Proxy Servers Based Seamless Mobility Management

Zhimei Jiang Kin K. Leung Byoung-Jo J. Kim Paul Henry

AT&T Labs Research
Middletown, NJ

Abstract- **We devise an integrated wireless network architecture using proxy servers to support mobility management in this paper. The technique takes advantage of the existing functionalities of proxy servers to provide mobility support for applications such as Web browsing and ftp, without modifying the IP protocol stack. The architecture uses proxy servers to provide application layer functionalities by directing all packets originated from the mobile hosts to a close-by mobility-aware router so that the latter can maintain active data connections during handoffs across different networks. The network architecture is scalable by deploying multiple proxy and mobility-aware router pairs. By assigning mobile hosts to proxy servers dynamically, the proposed architecture provides efficient mobility management functionalities, and is inherently scalable.**

I. Integrated Macrocell and Wireless LAN Architecture

The CDPD (Cellular Data Packet Data) service is our current wireless data services for mobile users. In the near future, wireless service providers will start to provide new, enhanced wireless data services using the 3rd generation (3G) wireless technologies such as the EDGE (Enhanced Data Rates for GSM Evolution) and W-CDMA (Wideband Code Division Multiple Access) networks. However, it is clear that customers will expect services with data rate higher than that to be provided by the 3G networks. To meet the growing demand for better data services, many companies (e.g., www.mobilestar.com and www.wayport.com) have started to provide high-speed data services using wireless local-area networks (WLAN) in places such as airport, convention center, hotel, etc. Such an approach is particularly feasible and attractive due to the maturity of WLAN technologies such as the IEEE 802.11b, which can provide a data rate of 11 Mbps, for example, far exceeding the maximum data rate of about 480 Kbps to be offered by the EDGE system.

We note that one major shortcoming of the WLAN services is its limited *service coverage*. That is, due to the lack of national wireless infrastructure, the WLAN service providers can offer services only in limited areas such as airport and convention center. As a result, users traveling outside of the WLAN areas cannot obtain any services at all. On the other hand, many cellular network operators have a nation-wide footprint, although the data rate will not be comparable with that of the WLAN. For the reason, an integrated macrocell and WLAN network for enhanced, seamless wireless data services that take advantage of the wide coverage of the cellular networks and offer high data rate through WLAN whenever it is possible provide an ideal combination of solutions.

A schematic diagram of the integrated (or overlaid) macrocell/WLAN architecture is presented in Figure 1. The macrocell network (e.g., the CDPD, EDGE or W-CDMA network) continues to be used to provide services for vehicles and pedestrians with moderate data-rate services in wide areas and roads. In addition, the architecture overlays WLAN's at places such as airport, convention center, and stadium, on top of the macrocell network. The idea is that users located inside the service area of WLAN are expected to be less mobile and equipped with sufficient computation and radio capability (e.g., laptop computer with WLAN air-interface card) for high data-rate services. Thus, these users are served by the WLAN instead of the macrocell network for enhanced services when possible.

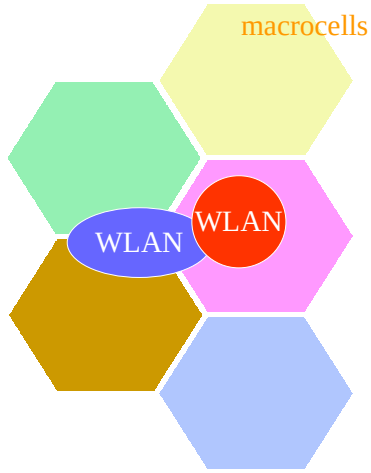

Figure 1: Integrated macrocell and WLAN architecture

The rest of this paper is as follows. In the next section, we further elaborate the advantages of using such integrated macrocell/WALN architecture. We also relate our work to previous research in the same general area. Section 3

describes in detail the architecture design, components of the system, as well as it basic operations. In section 4, we extend the architecture for wide area cases where multiple proxies are desired. We conclude in Section 5.

II. Advantages and Motivation for Integrated Macrocell/WLAN Architecture

It is well known that mobile users expect high-speed wireless data services to meet their needs for mobile communications and computing. In particular, there are an important market segment of wireless data services for business users (e.g., traveling warriors with laptop computers) who demand for mobile computing services at public places such as airport and convention center. Due to radio propagation difficulty, such indoor areas may not be served well by future macrocell networks. In contrast, the indoor environment can be served much cost effectively in terms of coverage and data rate by WLAN. In addition, a single macrocell may not be adequate to handle traffic for a potentially large number of users in a concentrated area such as an airport. WLAN can be deployed easily to provide enough capacity to meet customers' traffic demand in the indoor areas.

One important advantage of the integrated macrocell and WLAN network is that the WLAN technologies have become mature and the 3G networks will be available very soon. Another crucial advantage is that although the WLAN service providers can offer services at selected areas, only carriers with a national footprint can provide seamless wireless data services to users using the integrated network anywhere throughout the nation. Customers will be served well with much improved data rate by the WLAN at locations where they are less mobile, have time to perform serious mobile computing (e.g., read emails or log onto computers at office) and expect premium services. While customers are located outside of WLAN areas, they can still access services through the macrocell network with moderate data rate.

The idea of supporting users crossed different types of networks has been tossed around for several years [2]. A number of experimental systems have been tested including research projects at Stanford University and CMU [4][5]. The momentum of WLAN and the expected popularity of 3G wide area cellular data systems significantly increase the chance of deploying such architecture in a wide scale. Clearly, at the core of such systems is the mobility management scheme, which can maintain user connections after a vertical handoff, a handoff between different types of networks. Mobile IP is the most widely studied approach for handling mobility, where packets from and to the mobile host are tunneled through a home agent at its home network so that the server that is corresponding with the mobile hose can be shielded from the mobility of the mobile host [1]. To resolve certain scalability problem associated with mobile IP, several new schemes have been proposed, most noticeably the Cellular IP from Columbia University and the Hawaii project from Bell Labs [3][6]. All of these solutions, including Mobile IP, require significant changes on the mobile host and system architecture. In this paper, we propose a scheme that minimizes these changes, and yet is able to provide mobility management to a wide range of applications.

III. Integrated Network Architecture Using Proxy Servers

In the integrated network environment, mobile terminals are allowed to move between the macrocell network and the WLAN. Since both types of networks use their own IP addresses for routing, there is a need for a scheme that can maintain the data flow (connection) between the server and the mobile terminals, regardless of the actual serving network. Traditional mobility management schemes, such as Mobile IP, reside at the IP layer of the protocol stack, thus requiring significant changes on the mobile side, which has hindered the wide deployment of such mechanisms. In this paper, we focus on web-based applications in the integrated macrocell/WLAN networks. That is, users make use of the http (Hyper-text transfer protocol) to access required data. Examples of such applications include access to web page and web-based email. An architecture of the integrated network is presented in Figure 2, where mobile terminals access data from a remote server via the wireless access network in use (i.e., macrocell network or WLAN) and possibly through the public Internet.

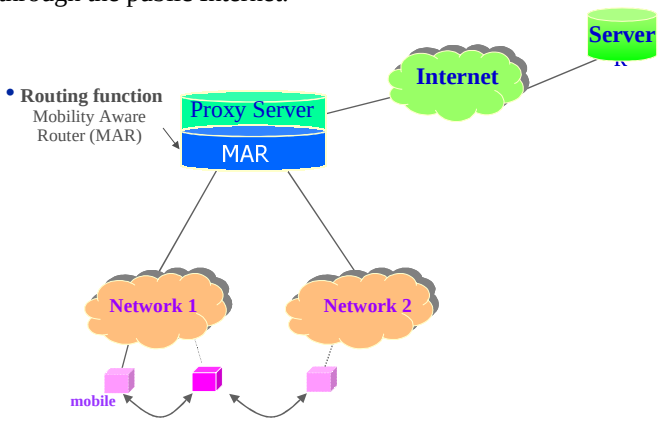

Figure 2: Proxy based mobility management architecture.

An approach to supporting continued data connection for mobile users is by use of proxy server, as shown in the figure above. The proxy is connected directly to a mobility-aware router (MAR), which is a router with mobile management and related functionalities and the MAR is then connected to various wireless access networks. One can view the proxy and the mobility-aware router as single, combined entity, although the two can be

two physically separate pieces of equipment. The essence here is that all the traffic going to and coming from proxy server must pass through the MAR. For instance, the MAR can be the gateway of the subnet where the proxy server is located. The proxy handles functions traditionally reside in a proxy server, such as transcoding, caching, etc., and it can be completely unaware of the mobility of the client; whereas the MAR handles routing functions and mobility management, and provides a "static" location of the mobile client to the proxy application layer.

A. System components

The above figure indicates that the system has four key components: mobile host (as referred to as client), proxy sever, MAR, and destination data server. Their roles in the system are as follows:

Mobile host:
Each mobile host is assigned a unique ID across the system. For instance, this ID can be a routable permanent IP address of the mobile at its home network. The same terminal can be assigned with different (temporary) IP addresses for routing purposes while being served by the macrocell network and WLAN. Such IP addresses are referred to as routing IP addresses in the following.

The client software is set up to use proxy for its WEB, ftp, and other network accesses. The proxy feature has been incorporated in almost all the WEB browsers currently on the market, thus making the proposed scheme accessible to most of the users. The client may need to go through certain registration and authentication procedure prior to gaining access to the proxy. A demon is also run at the background such that whenever the routing IP address for the mobile host changes, a notification is delivered to update the record at the MAR. Details about the registration and location update process are discussed later in the paper.

In order to maintain an active connection during a handoff, where the routing IP address of the mobile host is changed, it is assumed that the client software identifies a TCP connection only through its port sequence and port number. In other words, the host machine does not relate TCP connections with the IP address in any fashion. This way, after an address change, the mobile host can continue to send and receive through the same TCP connection created before the IP address change, albeit using the new routing IP address. The TCP connection related information, such as sequence and port number, remains the same during the handoff.

Proxy Server:
The proxy can be a generic proxy, which handles application layers functions. For example, upon receiving a request from a mobile host, it checks whether the requested content is already cached locally. If so, the cached information is sent to the host without issuing a real request to the destination server; otherwise, the proxy fetches information from the destination server and then forwards it to the mobile host.

The significance of the proxy in the architecture is that, it provides a means to direct traffic from and to a client to a single point, where mobility management functions can be carried out. In our design, this is accomplished by the MAR.

Mobility Aware Router:
The Mobility Aware Router (MAR) is the core of the system. Together with the mobile host, it provides the mobility support such that packets can be delivered to the mobile even when it moves from one network to another.

Depending on whether the mobile is aware of the existence of the MAR, the mobile may send the control messages, such as registration and location update, directly to the MAR; or to the proxy server, in which case, the MAR will intercept these packets and act accordingly. Note that since MAR can see all the packets exchanged between the proxy and the outside machine, it is fairly easy for it to scan for particular port numbers set aside for the control messages and intercepts those packets.

To keep track of the mobile hosts, the MAR maintains a table that maps the permanent unique ID of the host to its current IP addresses. MAR also needs to include in the table IP addresses used by the client in the past, in order to maintain the active connections after a handoff. The past IP addresses in the table may be removed upon closing of all the connections using the IP address or after all the associated connections have stayed idle for a certain period of time. If IP addresses are reused very often, the TCP port number may also be used to identify connections.

Destination Data Server:
No modification is required at the destination data server. In fact, the data server is not aware of whether it is accessed by a proxy or a client directly, neither is it aware of the client mobility.

B. System operations

In this section, we illustrate the key operations involved in this architecture, namely registration, packet delivery, mobility update, and contact initiation.

Registration and authentication:
To begin a service session from a wireless access network (e.g., either the macrocell network or the WLAN), a mobile host initiates the authentication and authorization process with the access network. At the successful conclusion of the authentication process, a (temporary) routing IP address is assigned to the mobile host by the network in use.

Permanent ID	Current routing IP address	Past routing IP addresses
P_ID_1	IP_{m1}	$T\_IP_{m11}, T\_IP_{m12, ..}$
P_ID_2	IP_{m2}	$T\_IP_{m21}, T\_IP_{m22,...}$
…	…	…

Figure 3: Mapping of terminal permanent into routing IP addresses.

The information regarding which proxy to use may be programmed on the mobile host or may be dynamically obtained from the system during the registration process. In the first case, a script may be written to indicate the proxy server to be used based on the interface that is accessed, e.g., Ethernet, WLAN, and GPRS, and/or, based on the network domains, e.g., WLAN of MobileStar vs. that of WayPort. The second option allows the network to inform the client of the proxy to be used as part of the registration process. The advantage of the second option is that the system can assign the mobile the proxy dynamically based on its needs. For instance, the one that is closest to the mobile's current location may be assigned to improve the efficiency of the routing. This aspect will be further elaborated later in this paper.

After gaining access to the access network, the terminal may need to register with the proxy server, and pass the authorization and the authentication process, prior to receiving services from the proxy server. In this process, one piece of the information for identifying the mobile host is its unique ID. The terminal also forwards the current routing IP address, together with its unique ID, to the MAR, so that the MAR can establish an entry for the mobile in the table shown in Figure 3. MAR can then use this table to handle packets sent to and from the mobile host.

During the registration and authentication process, the proxy server may further obtain user information from a centralized server, where user account information is maintained. In the mean time, the proxy server may inform the centralized server that it is the one serving this particular mobile, in case any other machine needs to contact the mobile hosts.

Mobility update:
The mobility update procedure takes place when an active terminal moves to a different network that requires a new routing IP address, e.g. from the macrocell network into a WLAN or vice versa. It is understood that the terminal has the capability to detect the existence of the new network that it has just entered into before initiating the authentication process. Upon receiving the new routing IP address, the mobile sends to the MAR a message that includes mobile's unique ID, the new IP address, and possibly its previous routing IP address. The MAR then updates its table accordingly, i.e., fill in the "current IP address" field with the new address, move the old IP address to the "past IP addresses" list if there are active connections tied to this IP address, or otherwise, remove it from the record completely.

Packet delivery:
The packet delivery process is indicated the figure below. The IP addresses use in the figure represent the following: IP_m: current address of the mobile; IP_p : address of the proxy server; IP_s : destination server address. Let's now discuss this process step by step.

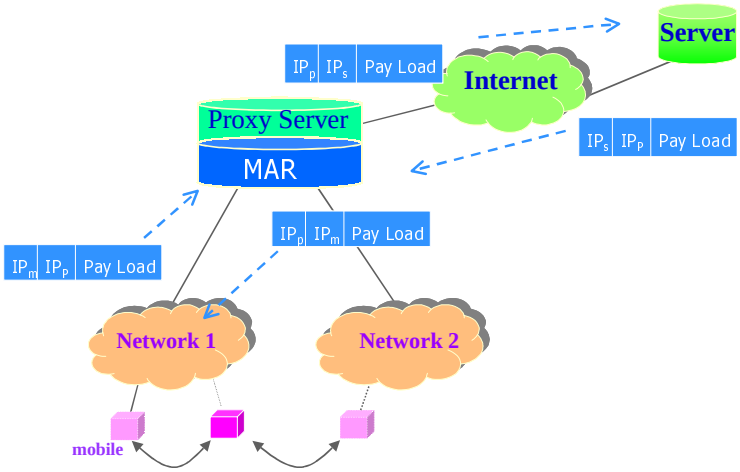

Figure 3: Packet delivery procedure

1. Mobile host sends a packet to the proxy server indicating the content that it would like to access. For example, the payload of the packet can be: "get http://www.cnn.com". The source and destination addresses of the packet are IP_m and IP_p respectively, as shown in the figure above.
2. On its way to the proxy server, the message first reaches the MAR. The MAR updates its table to indicate that there is a new connection from this mobile at the IP address IP_m.
3. The packet is then forwarded to proxy, which either returns a cached copy of the requested information or forwards the request to the actual destination server. In the latter case, the new packet has source and destination IP address IP_p and IP_s respectively; and the reply from the destination server is naturally destined for the proxy server. In both cases, at some point, the

proxy will have the data ready to send back to the mobile host.

4. The proxy sends the data back to the mobile host using the same IP address from which the initial request was received. The MAR, upon detecting a packet destined for the mobile host looks up its table to find out the current routing IP address of the mobile host, by matching the destinations IP addresses with the ones in the table. If the mobile has obtained a new IP address, the destination IP address of the packet is replaced with the current routing IP address of the mobile host, while keeping the TCP header information unchanged. This way, packets destined to a given mobile host can always reach the mobile's current location, regardless of which access network the mobile is in. Alternatively, the proxy may be partially mobility aware, i.e., it can look up the table itself and use the current IP address for reply, which would accomplish the same result.
5. The client removes the IP header information from the packets received and forwards them to the upper layer such as TCP. As mentioned previously, the client has been setup such that it can continue to send and receive packets through the same TCP connection after the IP address change.
6. If there are new packets need to be delivered, the mobile will always use its current IP address, regardless of the IP address that was used to create the TCP connection.

Contact initiation:
It is often desirable to be able to contact the mobile host, without having the mobile first initiate the contact. For example, the mail server may want to push an urgent email message to the mobile host. To accomplish this task, the contacting machine needs to contact the centralized server that coordinates among proxies serving mobile hosts, to obtain the information as to which proxy is currently serving the targeting mobile host based on its unique ID, and then contacts that proxy server directly. The proxy server, together with the MAR, upon receiving such packets will go through the same table lookup procedure as to the regular data packets, to obtain the mobile's current IP address and delivers the request to the host.

In summary, the key function of the proxy server in the integrated architecture (Figure 2) is to force all packets originated from a given mobile terminal for the web-based applications to go through the proxy, thus passing through and being processed by the MAR. Packets destined for the terminal are routed properly from the MAR to the destination server. The use of the proxy server in this architecture takes advantage of the existing functions of Web proxy to enable mobility management, with no modification requirement at the IP protocol stack. This will certainly enable easy deployment of this mechanism. Furthermore, the wireless access networks in the architecture under consideration in Figure 2 are not necessarily macrcocell networks and WLAN, respectively. In fact the techniques described above also apply to terminal movement between one macrocell network and another, as well as between one WLAN and another, as long as the access networks use different IP addresses for routing.

IV. An Integrated Architecture Using Multiple Proxy Servers

Another advantage of the architecture presented in this paper is that it is inherently scalable. An integrated macrocell and WLAN architecture using multiple proxy servers is shown in Figure 5, where each of the access networks can be a macrocell network, a WLAN and a combination of types of wireless networks, as long as each access network uses one IP address for routing. The proxy/MAR pair can be placed at different geographic locations. The solution techniques described above apply to the architecture in this figure as well.

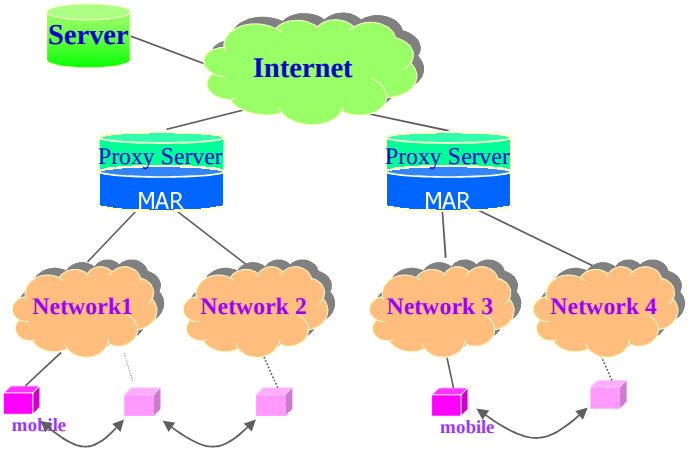

Figure 5. An integrated architecture using multiple proxy servers.

To provide scalability to the system, given the existence of multiple proxy/MAR pair, there is a need for an assignment algorithm to associate a mobile host with a particular proxy/MAR pair. One such assignment is to allow an MAR to serve as a gateway route for a given, distinct set of permanent IP addresses for a number of mobile hosts. This way, regardless of the network attachment point for a terminal, a fixed, unique proxy/MAR pair provides the needed packet routing and processing capabilities. The mobile host is informed of the proxy assignment during the registration process, as it tries to gain access to the access network and the proxy service. If during the period when the mobile remains registered, it roams a long distance and it becomes inefficient to keep the mobile on the same proxy server, the mobile can be

"handed off" to a new proxy server. From this point on, all the new connections will go through the new proxy, while the old one continue to use the old proxy.

Just as it is with traditional WEB proxies, multiple proxy/MAR pairs may form a hierarchical structure. In other words, requests can be sent from one proxy to another proxy in the case the originating proxy cannot provide the service required by the mobile host.

V. Summary

In summary, this paper presents a proxy based mobility management scheme. The system takes advantage of the existing proxy supports in network applications to maintain network connection during a handoff. More specifically, a mobility aware router that is coupled with the proxy is able to detect the network change of mobile host and forward the packets based on the mobile's current IP address. The main advantage of this scheme is that it requires little changes at the client, hence will be easier to implement. Because it depends on the availability of proxy support, applications without such supports will not be able to take advantage of this architecture. However, as more and more applications, from email to multimedia applications, are becoming Web based and are accessible through Web browser, this architecture is able to provide mobility support for a wide range of applications.

References:

[1] P. Bhagwat, C. Perkins, and S. Tripathi, "Network layer mobility: An architecture and survey,", *IEEE PERSONAL COMMUNICATIONS*, PP. 54-64 JUN 1996.

[2] E. Brewer, R. Katz, Y. Chawathe, S. Gribble, T. Hodes, G. Nguyen, M. Stemm, T. Henderson, E. Amir, H. Balakrishnan, A. Fox, V. Padmanabhan, S. Seshan, "A network architecture for heterogeneous mobile computing," *IEEE PERSONAL COMMUNICATIONS*, PP. 8-24 OCT 1998.

[3] Campbell, S. Kim, A. Valko, C. Wan, Z. Turanyi, "Design, implementation. and evaluation of cellular IP," *IEEE PERSONAL COMMUNICATIONS* , pp. 42-49 AUG 2000.

[4] A. Hills and D. Johnson. A Wireless Data Network Infrastructure at Carnegie Mellon University. IEEE Personal Communications, 3(1): pp. 56-63, February 1996.

[5] Kevin Lai, Mema Roussopoulos, Diane Tang, Xinhua Zhao, Mary Baker, "Experiences with a Mobile Testbed**,**" *Proceedings of the Second International Conference on Worldwide Computing and its Applications (WWCA'98)*, March 1998.

[6] R. Ramjee, T. La Porta, L. Salgarelli, S. Thuel, K. Varadhan, L. Li, "IP-based access network infrastructure for next-generation wireless data networks," *IEEE PERSONAL COMMUNICATIONS*, PP. 34-41 AUG 2000.