

Multi Layer Rules Based Framework for Vertical Handoff

A novel approach to trigger Vertical Handoff in multihomed devices

Abhijit Sur, Douglas C. Sicker

University of Colorado at Boulder, USA

{abhijit.sur, douglas.sicker} @colorado.edu

Abstract— Seamless interoperability between two dissimilar networks require handoff from one network to the other. Such handoffs are known as vertical handoffs. Vertical handoff introduces a shift in the approach to handoffs. It deals with handoffs between dissimilar networks, such as from an access point to a base station or vice versa. A horizontal handoff, on the other hand, involves the scenario wherein the handoff is within the same network such as handoffs between cells. Most of the existing research proposes to implement vertical handoff on an overlay network, wherein a user moves from one geographic location to another, such as moving from a WLAN coverage area to a cellular coverage area, or vice versa. This research discusses the need to trigger a vertical handoff based on changes in the mobile environment independent of geographic changes of a user. We develop a novel framework that executes vertical handoff based on a multilayer approach coupled together with adjunct information such as pricing and user preference. This framework allows changes in applications, variations in network conditions and user preferences to trigger vertical handoffs. We conclude by providing a general evaluation of the framework.

Keywords Vertical handoff; Rules-based;

I. INTRODUCTION AND PURPOSE

Wireless Local Area Network (WLAN) and Wideband Code Division Multiple Access (WCDMA) are two promising technologies for data networks. Engineers and entrepreneurs are considering the option of integrating these two networks in order to reap the potential benefits of both individual networks. Though a compelling idea, a key challenge to integrate these two dissimilar networks is to solve mobility management. Mobility management, in a hybrid network scenario, actually implies handoff between two technologies. In addition to the technical challenge posed by such handoffs, known as vertical handoffs (VH), there are business challenges as well, such as cost models, competition, etc. This paper explores the technical aspects of VH in an integrated cellular data network comprising heterogeneous access networks.

There are existing solutions to solve the problem of VH. Most of these existing solutions solve the problem of mobility at a particular layer of the protocol stack. These solutions ignore the potential importance of information from other layers in the protocol stack and the roles they can play for VH. In other words, they lack an integrated approach to solve the problem. This paper discusses the need for a solution which requires the involvement of multiple layers in the protocol stack. It highlights the need for expanding existing solutions in the industry for VH and what needs to be done in order to employ a holistic approach to solve the problem. We propose a novel framework that is *rules based* and allows *multiple layers* and *adjunct information* such as pricing and user preferences to trigger a VH between 3GPP Universal Mobile Telecommunications System (UMTS) network and IEEE 802.11 WLAN for data communications. This framework is rules-based and attempts to address some of the shortcomings of past research. This framework expands the initiation criteria when a VH can be triggered by allowing multiple layers (application, session, network and data link layers) to trigger vertical handoff.

We see value in the applications being able to trigger a handoff and adapt to vertical handoff at the peer/correspondent node (if there is any need). Applications running in mobile devices should be able to adjust the receiving rate to reflect the highly fluctuating network conditions. Furthermore, an application should be able to adjust its transmitting rate in case the peer node has undergone a VH from a high capacity network (say IEEE 802.11 WLAN) to a low capacity network (say 3GPP UMTS).

The layout of the paper is as follows. Section 2 discusses the assumptions and scope of our research. We present a literature review in Section 3. Sections 4-7 discuss the various aspects of the rules based framework. We conclude in Section 8 by providing a general evaluation of the framework.

II. ASSUMPTIONS AND SCOPE

This research assumes that the mobile host will be multi-homed, and the entire data traffic (user payload) will be communicated using one radio technology at a given point of time. We also assume that there is no such change in device during the handoff. The choice of radio technologies for accessing the network is limited to the two options already installed in the device - IEEE 802.11 WLAN and WCDMA. Further, implementation of VH in a network does not exclude horizontal handoff. Hence, if a network decides that a horizontal handoff is required then the system will not proceed for a VH. We, however, will solely focus on VH. This paper relies on the different radio access techniques for each interface and will also not delve into the details of any radio access mechanism of a specific technology.

The scope of our research is limited to data services only. One may argue that voice in UMTS (beyond R5) or WLAN will be transmitted in packet format as well (using VOIP technology) and can be considered as a data service. However, there are additional complexities of voice encoding during a handoff between two systems; hence it is considered as a separate research topic. VH is closely tied to the concept of session management. We will assume that a session is already in place and shall not discuss in detail the concepts of session establishment. We consider handoff of sessions engaged in communicating through a VPN tunnel as a separate topic for research and has left it for further study. This paper further assumes that there exists an IP backbone as the transport mechanism within the UMTS Terrestrial Radio Access Network (UTRAN). It should be noted that the air interface, such as a Bluetooth link, between the mobile phone (Terminal Equipment) and a wireless headset (Mobile Terminal) doesn't fall under the domain of UMTS bearer service, and UMTS quality of service (QoS) will not include the QoS of this air link. For purposes of this study, we will assume that the terminal equipment and the mobile terminal are included in one device – denoted as the *Mobile Equipment* (ME) in this research.

III. LITERATURE REVIEW

The integration of a WLAN and WCDMA network needs to incorporate some core capabilities. As Salkintzis *et al.* explain in [1], interworking mechanism should include integrated authentication, integrated billing, roaming, and a seamless terminal and service mobility. The authors, in the above referenced paper, mainly discuss handoff between a WLAN network and a cellular data network, as a mobile node moves out of one coverage area to another. Though such a situation is absolutely feasible, a handoff between a WLAN and a WCDMA network may be necessary even when the mobile node is stationary. For example, when the effective throughput for the user accessing a WLAN falls below a certain level, the user may force a VH to a 3GPP UMTS network for a guaranteed QoS. Gustafsson *et al.* discuss the scenario of “Always Best Connected” in [2]. As they explain, the definition of “best” depends on personal preferences, size

and capabilities of device, application requirements, security, operator or corporate policies, network resources and coverage. The aim of this paper is to present a framework that would allow a mobile device to remain “best connected”, at all possible times (mobile/stationary). This would require a capability for executing VH when there are changes in the network conditions or the session level QoS. The fundamental principle, however, remains the same - irrespective of when and where the handoff occurs: preserve the IP based session during the handoff. The user should not be required to associate and set up a new connection after the handoff process.

Pahlavan *et al.* [3] have studied the area of handoff in hybrid mobile data networks for non real time applications in wireless networks. The authors talk of different overlay network systems and prioritize the handoffs based on whether the handoff is happening from a General Packet Radio Service (GPRS) Base Station (BS) to a WLAN Access Point (termed as *upward vertical handoff*). According to them a handoff from a GPRS BS to a WLAN Access Point should be done whenever it is possible, whereas the reverse scenario (*lower vertical handoff*) should be used sparingly. We consider that both directions of vertical handoff are equally critical. The proposed rules-based framework uses the QoS requirements of an application along with other criteria, and does not make a decision solely based on the air interface. For a related work on VH on multiple overlay networks, please refer to [4, 5].

Similar research has been done by Zhang *et al.* [6]. Their research also assumes an overlay of the two networks. Their proposed solution with an additional entity – Location Connection Translation – can only be used when the original connection is always present, as in the case of ubiquitous coverage by WWAN (Wireless Wide Area Network). Moreover, they discuss scenarios when a VH is triggered as a user walks into a coverage area of higher WWAN signal strength or WLAN signal strength. As mentioned earlier, there would be other factors that could trigger VH. Nevertheless, their proposal of using Fast Fourier Transform (FFT) based decay detection is interesting and could be used to detect the possible handoff situation as an ME steps away from a WLAN area to a cellular coverage area.

There has been a significant research effort on VH by a collaboration of Ericsson Finland, Helsinki University of Technology, RadioNet and TeliaSonera, known as the VHO project. They have prototyped a service known as VHO Ämpäri, a music service that is VH enabled. This music service can adjust the quality and the required bandwidth based on the available network conditions and user preferences. Once again this goes on to show that VH is not just a phenomenon at the lower layers of the network [7]. A key challenge to the VH is to detect the change in the network parameters once a handoff takes place. For the Ämpäri project, the team developed an additional API layer (known as VHO API) that abstracts the network parameters which are passed on to the higher layers. The team also devised a control protocol for the Ämpäri service. The control protocol

runs at the application layer on top of TCP/IP in a client/server mode. This service uses the most stable connection available for the control channel and the one with the highest throughput for the music service.

Recently there has been a significant focus on research efforts for a joint scheduling mechanism for heterogeneous networks with different radio access technologies. These research efforts use the concept of software defined radio techniques to schedule the traffic and wherever applicable split the traffic into multiple streams over different radio access bearers [8]. Our research, however, assumes that the entire data traffic will be communicated using one radio technology at a given point of time. We, however, discuss the need to use both radio interfaces to engage in signaling (either actively or passively) simultaneously during the active state of the device. Our research, also does not adopt the reconfigurability approach, as discussed in [9]. It assumes that the device can reconfigure itself to choose the access mechanism only between the two options already installed in the device.

During our research we came across another study by Ylitalo *et al.* that uses a policy based approach for dynamic network interface selection in a multihomed mobile host [10]. Though the approach of our study is similar to the concept used in the referenced work, it is important to point out some of the differences in our approach. First of all, the referenced authors use a policy based approach to perform an interface selection. Even though it can take care of pre-call handoffs, we would like to cover the concept of a rules based approach during a mid-call handoff as well. It is important to distinguish between policies for accessing a network versus policies that would guide a handoff. The final outcome of the policy based search in their research would choose one interface over the other for accessing an application, whereas in our research we are concerned with whether there needs to be a handoff or not. Our research assumes that a node has already made a decision on which interface needs to be used initially. Hence the referenced research can be treated as a precursor of this research. Secondly, in our work we will not deal with the concept of “simultaneous access” wherein the transport can be split into more than one interface. Our study will use only one interface for transport at a time, but both for control messages. If there needs to be a handoff the entire transport will be handed over from one network access scheme to another. Finally, [10] attempts to solve the mobility problem at the network layer, using Mobile IP (MIP). Using MIP would mean that there would be no scope of involving the application layer to play a role in the VH. As we will see, applications need to play a lead role in deciding whether a VH needs to take place.

Existing research employs two approaches for implementing vertical handoff – Session Initiation Protocol (SIP) and MIP. Though MIPv6 takes care of the problem of triangle routing, there are additional overheads due to encapsulation. Moreover, MIPv6 doesn't provide a multi layer solution. Even though one can argue that MIPv6 can be an

effective solution for pre-call mobility, it will not be an effective one for mid call mobility (given the handoff latency that would be involved). Moreover, in case of MIPv6, the home agent will be able to tunnel packets to the Care-of address (COA) only after the ME updates the HA of its new COA. During this time, the HA will continue to send the packets to the old address. This will not only create link layer congestion but also cause disruption in the session. Finally, mobility management by MIP follows the MCHO (Mobile Controlled Handoff) approach. For our purposes, as discussed in [11], we need a handoff scheme that follows a MAHO (Mobile Assisted Handoff) model. For related work on using MIP for mobility management, please refer to [12, 13, 14, 15, 16, 17].

Implementing SIP for mobility would not require any installation of home agents. However, both mobile host and the correspondent node should have the SIP User Agent (UA) installed. We will extend the approach by Schulzrinne in [15] to use SIP for supporting mid-call mobility. As suggested in [18], mobility management by SIP is better suited as a mobility solution for next generation heterogeneous networks, even though it increases the overheads of signaling. However it should be noted that we do not add any extra bits for the SIP messages if we use it for supporting mobility, as mentioned in [19]. There is one drawback of using SIP. Since we are relying on protocols at a higher level in the protocol stack, using SIP will also add its share of latency for handoffs. However, if we take the “level of modification, infrastructure change and inherent operational complexity” into account, application layer mobility management fares a lot better than other schemes. For related work on using SIP for mobility management please refer [20, 21, 22]. We have also used our rules based framework, in another work [11], to execute VH with the use of SIP messaging. This would allow an application to adjust its session parameters in the middle of a session in case its peer node undergoes a VH.

IV. COMPONENTS OF RULES BASED FRAMEWORK

The core component of the rules-based framework is a **Rules Engine** (RE). It would be receiving inputs from multiple layers of the protocol stack and other databases. It would process these inputs based on predefined rules. This framework will use a combination of threshold parameter and dwell timer algorithms to monitor the parameters. In addition to these algorithms, the RE can also use these inputs for predictive algorithms to predict a handoff.

This framework can be used by SIP messaging to execute the actual VH (Please refer to [11] for more details). Once the RE gets a go-ahead signal from the network for a VH, it sends a message to the SIP UA of the interface engaged in the data communication to execute the VH. As mentioned earlier, using SIP allows mid call session re-negotiation (through SDP) in case any node has to change its receiving/transmitting capabilities. Figure 1 describes the entities involved in our proposed rules based framework.

Based on the outputs of the monitoring algorithms, RE decides whether a handoff needs to be executed. If so, it issues a request to the link layer to seek approval from the network (access point in case of IEEE 802.11 WLAN and RNC in case of UMTS) for a VH. The network can agree for the VH, if the QoS of the overall system is not affected. Once the ME gets an approval back from the network, the RE issues a request to the SIP UA of the active interface (that is being used for data session) to issue a SIP REFER method to the corresponding node. SIP REFER method is described in IETF RFC 3515 [23]. SIP REFER method switches the media session from one interface to another. It should be noted that the handoff follows the MAHO approach. The RE only assists the network to decide whether a vertical handoff should occur or not. Use of SIP REFER method for executing VH is discussed in [11]. Figure 2 gives an idea of the monitoring algorithms and input parameters for each layer.

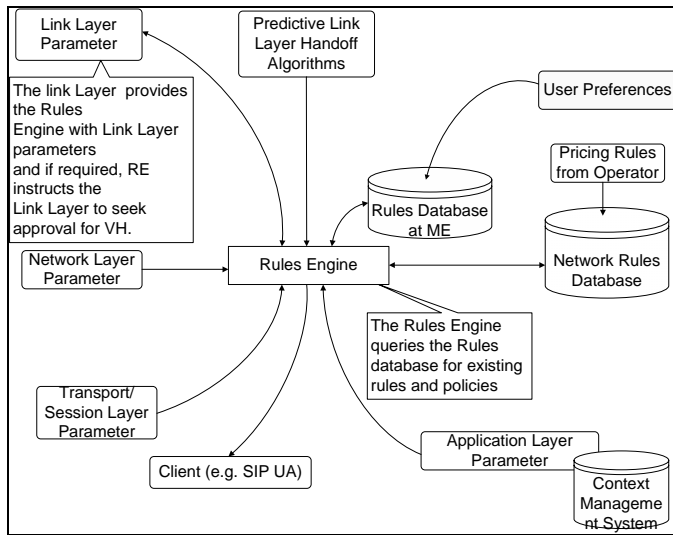


Figure 1: Context Diagram for Rules Based Framework

The RE will be supported by the **Rules Database (RD)**. The RD is a repository of user preferences (e.g. choose the cheapest connection if all else is equal), pricing plans and network QoS policies. The RD need not be centralized. In fact, there would be a RD at the local ME and a remote RD within the core network. In order to save excessive signaling between the local RD and the remote RD, the local RD can have a snapshot of the rules that are relevant for handoff.

The local RD will store the user preferences. Some of the user preferences that can be used for triggering vertical handoff are as follows:

1. Need for security: In case RE finds that a network is not secure, it will always transfer a session to a more secure network.
2. Need for capacity: If a user/application always prefers to be on a high capacity network, RE will initiate a request for VH if it finds a higher bandwidth network.

The remote RD will contain rules based on the overall system, such as throughput of the system. It will also contain rules for admission control for sessions. The remote RD can also play the role of QoS Policy Decision Function (PDF) or the Handoff Control Agent in the network. Concepts of Policy Decision Function (PDF) and Handoff Control Agent in the network, are discussed in [24]. The remote RD can be a part of the SIP Proxy server which stores preferences of a mobile user. Proxy Server, which in a SIP based architecture, normally stores preferences of a user, such as address of device where an incoming call needs to be forwarded. In order to allow vertical handoff to a cheaper network, we need access to rate plans subscribed by a user. We propose to store the rate plans within the core of a network (such as SIP Proxy server) since it may be difficult or unnecessary for a network to transmit a replica of the rate plan to a mobile device. The rules based framework can take advantage of SIP's capability to support the concept of *Virtual Home Environment (VHE)*. This would allow users to use SIP for storing their preferences in either the mobile device (local RD) or in a proxy server (remote RD).

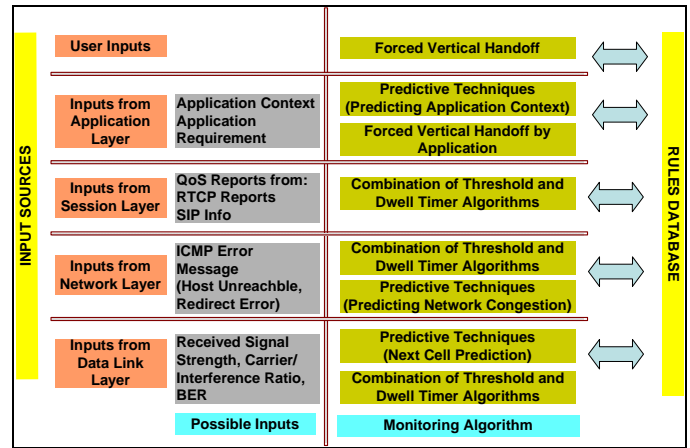


Figure 2: Analyzing the input data by Rules Engine

V. TRIGGERS OF VH IN RULES BASED FRAMEWORK

The authors in [22] suggest that it is not necessary for a device to change its location in order to undergo a VH. We would like to expand the thought and incorporate scenarios such as changes in network conditions, link layer parameters of air interface, quality of service parameters for the session, which can trigger VH. We saw in the previous section that the RE could be getting inputs from the following sources:

1. Data Link Layer
2. Network Layer
3. Transport Layer
4. Application Layer
5. User Inputs (Forced by mobile user)

The data link layer could be monitoring the thresholds for several parameters for the radio environment, such as Received Signal Strength (RSS) threshold level, Carrier to Interference Ratio (CIR), Bit Error Rate (BER). The technique

for measuring the signal parameters such as RSS is different for 802.11 versus WCDMA. While in WCDMA we use a dedicated control channel to measure the signal, we use beacons in the case of WLAN to assess the signal parameters. Beacon messages, in case of IEEE 802.11 WLAN, occupy the same channel as the data in case of WLAN. Hence care should be taken so that too much of the signaling doesn't impact the actual throughput of the data.

In order to avoid the ping-pong effect of handoff, we suggest the use of an additional parameter - dwell time - for the monitoring algorithms in all layers. Each of these parameters can be monitored for their threshold levels. If they exceed (or fall below, as the case may be) the threshold, and remain so for a period exceeding the dwell time, the monitoring algorithm for that layer raises an alarm to the RE. Such alarms would indicate to the RE that a VH could be required. This mechanism can be used in conjunction with algorithms for predicting handoffs based on inputs from data link layer, such as Signal Strength Method [25] and Next Cell Prediction [26].

A change in the network conditions, such as a malfunctioning node can also prompt the network layer to initiate a VH. The network layer receives ICMP messages, e.g. host unreachable, redirect error messages, which can be used to notify RE of possible network "stress". For instance, say an ME is engaged in a session through a WLAN. If the edge router starts experiencing network congestion, it will send ICMP messages to the ME. If the situation continues for a period exceeding the dwell time, the RE can explore whether there is any alternate route through the WCDMA network. If so, it can initiate a request for a VH. Thus based on the inputs from the network layer, an RE can request a VH for load balancing a heterogeneous network system and relieving an overloaded network. A predictive algorithm can also be used. Instead of waiting for the network to actually get congested and then issue a VH, an ME can also predict network congestion based on ICMP messages. If there is no other path available in that network, the RE can instruct the other network access system to discover an alternative route using the other access network. This way, the session doesn't really have to wait until the old route crashes and then look for possible alternatives.

QoS parameters at the session layer, such as Round Trip Time (RTT), can also trigger a VH. During a session a client will receive several RTP Control Protocol (RTCP) Sender Reports and SIP Info messages. These messages will give valuable inputs about the ongoing session. Certain parameters, such as RTT, can indicate whether there has been a deterioration of a session. If the RTT exceeds a certain threshold and remains so for a certain dwell time, the monitoring algorithm for the session layer will raise an alarm to the RE.

In addition to the link, network and session layers, an application can also request a VH. The RE will receive inputs from the application layer if an application wants to force a

handoff. Based on its needs, an application can request to force a VH. An example will make this clear. For instance a user, while being connected through WCDMA network, is surfing a website and wants to download some documents from this site. So he accesses the FTP web page for downloading the documents. When the user clicks "Download" on the web site to download a particular document, the client running the application judges a need for a high capacity network in order to have a prompt download. In that case the application can request the RE to initiate a VH to a high capacity WLAN. An application can also request a "conditional handoff" to the RE. This would mean that as soon as the FTP session is over, the session should be handed back to the WCDMA network.

An application can also predict the need for a vertical handoff, based on the context of the application. These techniques would require the use of the *Context Management System* (CMS) to understand the context of the application that is running and predict a VH based on the future requirements of the application. There is a considerable ongoing research concerning context management systems and how it can be applied to mobile computing (Please refer to [27, 28]). The authors of one of the referenced research paper define context as "collective information pertaining to user's environment aspects, application or computing characteristics or the physical environment of the user/mobile device". If implemented with our rules-based framework, a CMS would allow the RE to interact with one of the proposed *Context Brokers* (CB). The RE would query the CB (which would be residing in the ME) to get an idea of the context of the application that is running, whether it is a streaming news video or a streaming video of an advertisement that is based on the location of the mobile. The CB could also notify the RE on certain changes in the context of the applications that are being run and whether those changes would require a vertical handoff. In the above example, an application need not wait until the user clicks on the "Download" button to request for a VH. Based on the context of the application (that a user is currently accessing an FTP site), the CB can request the RE to initiate a vertical handoff, assuming that the user will download a document if he is currently surfing an FTP site.

Applications can get affected by vertical handoff, even if it doesn't request one. It is possible that the application (such as a streaming news video from a content server) that is currently accessed by a user may not be available once the vertical handoff takes place. In that case, a user could be directed to a similar news service from another content server, under the administrative domain of the new network. In order to have that feature, we need to know the context of the application running in the device. For the purpose of this paper, however, we will assume that the same application can be accessed after a vertical handoff is executed.

In addition to the inputs from multiple layers, a user can also request to force a vertical handoff. A user could be very happy with the current data session that he is accessing

through a public WLAN hot spot. However, let's suppose at one point of time he wants to initiate a money transfer to his bank. Since he is not sure about the security of the public WLAN hot spot, he specifically requests a handoff to his home WCDMA network, which can guarantee him a certain level of security.

VI. AN EXAMPLE OF RULES BASED FRAMEWORK

Until now we have discussed the input sources which can play a role in triggering vertical handoff. Before we proceed on the actual steps followed by the RE, let us first go through a sample structure of the rules that will be stored in the rules database.

In our rules structure, each layer/input source will be allocated a certain priority. Parameters that would be monitored within each layer would be assigned a certain weight. Layers and parameters will be used to formulate a rule. Rules will have a hierarchy as well. In case of a conflict, a rule at a higher level will be enforced. Let us now take an example of a rule and see how RE undergoes the discovery process. Table 1 shows a sample rule structure.

Input Parameter	Source	Max Threshold	Min Threshold	Dwell Timer
RTT	Session Layer	5 ms	NA	3 min
RSS _{current}	Data Link	NA	-85 dBm	4 min
RSS _{new}	Data Link	NA	-100 dBm	4 min

Table 1: A sample rule structure

The suffix of "current" and "new" denotes the current active network and the prospective new network. Suppose we have the following rules:

Rule 1: If ($RSS_{new} > 1.5 * \text{Minimum Threshold}$ for a dwell time = 4 min) and ($RTT_{current} > \text{Maximum Threshold}$ for a dwell time = 3 min), then Initiate VH.

Rule 2: If ($RSS_{current} < \text{Minimum Threshold}$ for a dwell time = 4 min) and ($RTT_{current} > \text{Maximum Threshold}$ for a dwell time = 3 min) and ($RSS_{new} > \text{Minimum Threshold}$ for a dwell time = 4 min), then Initiate VH.

The first rule suggests that when the received signal strength of the new network is greater than one and a half times of the minimum threshold (-100 dBm) for a dwell time of 4 min and the RTT exceeds the maximum threshold (5 ms.) for a dwell time of 3 min, then a request for vertical handoff will be issued, irrespective of the RSS value for the existing network condition. Rule 2 suggests that if the received signal strength of the current network stays below the minimum threshold (-85 dBm) for a dwell time of 4 min, received signal strength of the new network is greater than minimum

threshold (-100 dBm) for a dwell time of 4 minutes, and the RTT exceeds the maximum threshold (5 ms) for a dwell time of 3 min, then a request for vertical handoff should be initiated.

With these rules let us follow an example and see how the RE performs its tasks. We will follow the steps outlined in Figure 3 for three cases. In the first case, suppose RE gets an input from the session layer which indicates that the RTT is 6 ms. Let us suppose that this is the first instance that RE gets an input on RTT. RE feeds this data into the monitoring algorithm for the session layer. When it fetches the rules related to RTT for the ongoing session, RE gathers that an alarm should be raised only if RTT exceeds the maximum threshold and remains so for a dwell time of 3 minutes. Since this is the first instance that RE receives an input (exceeding the threshold) on RTT, it does not raise any alarm but continues to monitor the parameter.

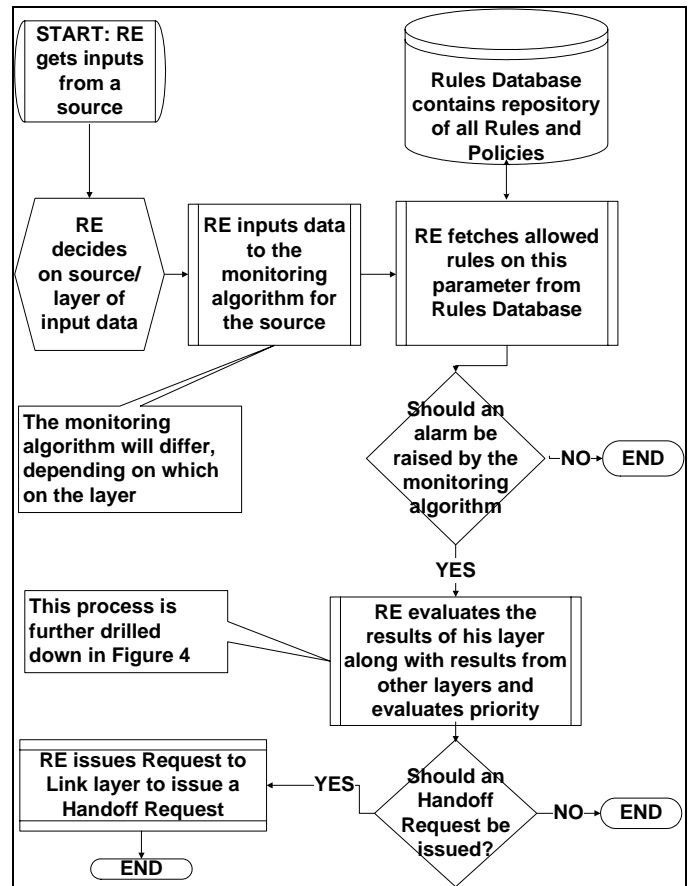


Figure 3: Process Flow for Rules Engine

Let us now suppose that after 3 minutes, RE once again gets an input on RTT. Even this time the value of RTT exceeds its maximum threshold of 5 ms. Since it exceeds the dwell time of 3 min, at this point of time the RE will raise an alarm. It will fetch all rules in the rules database that deal with RTT. Figure 4 shows a drill down of the steps at this stage. RE retrieves Rules 1 and 2, which are mentioned above. It realizes that it needs to find out what are the RSS values for the

existing network connection and the potential new connection. Supposing the most recent value of RSS for the current interface has a value of -75 dBm (above the minimum threshold) and that of the potential new interface is -120 dBm (below the minimum threshold), the RE will ignore the alarm generated by RTT.

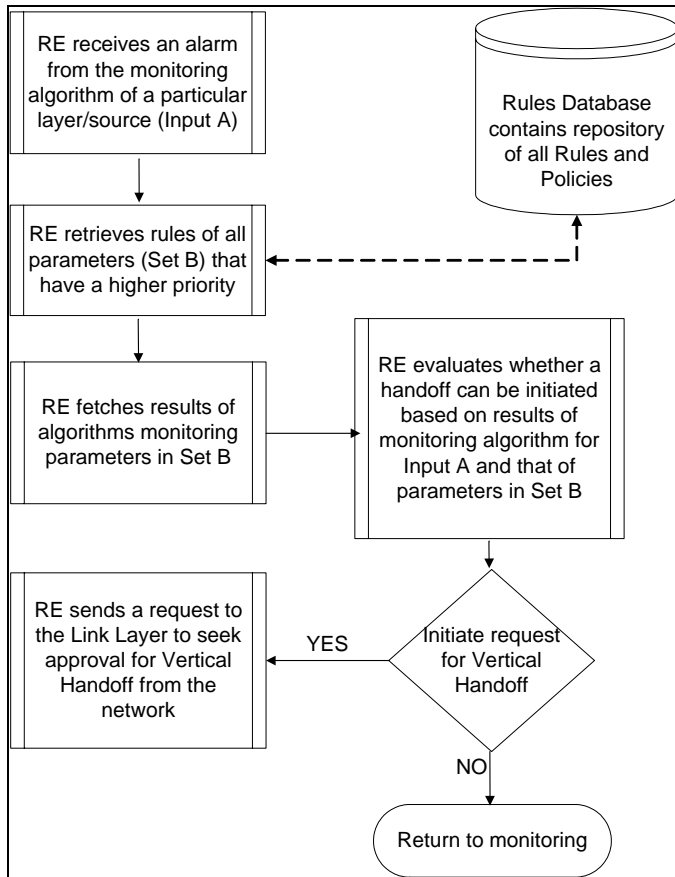


Figure 4: Processing of an input from one layer along with other inputs

Let's discuss a third possible scenario. Once again we receive an input of RTT which has now remained over the maximum threshold for over 6 minutes. Now once again the RE fetches the values for RSS for both interfaces. At this point the RE finds out that the RSS for the new interface has been -80 dBm (higher than the 1.5 times of minimum threshold) for more than 5 minutes. The RE will now issue a request for vertical handoff, irrespective of the RSS value for the current interface (based on Rule 1).

These examples give an idea of the possible rules in the rules database. The threshold value for data link parameters may be set to default values, depending on the network. However, threshold levels for session layer parameters need to be set up for each individual session based on the QoS parameters negotiated during the session establishment. Based on the discussions so far, we can now discuss the three stages of vertical handoff process and the role of our framework.

VII. EXECUTION OF VERTICAL HANDOFF

Conceptually, a handoff process can be broken into three following steps:

1. Discovery and initiation
2. Association with new network
3. Dissociation from old network

In this section, we will briefly discuss how our rules based framework manages each of these stages of a handoff.

A. Discovery and Initiation of Vertical Handoff

Our rules based framework plays a key role during the discovery and initiation stages of a vertical handoff. An ME uses the control channel to discover any new networks and monitor the current network and evaluate whether it is able to maintain the QoS, as defined in the rules database. The rules engine (based on inputs from multiple sources) decides whether a vertical handoff needs to be initiated.

The discovery process can be done in a reactive or proactive fashion, as proposed in [9]. There is a tradeoff with the power drain in the case of proactive scenario. The proactive scenario, when applied in our rules based framework, would mean that the device is constantly on the lookout for new networks. If it finds a better network, the ME would monitor it and decide whether (based on the rules) a handoff would be advised. The proactive mechanism will help in reducing the latency required in vertical handoffs. The proactive mechanism will have the capability of predicting handoffs as well. On the other hand, a reactive scenario would wait for the deterioration of the existing session below certain thresholds (again defined in the rules based engine) and only then would look for other available networks. Table 2 presents a summary of the pros and cons of each method. We propose the use of both proactive and reactive methods in vertical handoff, as required.

Method	Pros	Cons
Proactive Methods for initiating VH	1. Better chances of adapting to network, application and data link changes 2. Lower chances of dropped sessions	Higher battery drain
Reactive Methods for initiating VH	1. Preserves battery power 2. Lower requirements of computational processing	Higher chances of dropped sessions

Table 2: Pros and Cons of Proactive & Reactive Techniques

Using proactive techniques does not exclude the use of reactive techniques after all. Reactive techniques also involve monitoring certain parameters and waiting until they reach certain thresholds. Based on the parameter and the protocol layer of monitoring, the monitoring algorithm in the RE can differ. However, in order to keep it simple for the purposes of this research, we have used a combination of threshold and dwell timer algorithms to monitor the parameters from all sources. In [29] Pollini discusses other algorithms (along with

dwelling timer and threshold level) that can be used to monitor link layer parameters.

There are additional algorithms that can be used to monitor these parameters, such as Hypothesis Testing, Dynamic Programming, Pattern Recognition (based on Fuzzy Logic and Neural Networks) and Reward/Cost based on optimization techniques. In a simplistic case for vertical semi-soft handoff, the rules engine uses a combination of threshold and dwelling timer techniques to monitor parameters (for inputs from all layers). Of course, as mentioned earlier, we would need two different levels of threshold for each network while measuring parameters for the data link layer (since the data link layers are different for each network). While monitoring parameters above the data link layers, we don't need multiple values for the threshold and dwelling timer.

When an ME comes in the coverage of a new network, its RE starts monitoring the parameters of the new network as well as the parameters of the existing network (which has an active session). Once the RE discovers that the new network can be a good candidate for handoff (such as excellent air interface conditions), the RE adds the network in its list of *active set*¹. An active set would contain the list of networks where an existing session can be handed off. Thus an active set will include the networks with which the ME will be exchanging control messages (Note that we limit our model to two dissimilar networks). A network is removed from the active set once it falls below the threshold for a period of time exceeding the dwelling timer.

Once an RE finds that an existing session needs to be handed off to another network in its active set, it sends a message to the link layer to request that the network entity (base station or access point) allows a vertical handoff. The network entity also has to monitor certain parameters (such as system level QoS) before it grants permission. Once the RE receives an approval message from the network, it sends a message to the client application to initiate the vertical handoff (possibly by sending SIP REFER message to the peer/corresponding node).

B. Association with New Network

The association process will differ based on which network the ME needs to associate itself with. However both networks can use a common authentication system, such as a 3GPP AAA (Association, Authentication and Accounting) Server. In order to reduce the handoff latency, we propose the concept of a **Semi-soft** handoff (Please refer to [11] for more details). According to this scheme, an ME will undergo the association and authentication with the new network in its active set

before any handoff request is initiated. This will reduce handoff latency significantly.

The authentication mechanism will depend on whether the target network (potential candidate for handoff) is the home or visiting network for the ME. If both networks have an AAA server, then a peer-to-peer communication between the servers can authenticate the mobile user. Since we are doing an authentication before a handoff is actually initiated, it is possible that the authentication can timeout. If such a situation does arise, then the mobile can get re-authenticated. We discuss more details on association with the new network in [11].

C. Dissociation from Old Network

While the ME is undergoing the stages of discovery and initiation, authentication and association, it continues to transport the data in the old transport channel. In other words, the existing transport channel is not destroyed until the new one is established. Once a new route is established after the initiation process, the ME is ready to give up the old transport channel. The disassociation process differs based on which network the original transport channel resides. As discussed earlier, this process follows the Semi-Soft approach of "Make Before You Break" scheme. Repeating something that has been mentioned earlier - even though the transport channel is destroyed, signaling continues in both networks. We discuss the disassociation process in more detail in another work [11].

VIII. EVALUATION OF MODEL

We have discussed the need for multiple layers to play a role in the initiation process of VH. We have seen that our rules-based framework allows **multiple layers** in the protocol stack to participate in the decision for vertical handoff. Moreover, this model allows **adjunct information**, such as pricing and user preferences, to trigger vertical handoff. This model makes the handoff process (using MAHO model) more **robust**. The rules-based framework, along with SIP messaging, supports both **pre-call and mid-call mobility**.

However, the rules-based framework has certain challenges that need to be addressed. Since we are allowing the higher layers to play a role in deciding on the vertical handoff, it would contribute to the **latency** of a handoff. We had suggested a semi-soft approach to initiate a vertical handoff. The use of a rules engine would mean the use of an intelligent processor that would require a lot of **computational processing**. This research has presented the framework from a functional perspective, and we do not have any quantitative data on the exact processing requirements of the RE. However, we do not believe that computational processing would be a challenge given the fact that mobile devices are slowly developing significant processing capabilities.

In order to fully utilize the benefits of a rules-based framework, we should have **network-aware applications**. That would require re-writing of existing applications for

¹ Active Set is a term that is actually used for soft handoffs. The cells in an active set form a soft handoff connection to the ME. Thus a mobile will be communicating with all Base Stations in the active set. In our case, only signaling messages will be communicated with networks in active set.

supporting VH. On the up side there are already network aware applications in the market [31] and we expect that this trend will continue. Finally, **additional communication protocols** need to be defined between the layers, such that an application is aware of network conditions. There is always a danger of lack of standards in this regard. However, given the strong interest of the industry in other areas of research that involve similar layer violation (such as cognitive software defined radio), we think this challenge can also be dealt with.

The overheads in the rules based framework will be incurred primarily in three forms. First of all, the RE will involve computational complexity to maintain and evaluate the rules and decide whether a VH is necessary. Secondly, it will also involve the processing of the incoming messages to extract the signal parameters. Finally, network design factors will contribute to how much of network bandwidth can be consumed for control traffic. A network should not incur a high overhead of excessive control traffic, that in turn would make the quality of actual data traffic worse.

In a very simplistic case, let us consider a multimedia messaging session between two parties using RTP. We now consider the overhead incurred if we were to incorporate only Layer 4 (Transport Layer) into our model. We would use Sender Reports and Receiver Reports of RTCP for monitoring the RTP session. Collins [32] recommends that RTCP traffic should be set to a maximum of 5% of the bandwidth for an ongoing session. Additional processing will be involved to calculate the RTT and jitter for the RTP packets from the sender and receiver RTCP reports. Finally the RE would use the input values of RTT and jitter and compare it with its rules regarding whether a VH should be executed.

It is part of our ongoing research to quantify the cost associated with the monitoring of control parameters across all levels. From an implementation perspective, we recommend that the model should be incorporated incrementally – first incorporate the model at layer 2 and then implement include the higher layers in a phased manner.

IX. CONCLUSION

The intention of this paper has been to discuss the unique challenges of vertical handoff for data communications and to propose solutions to these problems. We have presented reasons why we need a multiple layer approach for executing vertical handoff. For example, if a node undergoes a VH from a high capacity network to a low capacity one, the corresponding node needs to know that there has been a change in the capabilities of its peer node. Thus, an application running at one end needs to adjust based on the changes at the other end. We also discussed the need to trigger a vertical handoff based on changes in the mobile environment independent of geographic changes of a user. An application, based on the context of the application (e.g. need for higher capacity), should be able to trigger a vertical handoff to a high bandwidth network. This presents two requirements:

1. Multiple layers and adjunct information should be able to trigger vertical handoff.
2. A higher layer (e.g. the application layer) should be able to adjust and re-configure its parameters, if required, due to a vertical handoff at a peer node.

Combining these thoughts, we can say that layers should be able to adjust to vertical handoffs, as well as trigger vertical handoffs. The novel rules-based framework suggested in this research paper can meet the first challenge. Using SIP messaging would allow applications to negotiate the session parameters mid-session after the VH to address the second challenge (Please refer to [11]).

As we rely on higher layers of a protocol stack to trigger handoff, we increase the vulnerabilities of malicious attacks on the system as well. For instance, a rogue user could inject false ICMP (e.g. host unreachable) messages to an ME, which would suggest that a particular route is going down and would hence need a VH. We recognize this threat and believe security features could be put in place to address these concerns. However, we leave this for future research.

This study has been based on a functional approach to solving the challenges of vertical handoff. We have presented qualitative discussions on the need for this framework and how it will meet the challenges. Work is presently underway to study the quantitative aspects of this framework.

REFERENCES

- [1] A. Salkintzis, C. Fors, and R. Pazhyannur, "WLAN - GPRS Integration for Next-Generation Mobile Data Networks." *IEEE Wireless Comm.*, vol. 9, no. 5, Oct. 2002, pp. 112-124.
- [2] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEE Wireless Comm.*, vol. 10, no. 1, Feb. 2003, pp. 49-55.
- [3] K. Pahlavan et al., "Handoff in Hybrid Data Networks," *IEEE Pers. Comm.*, vol. 7, no. 2, Apr. 2000, pp. 34-47.
- [4] E. Brewer et al., "A Network Architecture for Heterogeneous Mobile Computing," *IEEE Pers. Comm.*, vol. 5, no. 5, Oct. 1998, pp. 8-24.
- [5] M. Stemm and R. Katz, "Vertical Handoffs in Wireless Overlay Networks," *ACM Mobile Networking (MONET), Special issue on Mobile Networking in the Internet*, Spring 1998.
- [6] Q. Zhang et al., "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN," *IEEE Com. Mag.*, vol. 41, no. 11, Nov. 2003, pp. 102-108.
- [7] Vertical Handover Project, 2004, Product Modeling and Realization Group, Dept. of Computer Science and Engineering, Helsinki University of Technology. 21 Aug. 2004<
<http://www.cs.hut.fi/~pmrg/pdf/PrototypingVHOServices.pdf>>
- [8] J. Luo et al., "Investigation of Radio Resource Scheduling in WLANS Coupled with 3G Cellular Network." *IEEE Comm. Mag.*, vol. 41, no. 6, June 2003, pp. 108-115.
- [9] P. Demestichas et al., "Evolution in Wireless Systems Management Concepts: from Composite Radio Environments to Reconfigurability," *IEEE Comm. Mag.*, vol. 42, no. 5, May 2004, pp. 90-97.

- [10] J. Ylitalo et al., "Dynamic Network Interface Selection in Multihomed Hosts." Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), Jan 6-9 2003, pp. 315-324.
- [11] A. Sur, "A Rules-based Framework for Vertical Handoff", Master's Thesis (University of Colorado), 2004, Publication No. 1424215, to be available August 2005 in www.lib.umi.com/dissertations/.
- [12] S. Hui and K. Yeung, "Challenges in the Migration to 4G Mobile Systems," *IEEE Comm. Mag.*, vol. 41, no. 12, Dec 2003, pp. 54-59.
- [13] D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPV6", quoted in "IP Multimedia Services: Analysis of Mobile IP and SIP Interactions in 3G Networks," *IEEE Communications Mag.*, Jan 2004, pp. 113-120.
- [14] H. Lach, C. Janneteau, and A. Petrescu, "Network Mobility in Beyond-3G Systems." *IEEE Comm. Mag.*, vol. 41, no. 7, July 2003, pp. 52-57.
- [15] H. Schulzrinne and E. Wedlund, "Application-Layer Mobility Using SIP," Proceedings of Service Portability and Virtual Customer Environments 2000, Dec. 1 2000, pp. 29-36.
- [16] D. Trossen and H. Chaskar, "Seamless Mobile Applications across Heterogenous Internet Access," Proceedings of IEEE International Conference 2003, vol. 2, May 11-15 May 2003, pp. 908-912.
- [17] K. Wong et al., "A Multilayered Mobility Management Scheme for Auto-Configured Wireless IP Networks," *IEEE Wireless Comm.*, vol. 10, no. 5, Oct. 2003, pp. 62-69.
- [18] N. Banerjee et al., "Mobility support in wireless Internet," *IEEE Wireless Commun.*, vol. 10, no. 5, Oct. 2003, pp. 54-61.
- [19] Nakajima, Ashutosh Dutta et al., "Handoff Delay Analysis and Measurement for SIP based Mobility in IPv6," *Proc. IEEE International Conference on 11-15 May 2003*, IEEE (2003):1085-1089 vol. 2.
- [20] G. Camarillo, R. Kantola, and Henning Schulzrinne, "Evaluation of transport protocols for the session initiation protocol," *IEEE Network*, vol. 17, no. 5, Sept./Oct. 2003, pp. 40-46.
- [21] S. Faccin, P. Lalwaney, and B. Patil. "IP Multimedia Services: Analysis of Mobile IP and SIP Interactions in 3G Networks," *IEEE Comm. Mag.*, vol. 42, no. 1, Jan. 2004, pp. 113-120.
- [22] Ma, Jian-Guo, et al., eds. Third Generation Communication Systems. Berlin: Springer, 2004.
- [23] R. Sparks, "Session Initiation Protocol (SIP) REFER Method," RFC 3515, IETF, April 2003.
- [24] W. Zhuang et al., "Policy-Based QoS Management Architecture in an Integrated UMTS and WLAN Environment," *IEEE Comm. Mag.*, vol. 41, no. 11, Nov. 2003, pp. 118-125.
- [25] R. Sankar and N. Savkoor, "A Combined Prediction System for Handoffs in Overlaid Wireless Networks," Proceedings of IEEE International Conference 1999, vol. 2, 6-10 May 1999, pp. 760-764.
- [26] V. Bhargavan, J. Mysore, "Profile Based Next-Cell Prediction in Indoor Wireless LANS", quoted in "A Combined Prediction System for Handoffs in Overlaid Wireless Networks," *Proc. IEEE International Conference on 6-10 May 1999*, IEEE (1999): 760-764 vol. 2.
- [27] M. Chantzara, et al., "Towards Context-Aware Services Provision in Active Network Environments", quoted in "Context Management for the Provision of Adaptive Services to Roaming Users," *IEEE Wireless Communications*, vol. 11, no. 2, Apr. 2004, pp. 40-47.
- [28] S. Xynogalas et al., "Context Management for the Provision of Adaptive Services to Roaming Users," *IEEE Wireless Comm.*, vol. 11, no. 2, Apr. 2004, pp. 40-47.
- [29] G. Pollini, "Trends in Handover Design." *IEEE Comm. Mag.*, vol. 34, no. 3, Mar. 1996, pp. 82-90.
- [30] 3GPP TS 22.228, "Service Requirements for the Internet Protocol (IP) Multimedia Core Network Subsystem, Stage 1, Release 7," Sep. 2004.
- [31] Birdstep Wireless Infrastructure Technologies. Birdstep Web Site. September 2004. <http://www.birdstep.com/wireless_infrastructure/index.php3>.
- [32] D. Collins, "Wireless Communications: Principles and Practice", New Jersey: Goodwin, 1996.