Omnicon: A Mobile IP-based Vertical Handoff System for Wireless LAN and GPRS Links

Srikant Sharma Inho Baek Yuvrajsinh Dodia Tzi-cker Chiueh

chiueh@cs.sunysb.edu

Computer Science Department, Stony Brook University

Stony Brook, NY, 11790-4400

Abstract

Wireless LAN technology based on IEEE 802.11 standard offers mobile users broadband wireless Internet connectivity in public work spaces and corporate/university campuses. Despite aggressive deployment of 802.11b based hotspots in recent years, high-speed wireless Internet access remains restricted to a small number of geographical areas because of limited physical overage of wireless LAN. On the other hand, despite their lower throughput, cellular networks have a significantly wider coverage and are thus much more available. Recognizing that 2.5G or 3G cellular networks can effectively complement wireless LAN, we set out to develop a vertical handoff system that allows mobile users to seamlessly fall back to such cellular networks as GPRS whenever wireless LAN connectivity is not available. The resulting handoff mechanism allows a network connection on a mobile node to operate over multiple wireless access networks in a way that is completely transparent to end user applications. In this paper, we present the design, implementation, and evaluation of a fully operational vertical handoff system, called OmniCon*, which enables mobile nodes to automatically switch between wireless LAN and GPRS based on wireless LAN availability by introducing a simple extension to existing Mobile IP implementation. We discuss the design issues in the proposed vertical handoff system, including connection setup problems due to network address translation, and the disparity in link characteristics between wireless LAN and GPRS. A detailed performance evaluation study of the OmniCon prototype demonstrates its ability to migrate active network connections between these two wireless technologies with low handoff latency and close to zero packet loss.*

Keywords: Vertical Handoff, QoS, GPRS, 802.11b (Wi-Fi), Wireless LAN, Network Address Translation (NAT)

1. Introduction

The proliferation of 802.11b [1] (Wi-Fi) based wireless LAN technology and the benefits arising from the associated tetherless broadband connectivity have given rise to an aggressive deployment of public hotspots that serve as broadband access networks. These hotspots are being deployed in prime locations like crowded restaurants, cafeterias, enterprises, university campuses, airports, etc. The connectivity services of these hotspots are primarily used to access and transfer data between a user's mobile terminal and her home network. The increasing number of such hotspots promise a *bandwidth and connectivity on move* culture that is to come in the near future.

Although 802.11 wireless LAN links boast Mbits/sec bandwidth, their physical coverage is fundamentally limited because of the engineering constraints of the underlying radio technology. To increase the coverage, one can deploy multiple wireless LAN segments in an overlapped fashion. As mobile terminals move across these overlapped segments, they can remain connected continuously by associating with appropriate access points based on the perceived radio signal strength and quality. The intelligence to measure the signal strength and switch among access points is built into the wireless LAN interface cards, which expose various status and control information to device drivers. That is, existing 802.11 wireless LAN hardware supports link-layer handoff without software involvement.

Mobile IP [2], an extension to the TCP/IP protocol suite, takes these wireless access networks one step further by providing mobile nodes the ability to roam across wireless IP subnets while assuming the same network address and maintaining network-layer connectivity. Any network application executing on a mobile host with Mobile IP support can continue to run regardless of any change in the mobile node's point of network attachment. With Mobile IP, mobile nodes do not need to reconfigure their IP addresses while migrating from *home* subnets to *foreign* subnets.

With the growing dependence on wireless devices, continuous network connectivity is a must. For instance, recent shift in PDA manufacturers' decision to equip their products with 802.11b support shows the growing importance of wireless connectivity. According to Gartner [3] group, a research and advisory firm for technology market, the number of wireless LAN users is expected to increase to around 31 million users in 2007 from the current count of 4.2 million in 2003. The number of Wi-Fi hotspots is expected to grow to around 100,000 within the next five years. At the same time, an increasing number of first responder services are starting to use wireless networks to relay information and data to central control locations.

While the roaming capability provided by Mobile IP is good within and across the coverage areas of hotspots corresponding to foreign subnets, it is not adequate for retaining connectivity when one moves out of coverage area for durations long enough to cause already established sessions to timeout. Thus, Mobile IP support is effective for retaining session continuity as long as there is some physical medium available to communicate with the home subnet. If the users are moving across non-adjacent hotspots or if the users pass through "dead zones" in enterprise or campus networks, the

connectivity is lost. The loss of connectivity for a prolonged period may lead to session timeouts, thus rendering the continuous network service of Mobile IP impossible.

General Packet Radio Service (GPRS) is a data service on GSM networks. In contrast to wireless LANs, GPRS provides wider coverage areas and *always on* connectivity. However, the bandwidth provided by GPRS, a maximum of 171.2 kilobits per second, is nowhere comparable to the broadband connectivity of Wi-Fi devices. Worst yet, empirical measurements show that the sustained bandwidth visible to end users is much smaller, especially for the upstream direction. Whereas the Wi-Fi services offered by hotspots usually follow a flat rate charging scheme, whereas, the charging scheme for GPRS services is usually based on the actual usage in terms of the amount of data transfer.

Typically GPRS connectivity is highly available but has lower performance, whereas wireless LAN service is high-performance but is weak in availability. Given the increasing popularity and complementary properties of Wi-Fi and GPRS services, it would be useful if a mobile user can seamlessly switch between these two services. In this paper we present a connectivity switching solution, called *OmniCon*, which aims to provide this switching capability in a transparent fashion to network applications. OmniCon is a simple extension to Mobile IP that takes into account the dissimilarities between the WLAN and GPRS technologies. More specifically, it provides a *vertical handoff* extension to existing Mobile IP implementations without requiring any modifications to them. Mobile IP is essentially a LAN based solution, whereas, GPRS provides WAN connectivity. As a vertical handoff extension to Mobile IP, OmniCon addresses several issues in WAN, such as, network address translation (NAT), differences in connectivity characteristics of WLANs and GPRS, and link QoS. Though we describe out experiences of handoff between WLAN and GPRS networks, OmniCon is independent of GPRS technology and is readily extensible to any other WAN technology with similar issues.

This paper is organized as follows. Section 2 presents an overview of some of the related work. In Section 3, we elucidate the design goals of OmniCon. Section 4 describes the design and architecture of OmniCon. In Section 5 we provide the implementational details of OmniCon. We present the performance evaluation of OmniCon in Section 6. Finally, in Section 7 we summarize the research work in OmniCon and discuss the possible future enhancements.

2. Related Work

The handoff research finds its roots in the cellular GSM networking. Research in handoff optimization is mainly driven by the needs of time-critical QoS enabled applications such as media conferencing and VoIP to have a lossless and low-latency handoff mechanism. Gregory Pollini [4] gives an overview of research on handoff performance and control. He also discusses the trends in handoff research specific to wireless telecommunication networks before the advent of wireless IP networks.

The handoff problem in IP networks can be considered as a special instance of the broader *Mobility Management*

problem, which arises because of the change in a mobile node's point of attachment to the network as it moves around. Mobile IP [2] is a routing based handoff solution for IP networks. In Mobile IP, there are *home* and *foreign* agents running on the wired network. These agents, commonly referred as *Mobile Agents* periodically broadcast Mobile IP advertisements on the wireless LANs. Whenever a mobile node migrates from one subnet to another (foreign) subnet, it starts receiving the Mobile IP advertisements from the corresponding foreign agent. The Mobile IP software running on the mobile node intercepts these advertisements and sends a registration request to the newly discovered foreign agent. After due authentication and consultation with the home agent, an IP-over-IP tunnel is established between the home agent and the foreign agent. From this point onwards, the home agent acts as a proxy for the mobile node, intercepts all the packets intended for the mobile node and transmits them over the tunnel. The foreign agent takes care of decapsulating the packets coming from the tunnel and forwards them to the mobile node. Similarly, all packets that a mobile node transmits are first received by the foreign agent and are tunneled over to its home agent, which further routes them to the true destination. This process is known as *reverse tunneling*. This is the most preferred mode of routing to avoid various issues like ingress and egress filtering at the routers and firewalls. Whenever a mobile node migrates to a new foreign subnet, it needs to bind with the foreign agent of the new foreign subnet, and needs to tear down the association with the foreign agent in the old subnet. When the mobile node returns to the home subnet, standard routing is resumed. The entire process of switching from one mobile agent to another as a mobile node moves across adjacent wireless IP subnets is called *Mobile IP handoff*. OmniCon leverages on Mobile IP to provide a vertical handoff mechanism.

Mark Stemm and Randy Katz [5] are presumably the coiners of the term *vertical handoff*. They point out that no wireless technology simultaneously provides a low latency, high bandwidth, and wide area connectivity to mobile users. They present a notion of ordering among wireless networks where, the high bandwidth – small coverage networks lie at one end and low bandwidth – wide coverage networks lie at the other end. They propose a Mobile IP based solution to transparently migrate across multiple overlay networks. The main focus of the work is to make the handoffs seamless while trying to achieve the best bandwidth. Further, an emphasis is laid on minimizing power consumption and network traffic while avoiding data loss. However, this work does not address the issue of inability of certain networks to communicate with each other in mutual direction. For example, because of network address translation, external hosts cannot initiate communication with hosts from GPRS networks. Also the emphasis is on transition between different networks and not continuous application operation under different networks. In contrast, OmniCon aims to facilitate continuous operation of important applications by enabling QoS on the lines of DiffServ prioritization.

Helen Wang et al. [6] describe a system for vertical handoff which allows users to specify policies to determine the appropriateness of a network and choose the best available network depending on policies. This work can complement well with the vertical handoff system like OmniCon where there are multiple networks to choose from.

Mary Baker et al. [7] suggest a mobility solution as part of the *MosquitoNet* project. The inherent assumption in this

setup is that the mobility management solution that is deployed should not expect any additional infrastructure support from any foreign network. The minimal support that is expected is allocation of an IP address. The setup is based on conventional Mobile IP setup with a difference that the mobile host acts as a foreign agent when the host is in some foreign subnet. The foreign agent module on the mobile node acquires the newly allocated address and the mobile node is connected to the network through the foreign agent module with which it communicates using a loopback virtual interface. This solution is an elegant solution which eliminates the need of Mobile IP support at the foreign network. However, this solution cannot be directly applied to GPRS network for the lack of public IP address assignment in these networks.

M. Buddhikot et al. [8] present an integration solution to 3G and Wi-Fi networks. They describe two integration approaches, namely, tight integration and loose integration, with an emphasis on loose integration. In tight integration approach, the Wi-Fi network is made to appear to the 3G core network as another 3G access network. This is done by deploying an 802.11 gateway at the Wi-Fi network to emulate the functions of a 3G network. This requires the WLAN network and the 3G network to be own by the same service provider. Further, the mobile nodes are required to implement a 3G network stack on top of the Wi-Fi device drivers to present the abstraction of 3G through Wi-Fi. The alternative loosely-coupled approach does not need a direct connectivity of Wi-Fi network with the 3G network but requires Mobile IP based roaming mechanism. In this approach, each Wi-Fi network is connected to an IOTA gateway which acts as a mobile agent and provide connectivity to nodes moving in and out of the network. IOTA gateways also provide QoS guarantees on wireless LANs. For QoS on 3G network, the system relies on enforcement of SLA by the network. This work also presents a comprehensive integration solution in terms of accounting of service usage. This mechanism treats 3G networks as the primary access network and Wi-Fi networks as secondary access networks. Further, this work does not address the issues that arise out of network address translation and the disparity in the characteristics of different networks. On mobile nodes, multiple interfaces are consolidated into a single virtual device and this virtual device is exposed to the OS. The actual handoff is carried out by the virtual device driver. For this purpose, it does not rely on the existing support in Mobile IP.

Rajiv Chakravorty et al. [9] analyze the performance of GPRS links from the perspective of WWW access. They provide a critical insight into several issues encountered by TCP, and especially HTTP, over GPRS links. They propose a design of a GPRSWeb proxy system to boost the WWW performance on GPRS. They use multiple optimization targeted towards GPRS networks, such as, caching web pages, data compression, and delta encoding. Though this work does not directly address the vertical handoff issue, it effectively addresses the operational issues for mobile nodes using GPRS. This mechanism can be easily retrofitted in OmniCon setup to enhance the web performance for mobile users.

A. Bakre et al. [10] propose Indirect-TCP as a solutions for TCP performance degradation on wireless links. I-TCP works by splitting the transport connection at the wired-wireless boundary, usually on base stations. I-TCP maintains two

separate TCP segments, one is over the wired network between the base station and the wired end of TCP connection, the other is over the wireless between the base station and the wireless host. This way, the losses occurring on wireless network are hidden from the wired nodes. To cover up for the losses on wireless segment the base station carries out retransmissions. A variant of this scheme can be used in systems like OmniCon to mask the dynamic characteristics of wireless media like GPRS links. After a handoff to GPRS network, the foreign agent can also work as an I-TCP proxy which monitors and masks the behavior of downstream traffic.

3. OmniCon Goals

Mobile IP provides a mobility solution to mobile users. It is essentially a LAN based solution where a mobile node needs to communicate directly with the mobile agents residing in the current network of attachment. The handoff in Mobile IP is *horizontal handoff*, because the link layer technology remains the same and only the point of network attachment changes. A typical example of horizontal handoff is when a mobile node moves from one wireless LAN cell to another. For Mobile IP to work for wireless WAN technology, such as GPRS, the service providers need to provide mobile agent functionality inside the WAN. Since GPRS network already supports mobility to users at the link layer, GPRS providers do not have much incentive to integrate their mobility system, which is already complex, with Mobile IP to support vertical handoff for other wireless technologies.

Further, existing Mobile IP implementations largely ignore link layer mobility issues, because Mobile IP aims to solve the routing problem during roaming in a way independent of the link layer. In a horizontal handoff between adjacent wireless LAN segments that are associated with different IP subnets, the wireless LAN interface (NIC) on the mobile node (MN) first initiates a link-layer handoff to switch itself to a new access point (AP) after it detects the radio signal from the new AP is stronger, then the MN receives an advertisement from the foreign agent (FA) associated with the new AP, and finally the MN initiates and completes a network-layer handoff using mobile IP. In contrast, in a vertical handoff between GPRS and wireless LAN, it is the software's responsibility, rather than any NIC's responsibility, to decide when the wireless LAN link is usable or unusable, and then to switch from wireless LAN to GPRS or vice versa. As a result, vertical handoff software needs to constantly monitor wireless LAN link quality so that it can use the wireless LAN link as much as possible to minimize the expensive per-byte charge associated with GPRS link usage.

The link characteristics of wireless LANs and GPRS are inherently different. WLANs are capable of providing broadband connectivity but GPRS links are comparatively slow in terms of latency and bandwidth. When a connection is switched from WLAN to GPRS, the downstream traffic needs to be carefully managed to effectively use the meager available bandwidth. Further, the round trip time on WLANs is of the order of a few milliseconds, whereas, on GPRS links the round trip delays can be of the order of thousands of milliseconds. Long round-trip delays have a significant

impact on the TCP connection behavior. These differences between wireless LAN and GPRS need to be taken into account to provide seamless communication service during the hand off period.

In summary, OmniCon needs to implement the following three tasks to support seamless handoff between GPRS and wireless LAN:(1) monitoring wireless LAN link status to decide whether to use wireless LAN or GPRS link, (2) modifying mobile IP implementation, which is designed primarily for horizontal handoff, to oblige the fact that the foreign agent for GPRS link and the mobile node are not on the same subnet, and (3) supporting additional packet scheduling and buffering mechanisms to accommodate transmission characteristics of GPRS link. Moreover, it is desirable to keep modifications to Mobile IP implementation to the minimum so that the resulting code can be easily portable across multiple implementations and versions. Therefore, link monitoring should be carried out by an external software module, which in turn triggers network-layer handoff supported by Mobile IP when a vertical handoff is required.

4. OmniCon Design

4.1. Architectural Choices

For a mobile node to move across IP subnets and still maintain all the active network connections, an indirection mechanism is needed to channel packets between the mobile node and the parties it is communicating with. This packet indirection mechanism also needs to address the issue that a mobile node may want to keep its IP address unchanged regardless of its geographic location. Finally, it is essential that this packet indirection mechanism be notified whenever the mobile node moves into a different subnet, so that it can adjust the redirection parameter accordingly.

Mobile IP is perhaps the most popular and comprehensive packet indirection mechanism. So it is natural to extend Mobile IP to implement vertical handoff. The first extension required is to add a wireless LAN link status monitoring module to the mobile node (MN). In the Mobile IP framework, a Foreign Agent (FA) is needed to channel packets between a mobile node and its Home Agent (HA), and an MN and its FA typically reside on the same subnet. Unlike wireless LAN, it is unlikely that the GPRS link's FA can reside on the same site as MN, because the GPRS service provider does not necessarily support Mobile IP. The fact that an MN and its GPRS link's FA are not on the same subnet means that a tunnel needs to be established between an MN and its GPRS link's FA to channel traffic between an MN and its communicating parties. Note that this is an additional channel than the IP-IP channel between an FA and HA in Mobile IP.

While one end of this tunnel always resides in the MN, there are three possible ways to implement the other end of the tunnel. First, it could be a special *communication agent*, which uses a proprietary protocol to talk to HA and MN. Although this approach gives the most flexibility, it is also least portable as it requires modification to the MN and FA. Second, it could be a generic Mobile IP FA that communicates with MN and HA in a standard way, and takes care of

issues such as setting up forwarding tunnel between HA and MN, authenticating the MN, and updating its own routing table to route packets appropriately. The only modification required here is to add the support for tunneling with MN to the existing FA implementation. Third, the HA can directly tunnel packets to an MN without the help of any intermediate agent. But this approach makes it difficult to introduce additional packet manipulation techniques to accommodate the transmission characteristics of GPRS links, such as intelligent packet dropping, efficient traffic shaping, data transcoding to effectively utilize the available bandwidth on GPRS links, etc. Eventually, we decided to take the second approach, thus restricting the modification to FA implementation. The resulting modified foreign agent is called the *GPRS foreign agent* in OmniCon.

It is also possible to implement vertical handoff in a way that is completely independent of Mobile IP. One possibility is to leverage network address translation (NAT) technology for packet interception and redirection. In this architecture, an enterprise-wide NAT gateway is used to intercept traffic going to mobile nodes belonging to the enterprise. A special tunnel is set up between each mobile node and the NAT gateway so that the communications between an MN and its communicating parties always go through the NAT gateway. Whenever an MN changes its IP address, it needs to inform the NAT gateway so that the old tunnel can be torn down and a new tunnel can be established. Since this special tunnel can run over a wireless LAN, as well as a GPRS link, this architecture provides a unified framework for both vertical and horizontal handoff. The major disadvantage of this architecture is that as an MN moves from one IP subnet to another, it needs to acquire a separate IP address, because there is no foreign agent support in the infrastructure as in the case of Mobile IP. Acquiring a new IP address, through DHCP for example, could be a time-consuming process, and thus could adversely affect the handoff latency. Another drawback is that this architecture requires special authentication protocol between MN and NAT gateway. Because of its relative immaturity, we decided to base the vertical handoff implementation on Mobile IP, and implement it as a simple extension to Mobile IP.

4.2. Network Address Translation in WAN

When a mobile node uses a GPRS link, it needs to obtain an IP address for its GPRS device. Because of lack of public IP addresses, the GPRS service provider typically uses an NAT gateway to translate between public IP addresses and addresses assigned to mobile nodes' GPRS devices. Because of NAT, a mobile node must initiate all network connections it has with the outside world. This restriction precludes the possibility of using conventional IP-over-IP tunneling mechanism to tunnel the traffic over the GPRS link. Consequently, we cannot use the MN decapsulation mode and co-located care of address mechanism, which is the preferred mechanism in Mobile IP when there is no foreign agent in the infrastructure. In MN decapsulation mode, an IP-over-IP tunnel is established directly between a mobile node and the home agent without an intermediate foreign agent. One end point of the tunnel is bound to the IP address of a mobile node, which is typically obtained through a standard protocol such as DHCP [11] or some other static address assignment

mechanism.

For inbound packets to reach a mobile node through an NAT gateway, they need to be part of a connection initiated by the mobile node. To address the NAT problem, OmniCon uses an IP-over-TCP tunneling mechanism instead. The TCP tunnel is initiated by the mobile node so that the NAT gateway on the GPRS network can allow bidirectional traffic in this connection. Use of TCP for tunneling has certain advantages. For example, TCP ensures that the tunneled data reaches the destination in a reliable manner. But at the same time, use of TCP introduces certain overheads such as unnecessary packet retransmissions, additional bandwidth consumption due to TCP ACKs, and underutilization of link bandwidth in TCP slow start phase. Since GPRS link is a point-to-point link, the impact of slow start on the TCP tunnel itself is not very significant as it is the only connection active on the link. However, multiple TCP connections could be tunneled through the TCP tunnel, and these connections may compete with each other for the overall link bandwidth.

Although most Mobile IP implementations support mobile nodes with multiple network interfaces, the GPRS interface cannot be used directly by Mobile IP software. The primary reason is that Mobile IP needs to update the address assignment and routing information associated with an interface. The address assigned to a network interface is typically that of the foreign agent and the default gateway route is updated to the foreign agent as well. The only exception is when the mobile node uses the MN decapsulation mode of tunneling and the tunnel is established using the co-located care of address obtained through DHCP. Since the address assigned to the GPRS interface is not a valid Internet address, it cannot be used as a co-located care of address. Further, any modification to the GPRS device's address could break the mobile node's connectivity with the GPRS network completely. Therefore, the GPRS interface on a mobile node cannot be accessed directly by Mobile IP. Instead, OmniCon introduces a virtual network device, *tcptun*, which is exposed to Mobile IP software. This virtual device implements the TCP tunneling mechanism over an already established TCP connection with the GPRS foreign agent. It also emulates the activities of a network interface for all inbound packets received over the TCP connection, thus creating a semblance of connectivity over a LAN with the GPRS foreign agent. When OmniCon needs to carry out a vertical handoff between the wireless LAN interface and the GPRS interface, it triggers a horizontal handoff between the wireless LAN interface and the virtual device for GPRS interface.

4.3. Handoff Procedure

In Mobile IP, there could be multiple network interfaces in a mobile node, but only one of the active interfaces is used for external communication. An interface is active if advertisements from some mobile agent are received on that interface. In a generic handoff, wireless network interfaces first carry out link-layer handoff, which in turn triggers the network-layer handoff supported by Mobile IP. Because no hardware exists to trigger a link-layer handoff between wireless LAN and GPRS link, OmniCon software needs to assume the task of triggering a Mobile IP handoff, where an MN switches from wireless LAN interface to the virtual interface or vice versa.

To make effective use of wireless LAN bandwidth and GPRS link bandwidth, OmniCon needs to make the handoff decisions intelligently. For this purpose, OmniCon implements a *decision module* which constantly monitors the wireless LAN signal strength, quality, and noise level. Whenever the communication over wireless LAN starts degrading and the signal strength falls below a certain threshold, OmniCon triggers a handoff from wireless LAN interface to the virtual interface. The handoff triggering is carried out by sending multiple foreign agent advertisements on behalf of the GPRS FA, through the virtual interface, up to the TCP/IP stack of the MN. This enables Mobile IP to carry out a network-layer handoff and start using the virtual interface. Once the wireless LAN signal strength becomes available again, Mobile IP switches back to the wireless LAN interface by holding off the foreign agent advertisements on the GPRS link.

It may so happen that the mobile device is in a region where the signal strength is close to the threshold value but fluctuates in a range. This would set off multiple handoffs back and forth. To address this issue, the decision module uses the conventional two level thresholding scheme. Instead of choosing just one threshold, OmniCon chooses a *high watermark* threshold and a *low watermark* threshold. The decision module triggers a handoff from WLAN to GPRS if the signal falls below the low watermark threshold. A reverse handoff of is triggered only when the signal value improves above the high watermark threshold. The distance between high and low watermarks should be more than the typical fluctuation range of radio signals. This ensures that a handoff from WLAN to GPRS is triggered only when the mobile node is moving away from the network, and similarly, the reverse handoff is triggered when the mobile node is moving towards the WLAN.

4.4. Link Disparity

The GPRS foreign agent needs to communicate with the mobile node over a GPRS link, which the mobile node initializes by invoking a dial out procedure. The dial up duration is typically several seconds and can range upto minutes. This duration is too long to provide a continuous network connectivity if a mobile node is migrating through dark spots in an enterprise network. Fortunately, the charging scheme most GPRS providers employ is based on date transfer rather than on connectivity duration. This enables mobile nodes to pre-establish the GPRS connectivity and the TCP tunnel in the system initialization phase. This GPRS link can then be used on an *as needed* basis. Since charging is based on the amount of data transferred, keeping the GPRS link active without using it does not incur any additional monetary cost.

The characteristics of wireless LAN link and GPRS WAN link are very different. The round trip delays on wireless LANs are of the order of milliseconds and those are more than thousand milliseconds for GPRS WANs. Also the available bandwidth is quite different. On WLANs it is in terms of several megabits per seconds, whereas, on GPRS network only a few kilobits per second are achievable. Thus, a vertical handoff may enable mobile nodes to retain their connectivity, but it is impossible to extend the same quality of service to the applications in the event of handoffs. This compels us to implement some traffic monitoring, filtering, and shaping over the GPRS link. Since all traffic is tunneled through

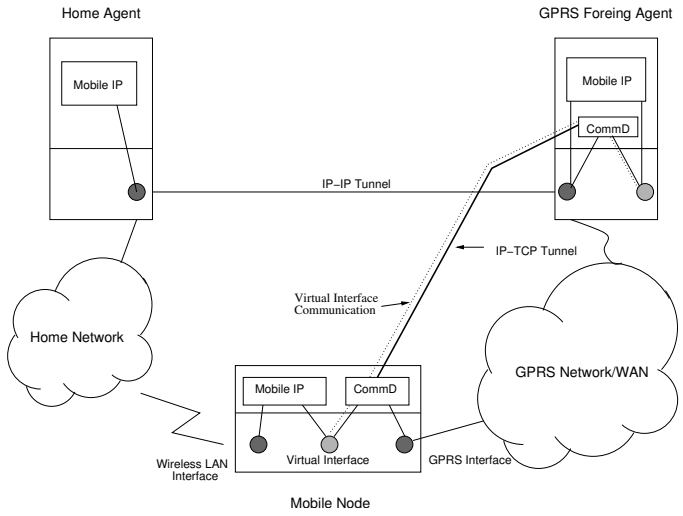

Figure 1. *OmniCon Architecture. The mobile node and GPRS foreign agent are enhanced with OmniCon Communication Daemons. The communication daemons establish a TCP connection with each other over GPRS link. OmniCon exposes a virtual interface to the Mobile IP software. The packets transferred through the virtual interfaces are tunneled over the TCP connection. The Communication Daemon on mobile node interacts with the decision module for triggering vertical handoff. The Communication Daemons also implement the traffic monitoring and filtering logic for effectively using the GPRS link bandwidth.*

the TCP connection between the mobile node and the GPRS foreign agent, the connection end-points are appropriate places where traffic prioritization can be carried out. OmniCon implements a traffic shaping mechanism to improve the utilization of GPRS link and to provide quality of service guarantee to critical applications.

In summary, the primary architectural components of Omnicon are (1) a GPRS foreign agent, (2) a TCP tunneling mechanism based on virtual network device, (3) a decision module to monitor wireless LAN link availability and trigger handoffs, and (4) a traffic shaping module to efficiently use the GPRS bandwidth. Figure 1 shows how these OmniCon components interact with one another.

5. Prototype Implementation

A prototype based on the OmniCon architecture is implemented under the Linux operating system. Because portability is an important goal of OmniCon prototype development, the implementation keeps its dependencies on operation system support to the minimum. The current implementation is readily portable across multiple platforms, such as Linux, Windows and Solaris. In OmniCon design, mobile node needs to be modified to support wireless LAN link availability monitoring, TCP tunneling, traffic shaping and multiple network devices, and foreign agent needs to be modified to support TCP tunneling and traffic shaping. OmniCon also needs Mobile IP software to support an *early-expiration* policy for mobile agent advertisements to facilitate the handoff process.

The primary entities responsible for communication between the mobile node and the GPRS foreign agent are the *Communication Daemons* (CDs) running on both nodes. The communication daemons on the mobile nodes act as clients to the communication daemon on the GPRS foreign agent. The CDs are responsible for establishing and maintaining the

TCP tunnel between a mobile node and the GPRS foreign agent. The CDs are also responsible for providing network Quality of Service to different applications running on mobile nodes. The CD on a mobile node is responsible for interacting with other local components like, the virtual device and the decision module. It is also responsible for triggering the Mobile IP handoff. The implementational details of various components of OmniCon mobile node and the GPRS foreign agent are described below.

Virtual Network Device

To simulate vertical handoff with horizontal handoff, OmniCon implements a virtual network device, called *tcptun*, This virtual device is exposed to the Mobile IP software and the communication daemon. OmniCon uses the generic Virtual Network Device support in Linux kernel [12] to implement TCP tunneling. Similar support is also available in Windows operating system through the NDIS [13] miniport abstraction. The tcptun device exposes an API to be used by the communication daemons to read and write network packets. The packets that are supposedly transmitted over tcptun are handed over to the CD to be tunneled over the TCP connection. The packets received over the TCP tunnel are given to tcptun, which in turn, are given to the operating system TCP/IP stack in a decapsulated form. Thus, from TCP/IP stack point of view, the virtual device is just another network interface. During the system initialization, the tcptun device configuration, such as IP address, subnet mask, etc., is assigned the same value as that of the WLAN NIC. From this point onwards, the Mobile IP software starts listening on this device for mobile agent advertisements. On reception of GPRS foreign agent advertisements on tcptun, Mobile IP software registers with the FA using the same device. Once the registration is successful, the routing table entries are updated to set the tcptun device as the default interface for all outbound packets bearing the mobile node home IP address as the source address. These outbound packets are read by the communication daemon and are tunneled over the TCP connection to the GPRS FA, which takes care of further routing.

During the handoff between WLAN and GPRS, there is a possibility of packet loss because of unavailability of the wireless interface and the delay in registration. OmniCon completely eliminates the packet loss for upstream traffic by buffering the packets and retransmitting them after handoff completion. In order to buffer the packets that are transmitted using the WLAN NIC, OmniCon depends on the NetFilter mechanism available in Linux kernel. Using NetFilters, OmniCon captures all packets that are going to be transmitted over the wireless NIC. These packets are buffered in tcptun internal buffers. After a handoff from WLAN to GPRS is complete, tcptun transparently hands over buffered packets to the communication daemon to tunnel to the GPRS FA. When a handoff from GPRS to WLAN takes place, the tcptun device retransmits the buffered packets that were sent using the tunnel during the handoff. The retransmission is carried out over the wireless NIC. This approach completely eliminates the data loss of upstream traffic. The amount of buffering can be configured by the communication daemon.

Mobile Node Communication Daemon

The TCP connection used by OmniCon needs to be setup by the mobile node as only the hosts residing behind NAT can initiate outgoing connections and no incoming connections are allowed. The mobile node communication daemon (MNCD) is responsible for initiating the dial out procedure on GPRS links and establishing the TCP connection. During system initialization, the MNCD initializes the tcptun virtual device and connects to the GPRS network using GPRS interface. It also updates the routing table entries to set the GPRS interface as the default device for communication with the GPRS Foreign Agent. Once the TCP connection with the GPRS FA is established, it starts receiving all packets that are broadcast by the GPRS FA. These packets include the Mobile IP agent advertisements. The MNCD caches these advertisements without forwarding them to the tcptun virtual device. It also interacts with the decision module regarding the signal strength and quality of wireless LAN link. Based on inputs from the decision module, it initiates a vertical handoff by triggering a horizontal handoff. The horizontal handoff is triggered by releasing the cached FA advertisements to the tcptun virtual device. On reception of these advertisements, Mobile IP software immediately initiates a handoff. From this point onwards, the mobile node starts using the tcptun virtual device for all its external communications. In effect, it starts using the GPRS link through the foreign agent communication daemon. When wireless LAN link later becomes available again, the decision module informs the communication daemon of this change. This results in MNCD filtering out the GPRS FA advertisements, and eventually a handoff from GPRS to WLAN. Subsequently Mobile IP software starts using the WLAN NIC again after due registration with the mobile agent on the wireless LAN. The MNCD also implements a traffic prioritization mechanism to effectively use the upstream bandwidth of the GPRS link.

Foreign Agent Communication Daemon

The Foreign Agent Communication Daemon acts as a server for MNCDs running on mobile nodes. It acts as a router and a Mobile IP Foreign Agent for all GPRS-capable mobile nodes in an enterprise. After startup, like MNCD, it initializes the tcptun virtual device, and starts listening on a well known port for incoming TCP connections from MNCDs. Once a TCP connection is established, the interaction with the virtual device is similar to that of MNCD. The FACD also needs to demultiplex packets that are supposedly transmitted using the tcptun virtual device over multiple TCP connections. For this purpose it maintains a mapping between the home address of the mobile node and the connections. Every outgoing packet is examined to determine the destination address, based on which, the appropriate TCP connection is used to tunnel the packet. All relevant broadcasts, such as, mobile agent advertisements are transmitted over all connections. FACD also implements a traffic prioritization mechanism to use the downstream bandwidth effectively.

Since the GPRS FA is expected to support multiple mobile nodes belonging to an enterprise, its scalability is an important design consideration. Assuming that the GPRS FA is connected to the enterprise network, the downstream traffic to mobile nodes is received by the GPRS FA from the enterprise network, typically home agent, and the upstream

traffic from the mobile nodes is received from the WAN. The upstream traffic from each mobile node is of the order of around 10 Kbps. Compared with current wired network capacities, this is relatively modest. Thus, the network scalability of GPRS FA is solely determined by the amount of downstream traffic it needs to service.

Decision Module

The decision module on mobile node is responsible for triggering Mobile IP handoffs. The handoff decisions are based on the monitored signal strength, quality, and noise levels for the wireless network interface. The Wi-Fi network cards expose these values to the device driver. The signal quality for wireless LANs is determined by computing the difference between the background noise and the signal strength. The decision module periodically polls the device driver statistics to observe a change in these values. The decision module computes an average of these values using a fixed number of previous samples. The averaging scheme is used to eliminate the effects of sudden surges or drops in the observed signal strength values. When the observed average signal level falls below a predetermined *low watermark* threshold, the decision module notifies the MNCD to initiate a vertical handoff. The MNCD responds to this notification by sending multiple cached GPRS FA advertisement over the tcptun virtual device. Mobile IP responds to these advertisements by sending registration requests to the GPRS FA and upon success, starts using the virtual device. A reverse process is carried out when the signal strength improves beyond a predetermined *high watermark* threshold. The decision module employs this two-level thresholding scheme to avoid oscillating handoffs between WLAN and GPRS interfaces.

Traffic Shaping

OmniCon communication daemons implement a traffic prioritization mechanism loosely based on DiffServ expedited forwarding [14, 15] specifications. Because an MN can have multiple TCP/UDP connections going on simultaneously, this traffic prioritization mechanism is designed to regulate these TCP/UDP connections, which share the same TCP tunnel. There are multiple priority queues associated with each TCP tunnel. Each packet sent over a TCP tunnel is examined for its type. If a packet is Mobile IP control message, such as, registration, deregitration, advertisement, etc., the packet is placed in the highest priority queue. Next, the mobile users are allowed to configure the relative priorities among different classes of traffic depending on the quintuple specification as used in Wireless Rether [16] protocol.

$$\{\texttt{SrcAddr, SrcPort, DestAddr, DestPort, Protocol}\}.$$

For each packet, a lookup is performed in the specified user policies and an appropriate priority queue is selected. The packets are transmitted over the TCP connection in a paced manner to match the GPRS bandwidth. The packets from higher priority queue are dispatched before the packets from lower priority queues. When the packets arrive in a queue faster than the dispatching rate the packets are dropped in a tail drop fashion. This priority mechanism ensures that the meager GPRS bandwidth is efficiently used by applications that are of greater interest to the mobile node users.

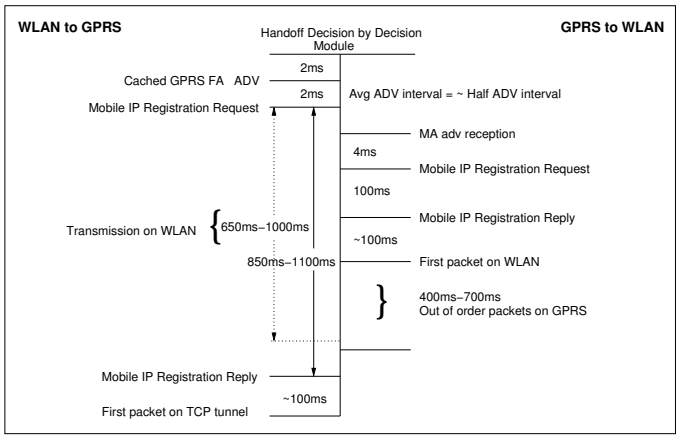

Figure 2. *Gant chart of WLAN to GPRS and GPRS to WLAN handoff.In WLAN to GPRS case, Mobile IP sends a registration request within 4ms of handoff notification. The handoff duration depends on the round trip delays on GPRS link. It is crucial for the decision module to anticipate the handoff well in advance to mask the GPRS round trip delays. The GPRS to WLAN handoff takes less than 250ms. Even after handoff, some packets are received on GPRS link because of latency.*

6. Performance Evaluation

We have implemented a fully operational OmniCon prototype that can seamlessly switch TCP/UDP connections between wireless LAN and GPRS link. To evaluate the prototype's performance, we set up a test-bed to measure the performance cost due to vertical handoff and the tunneling overhead because of additional TCP headers. The behavior of TCP and UDP traffic inside TCP tunnel was analyzed to study the impact of handoff. In addition, the traffic prioritization mechanism was tested for its ability to provide preferential treatment to important traffic.

The test-bed setup comprised of a Home Agent running on a Pentium class machine with 128MB RAM and a 1GHz Pentium-III processor. The GPRS Foreign Agent node was a 600 MHz Pentium-II machine with 128 MB RAM. Both, HA and the GPRS FA were connected to a 100 Mbps switched Ethernet LAN. The mobile node was a 500 MHz Pentium-II notebook with 64 MB RAM. The notebook was equipped with an 802.11b Orinoco PCMCIA adapter. The wireless network was setup in infrastructure mode with a lucent AP-1000 access point. The GPRS service was obtained from AT&T. The GPRS link's peak upstream bandwidth was around 10 Kbps and the downstream bandwidth was around 38 Kbps. The Mobile IP software used was HUT Dynamics Mobile IP version 0.8.1.

Figure 2 shows a Gant chart for Mobile IP handoff for both WLAN to GPRS and GPRS to WLAN scenarios. The starting point in the chart corresponds to the instance when the decision module notifies the CD to trigger the handoff. In WLAN to GPRS handoff, the cached GPRS FA advertisement is released by CD within 2 milliseconds of notification time. Mobile IP responds to this advertisement by invalidating the previous agent advertisement and sending a registration request within 2 milliseconds. The GPRS foreign agent responds with registration reply after approximately 800 to 1100 milliseconds. The length of the duration corresponds to the large round trip times on GPRS link. Figure 3 shows observed round trip time values for GPRS link. The registration reply completes the handoff and the subsequent packets are sent over GPRS network. The handoff duration is dominated by the round trip delay on the GPRS link. In order to carry out a

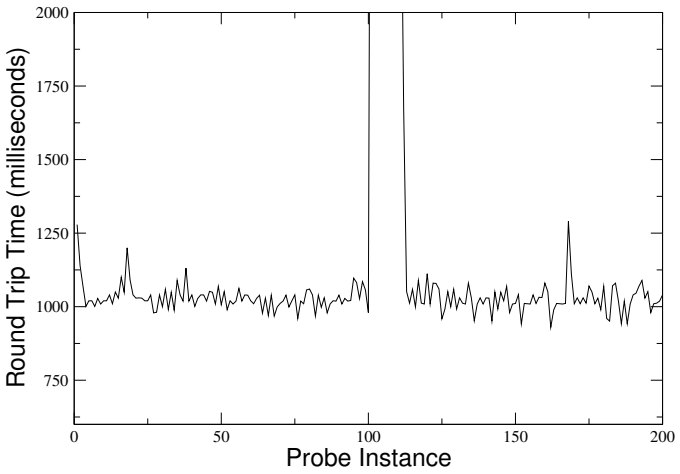

Figure 3. *The round trip times observed on GPRS link. The duration is constantly fluctuating between a range of 850ms to 1200ms with a median of around 1000ms. At times the GPRS link stalls leading to large delays of the order of multiple seconds. The spike in graph is because of stalling of the GPRS link.*

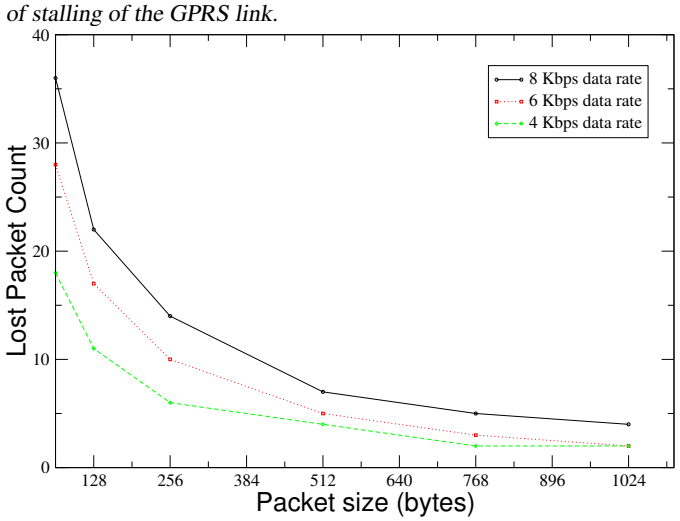

Figure 4. *Packet loss for upstream traffic when there is no packet buffering by mobile node. With no packet buffering there is a loss of upto 36 packets when the traffic comprises of 64 byte packets sent at data rate of 8 Kbps. The number of packets lost depends on the data rate and average packet size. This loss can be completely eliminated by buffering a maximum of 36 packets corresponding to the minimum packet size of 64 and maximum data rate of 8 Kbps.*

seamless handoff, the decision module must anticipate the need for handoff in advance. Thus it is imperative for decision module to also consider the rate of decrease in signal strength to determine the duration of WLAN connectivity in future. This can be done by observing the change in signal levels and extrapolating the values for around 1 second in future.

The GPRS to WLAN handoff is relatively short in terms of duration. The round trip delays on WLAN are short. The overall handoff duration is less than 250 milliseconds after an advertisement is received from the mobile agent. Since this duration is shorter than the latency on the GPRS link, the mobile node keeps receiving some out of order packets on GPRS link for a duration of approximately 400 to 700 milliseconds corresponding to the latency of the link. If the decision module notification is well in advance, there is no packet loss during handoff.

If the connectivity to the wireless LAN is suddenly lost before the decision module could anticipate the handoff there is a loss of data during handoff. This loss can be avoided by buffering previously sent packets to deal with unanticipated

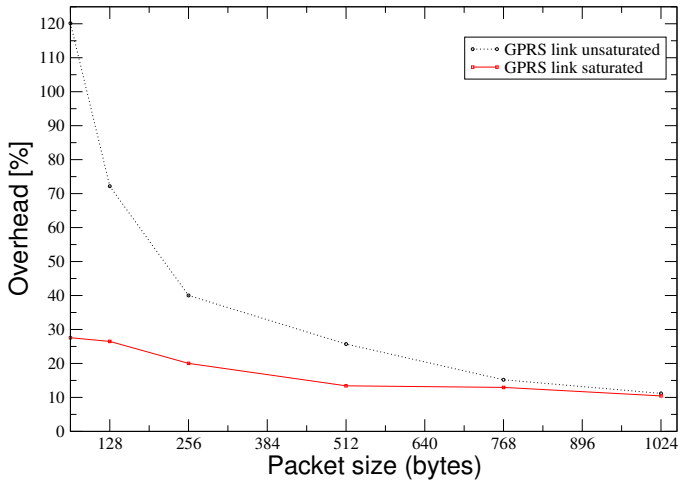

Figure 5. *Overhead incurred because of additional TCP headers and ACKs. When the link is unsaturated, the overhead is large for small packets. The overhead decreases with packet size and link saturation. Link saturation enables TCP to combine multiple packets into single TCP transmission reducing the overhead.*

handoffs. Figure 4 shows the packet loss behavior for upstream traffic during handoff when packet buffering by mobile node is disabled. When the upstream traffic is sent as maximum possible data rate of 8 Kbps with minimum possible packet size of 64 bytes, the number of packets lost is 36. The loss decreases with data rate decrease an increase in packet size. Thus, the tcptun virtual device needs to buffer atmost 36 packets to completely eliminate the data loss for upstream traffic in the event of unanticipated handoffs.

The data sent over GPRS link is tunneled through a TCP connection. We analyzed the overhead OmniCon has to pay because of additional TCP headers and acknowledgments sent on the link. Figure 5 shows the overhead incurred for different packet sizes. The overhead varies with the load on the link. In case of unsaturated link, small packets of size 64 bytes incur as much as 120% overhead. This is because each packet is sent separately and a TCP ACK has to be sent back over the same link. The overhead reduces to 10% for larger packets of size 1024. When the link is saturated, multiple small packets are combined in a single TCP transmission and the overall overhead is reduced. For 64 byte packets, the overhead reduces to a mere 27% and for large packets the overhead is relatively unchanged. The difference between the overhead at saturated and unsaturated link speeds reduces with increase in packet size. This is because of less scope of combining multiple packets in a single transmission.

Figure 6 and Figure 7 show traces of UDP and TCP traffic through the TCP tunnel. After handoff, there is a change in the rate at which packets are transmitted over the pre-established TCP connection. This sudden change invokes the slow-start mechanism for the tunneling TCP connection. The tunneled UDP traffic stabilizes quickly. However, for tunneled TCP connections, their congestion control mechanism is affected by the tunneling TCP connection. This interaction of tunneling and tunneled TCP connections results in a less uniform traffic. The spikes in Figure 7 represent the slow-start phase for tunneled TCP connection. The throughput is reduced in both cased because of difference in link bandwidths on wireless LANs and GPRS networks. The handoff shown is WLAN to GPRS handoff and the throughput shown is

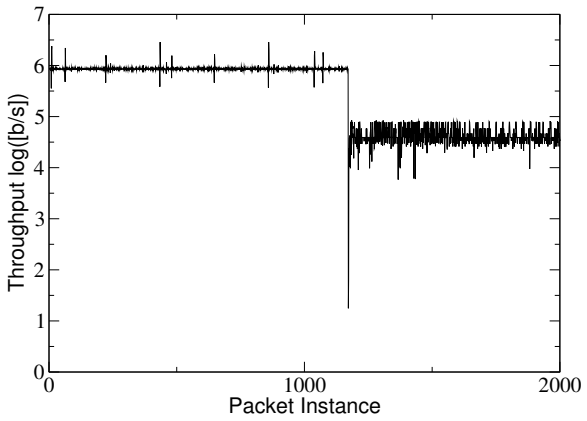

Figure 6. *Trace of UDP traffic during and after handoff. Even UDP traffic has to face TCP slow start because of sudden change in traffic characteristics for the tunneling TCP connection. After this the UDP packets are transmitted in an evenly spaced manner over the TCP connection.*

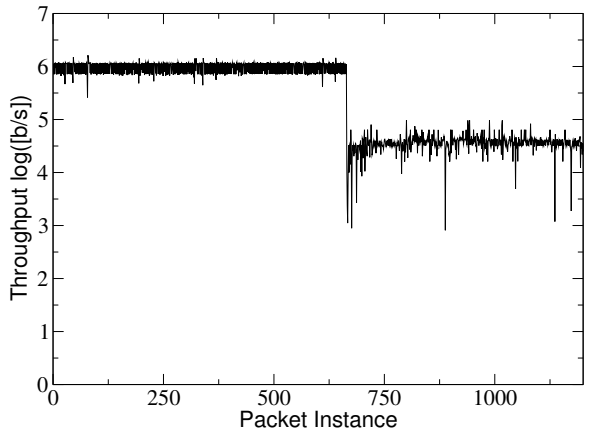

Figure 7. *Trace of TCP traffic during handoff. After handoff, TCP is tunneled inside TCP. The congestion control mechanism of tunneling TCP connection affects the congestion control of tunneled connection. As a result, there is a recurrence of slow-start phase multiple time. The spikes indicate the sudden drop in transmission rate and recovery.*

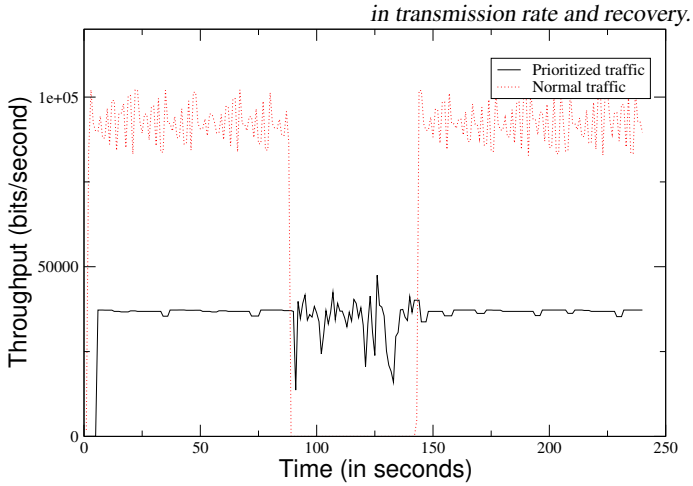

Figure 8. *Traffic prioritization in OmniCon. The prioritized traffic is being pumped at 40Kbps and the other normal traffic is around 1Mbps. After handoff to GPRS, the prioritized traffic gets the available bandwidth, whereas, the normal traffic is completely eliminated. The normal traffic transmission is again resumed after a handoff back to wireless LAN.*

logarithmic in scale.

Figure 8 shows a handoff scenario when a mix of prioritized and non-prioritized traffic are sent to the mobile node. After handoff, the amount of traffic exceeds the available bandwidth. The traffic shaping and prioritization mechanism in communication daemons drops the non prioritized traffic while servicing the prioritized traffic. The non-prioritized traffic is sent at 1 Mbps and the prioritized traffic is sent at 40 Kbps. After handoff, the non-prioritized traffic does not receive any bandwidth whereas the prioritized traffic is still sent at around 38Kbps.

7. Conclusion

Given the enormous popularity of Wi-Fi wireless LAN technology, including 802.11a/b/g, there is little question that that it will become an important component of the future wireless communication infrastructure. However, Wi-Fi

wireless LAN technology has a fundamental weakness of limited coverage, which can be mitigated to a certain extent through the deployment of multiple access points. However, it is unlikely that the resulting coverage can match that of existing cellular networks. Recognizing the complementary role of cellular data service, we set out to develop a vertical handoff system to bridge wireless LAN and GPRS such that mobile nodes can seamlessly fall back to GPRS link whenever wireless LAN connectivity is unavailable. The proposed architecture, called OmniCon, is a simple extension to existing Mobile IP implementations, and the associated prototype shows that it can indeed switch a network connection between these two access network technologies in a way completely transparent to network applications. In this paper, we discussed the design issues due to differences between these two network technologies and their solutions, as well as possible architectural choices and the rationale behind the choices we made.

The two primary issues that deserve repetition are the issues arising out of network address translation in GPRS networks and the disparity in the link characteristics of these two technologies. OmniCon addresses these issues effectively by using TCP tunneling and traffic prioritization mechanism. The performance measurements of an implemented prototype show the advantages of the proposed design. In particular, the *no packet loss* handoff mechanism validates our design choices. The overhead of 10% to 20% of additional required traffic is a necessary trade-off between tunneled connectivity and no connectivity at al. The effectiveness of the traffic prioritization mechanism validates our claims about effective utilization of available meager GPRS bandwidth.

Our ongoing work focuses on providing a generic vertical handoff scheme independent of underlying link layer mechanisms. The GPRS FA design does not depend on any of the GPRS network features. In fact, it is not necessary for the mobile nodes to connect to the GPRS FA using GPRS link only. Thus, on close observation the GPRS FA in OmniCon is in fact a WAN FA, where the incoming connections are from WAN. Thus this vertical handoff mechanism can be easily adapted to other 3G cellular network technologies, such as EDGE or WCDMA. On the performance front, we plan to pursue compression schemes for tunneled data to reduce the link overhead, and explore additional traffic control mechanisms such as I-TCP to use the GPRS link more efficiently.

References

[1] IEEE. “IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications”. Institute of Electrical and Electronics Engineers, November 1999.

[2] “IP Mobility Support”. IETF RFC 2002, October 1996.

[3] Gartner group. http://www.gartner.com.

[4] Gregory Pollini. “Trends in Handover Design”. *IEEE Communications Magazine*, March 1996.

[5] Mark Stem and Randy Katz. "Vertical Handoffs in Wireless Overlay Networks". In *ACM Mobile Networking (MONET), Special Issue on Mobile Networking in the Internet*, 1997.

[6] Helen Wang, Randy Katz, and Jochen Giese. "Policy-Enabled handoffs Across Heterogeneous Wireless Networks". In *WMCSA*, 1999.

[7] M. Baker, X. Zhao, S. Cheshire, and J. Stone. "Supporting Mobility in MosquitoNet". In *USENIX Technical Conference*, January 1996.

[8] M. Buddhikot, G. Chandranmenon, S. Han, Y.W. Lee, S. Miller, and L. Salgarelli. "Integration of 802.11 and Third-Generation Wireless Data Networks". In *Proceedings of the IEEE INFOCOM'03,* April 2003.

[9] Rajiv Chakravorty, Andrew Clark, and Ian Pratt. "GPRSWeb: Optimizing the Web for GPRS Links". In *ACM/USENIX MOBISYS 2003*, May 2003.

[10] A. Bakre and B. R. Badrinath. "A Comparision of Mechanisms for Improving TCP performance over Wireless Links". In *Proceedings of the second USENIX Symposium on Mobile and Location Independent Computing*, April 1995.

[11] "Dynamic Host Configuration Protocol". IETF RFC 2131, March 1997.

[12] Alessandro Rubini. "Virtual Network Interfaces". http://www.linux.it/kerneldocs/vinter/vinter.html, 2003.

[13] Windows DDK, NDIS 5.0 overview. Microsoft Corporation, 2003.

[14] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. "An Architecture for Differentiated Service". IETF RFC 2475, December 1998.

[15] B. Davie, A. Charny, .C.R. Bennett, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis. "An Expedited Forwarding PHB (Per-Hop Behavior)". IETF RFC 3246, March 2002.

[16] Srikant Sharma, Kartik Gopalan, Ningning Zhu, Pradipta De, Gang Peng, and Tzi cker Chiueh. "Implementation Experiences of Bandwidth Guarantee on a Wireless LAN". In *ACM/SPIE Multimedia Computing and Networking (MMCN 2002)*, January 2002.