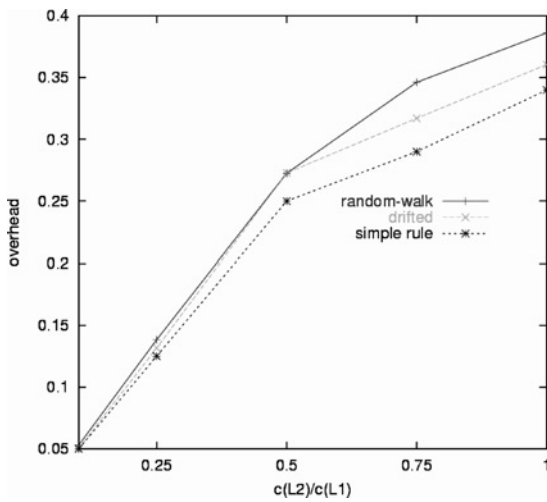
Fig. 11 *Simulated average overhead for verifiying the adaptive rule*

6 Conclusion

This paper 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 DCPs 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.

7 Acknowledgment

8 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 Fineberg, V.: 'A practical architecture for implementing end-to-end QoS in an IP network', *IEEE Commun. Mag.*, 2002, pp. 122–130
3 Talukdar, A.K., Badrinath, B.R., and Achaya, A.: 'MRSVP: a reservation protocol for an integrated services packet network with mobile hosts', Dept. Comp. Sci. tech rep. TR-337, Rutgers University
4 Chen, W.T., and Huang, L.C.: 'RSVP mobility support: a signaling protocol for integrated services internet with mobile hosts', Infocom 2000
5 Carolis, De., *et al.*: 'QoS-Aware Handover: Secondary Home Agent', IETF draft-decarolis-qoshandover-02.txt, April 2000.
6 Paskalis, S., Kaloxylos, A., Zernas, E., and Merakos, L.: 'RSVP mobility proxy', Internet draft, December 2001
7 Shen, Q., *et al.*: 'On providing flow transparency mobility support for IPv6-based real-time services'. Proc. MoMuC 2000, Tokyo, Japan, October 2000
8 Chiussi, F.M., Khotimsky, D.A., and Krishnan, S.: 'Mobility management in third generation all-IP networks', *IEEE Commun. Mag.*, 2002, pp. 124–135
9 Pepelnjak, I., and Guichard, J.: 'MPLS and VPN architectures' (Cisco Press, 2000)
10 Ross, S.M.: 'Stochastic process' (John Wiley & Sons, 1996, 2nd edn.)
11 Chen, Y.C.: 'Supporting vertical handover between UMTS and WLAN for real-time services', Master Thesis, National Chiao Tung University, 2004