A Mobility Enhancement for Switched Wireless Ethernet with Soft Handoff

*Raheem A. Beyah, *Corey S. Turner, Cherita L. Corbett, and John A. Copeland*

* Cisco Systems, Inc.
Herndon, VA 20171

Communications Systems Center
Georgia Institute Of Technology
Atlanta, Georgia 30332-0490

ABSTRACT

With the proliferation of mobile users, wireless LANs (WLANs) are quickly entering enterprise networks. These networks have traditionally been configured in a Shared Ethernet topology. Due to the tremendous performance increase over Shared Ethernet, most corporations' networks are slowly being migrated to a Switched Ethernet topology. Traditional Ethernet switches will not suffice to completely take advantage of this WLAN technology. When WLAN's use Ethernet switching as a backbone, service disruption may occur when attempting to migrate from one cell to another. This disruption does not occur on Shared Ethernet networks. The authors propose a soft handoff procedure to ensure optimal performance, minimal service disruption, as well as minimal broadcast frames.

I. INTRODUCTION

Traditional roaming/handoff techniques assume that there is a shared medium between each base station. The problem that arises is that Ethernet switches are quickly replacing Ethernet hubs in many applications. Ethernet hubs are logically equivalent to the standard bus topology. Hubs do not filter any traffic; rather, they forward traffic received on any port to all ports. In this configuration it is easy to support roaming by allowing the base station's tables to be familiar with the whereabouts of the migrating hosts. The frames sent to the mobile host located on a particular cell are broadcast to all the base stations on the shared medium, and it is the base station's decision to forward it over the air medium. Several questions arise when using a switched backbone. What happens when an Ethernet switch is being used as opposed to a hub? How is roaming handled with a switch? Most vendors have not adequately addressed this issue due to its neoteric nature. Furthermore, this capability must be implemented in the supporting switch and should not be the concern of the base stations. The authors have a protocol for handling the handoffs on indoor wireless LANs when using a switched backbone. The remainder of the paper is organized as follows. Background information is given on Ethernet technologies. A problem description and solution is presented. Finally, a detailed protocol description is given.

II. EVOLUTION OF ETHERNET

Originally, Ethernet was designed as a shared medium topology, meaning that all devices or nodes on the network would contend for access on a single shared piece of medium. A mechanism is used to manage contention for the medium in Shared Ethernet. Under certain conditions of sustained high offered load, network performance degrades because devices are more likely to attempt to transmit simultaneously, causing collisions. Now, switching techniques have been implemented in the Ethernet environment to overcome the limiting factors of Shared Ethernet. Switched Ethernet isolates separate collision domains on each port, so collisions between devices do not occur [1]. This concept of a collisionless network enables Ethernet to be the backbone to distribute high-speed data in the LAN environment.

A. Shared Ethernet

In Shared Ethernet all devices share the same logical medium and are said to belong to the same collision domain. The transmission behavior of these devices is governed by the Carrier Sense Multiple Access/Collision Detection (CSMA/CD) mechanism to resolve contention for use of the medium. CSMA allows only one station to transmit at a time if there is no existing traffic on the network. If two or more stations attempt to transmit simultaneously, their frames will collide. CD will detect the collision and stop all traffic. The stations must then recover from the collision by using a technique called exponential back-off. Instead of trying to retransmit immediately, which would of course result in another collision, each station will wait a random period of time from 0 to some maximum delay d to retransmit. Since the waiting period is chosen randomly, it is unlikely that the two computers will choose the same value. One will usually begin retransmitting sooner than the other and will seize control of the medium. In the unlikely event that the two machines choose the same waiting period, another collision will occur and the process will begin again. This time, however, the value of d will double to $2d$. If there is another collision d will increase exponentially to $4d$, $8d$, and so on until some station takes control of the network and transmits its data. At that time each station will reset its maximum delay value. It should be noted that each time *d* increases the probability increases that some computer will choose a delay shorter than all others and grasp the medium. All this mediation takes time, and the more users there are trying to access the network the worse the problem can become [2]. This situation has caused Shared Ethernet to be supplanted by switching technology that is capable of handling greater throughput without experiencing the collision-induced delays in a shared environment.

B. Switched Ethernet

When switches are in place on a LAN, each port can offer its users their own piece of dedicated bandwidth. When only one machine is connected to a port it can suspend use of the CSMA/CD MAC protocol because there is no one to compete with for access to the medium. The aggregate bandwidth of an Ethernet switch can be computed by multiplying the number of switched ports, n, by the data rate and dividing this number by two, since the communication involves two parties (a sender and receiver). A single station connected to a switch can also take advantage of full duplex mode. Full duplex allows transmissions to and from the station simultaneously, which is not possible when the medium is shared. The aggregate bandwidth equation in this case changes to simply *n* times the data rate, since both parties can transmit simultaneously. Aggregate throughputs for 10 Mbps Ethernet networks jump to 20 Mbps, from 100 Mbps to 200 Mbps [3]. There are various algorithms for implementing switching, including: port switching, configuration switching, microsegmenting, dynamic switching, etc. The authors choose not to discuss those specific techniques, as the proposed solution applies to all of the techniques in practice. For an extensive explanation of switching techniques see [4]. The next paragraph provides a brief overview of switch operations.

To begin, we will assume the switch has just been turned on and has no information in its MAC tables. After the first frame is sent from the source node the switch intercepts the frame and immediately adds the source to its central address table (similar to an Ethernet bridge). The switch uses the address table to locate the port to transmit the data to the receiver on [4]. The port of the destination node is retrieved from the central address table and data is forwarded over the switch's backplane to the appropriate port, to be later forwarded to the destination node by the Ethernet controller. If the destination address is not located via the table lookup, the switch will immediately flood the frame on all ports. If the unknown device replies to the flooded data frame, the address of the node is stored in the switch's central address table for future use. As an optimization, the table is constantly sorted so the address currently used by the switch is at the top, minimizing search time. The tables are kept manageable by some sort of table management algorithm. For example, at minimum, least recently used (LRU) or least frequently used (LFU) could be choices of the algorithm. In addition to specific table management algorithms, there is usually an aging time that, once exceeded, all relevant addressees are deleted from memory. This process keeps the address table in the switch relatively small and expedites table lookups.

III. PROBLEM DESCRIPTION

When WLAN's use Switched Ethernet as a backbone network, a service disruption will occur. Due to the service disruption, frames may be flooded which wastes bandwidth. As a user migrates from the home cell (cell A) to a neighboring cell (cell B), a handoff must occur. The handoff is the process where the mobile host breaks its session with the cell it is exiting from and initiates another session with the cell to which it is migrating. The initial service disruption occurs while the second base station (base station B) updates its host list. A second service disruption may occur while the switch updates its central address table. Until this happens, the mobile host's MAC address is still associated with the old port (port A) on the switch, and data sent to the mobile host will be incorrectly forwarded out of that original port (port A). If left to its own devices, the switch will not update its tables to reflect that the mobile user can be reached through port B until it receives a frame on the new port from the mobile host. If sufficient time passes before the host transmits a frame, the table entry associated with the mobile user on port (port A) will time out, and when a new frame arrives at the switch that is destined for the mobile user, the switch will waste bandwidth by inappropriately flooding the data frame to all ports because, to the switch, the host looks like a new user. Even if we assume that the user transmits a frame before the table entry times out, all potential problems are still not resolved. By the time the first frame is received from the mobile user on the new base station, data could already have been delivered to the wrong base station and the mobile user would have to rely on upper layer protocols such as TCP to attempt a retransmission of the data. If the data is some sort of real-time data stream relying on UDP, the data is simply lost. At least one wireless networking manufacturer [6] has sought to avoid these problems by buffering data for the mobile host at the first base station (base station A). When the host successfully associates with the new base station (base station B), base station B will send a message to base station A across the wired backbone requesting all buffered data for the mobile host. This message uses the MAC address of the mobile host as the source address and thereby implicitly prompts the switch to update its forwarding tables so that all subsequent data will be forwarded appropriately. This additional buffering and secondary "proxy" request for the data introduces additional delay and variance into the network which is already an issue for delay-sensitive applications such as IP video and IP telephony. Obviously, these problems are not encountered when a Shared Ethernet backbone is used due to its broadcast nature.

IV. SOFT HANDOFF SOLUTION

We must first consider the elements or enhancements needed. Fortunately, the creation of a mobility-enhanced switch (MES) is the only addition to the current network. A MES is nothing more than a current switch with software to allow it to support mobility. The two aspects of mobility that need to be supported in the switched environment are soft handoffs and roaming. Similar techniques for soft handoffs have been implemented in cellular technology and, more recently, wireless ATM [5]. Consider the following questions posed at Figure 1: What happens when laptop 1 migrates from base station A to base station B? If laptop 1 is transmitting or receiving data, has any data been lost? What happens now if laptop 3 attempts to transmit to laptop 1 in its new location? All of these issues are addressed by the following protocol.

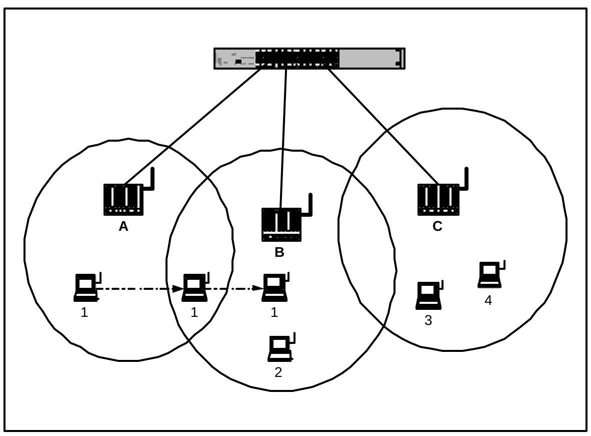

Figure 1. Wireless network with switched backbone

A. Design Overview

When laptop 1 migrates to the range covered by base station B, it must first pass through the zone covered by both base stations. This area is called the critical area and is where the majority of the action takes place as far as the handover is concerned. The handoff is a user-initiated handoff. As user 1 approaches cell B it notices that its signal is becoming weaker and must attempt a handoff soon. Once it reaches the area where it can receive beacon signals from base station B, the user radios to its home cell (cell from which it is migrating) that signals to the approaching base station to prepare for a seamless handoff. After the user enters the critical zone it begins to communicate with both base stations (to ensure no data loss). This is allowed because the MES has now updated its forwarding tables to allow data sent to user 1 to be distributed onto two separate ports (analogous to a multicast session). Finally, the user migrates to the range solely covered by base station B, the connection between base station A and user 1 has been terminated, the duplicate data queued by the switch is released, and the MES updates its tables again with the final location of user 1, which is cell B. If there is a sudden withdrawal when in the critical zone, the MES handles it the same as if the user is migrating to cell B. The forwarding tables in the switch are updated where both ports are receiving the data, however if the user withdraws back to cell A or remains in the critical zone for a pre-determined length of time, the MES will automatically assume that the user will be stable in cell A. Once this assumption has been made, the forwarding tables in the MES are updated to reflect the final location of user 1. No data has been lost; hence, this is a seamless protocol. Furthermore, this transit is unseen to other users who can send and receive frames as usual.

B. Protocol Description

Now that we have seen the overview of the handoff protocol, the technical issues are illustrated further.

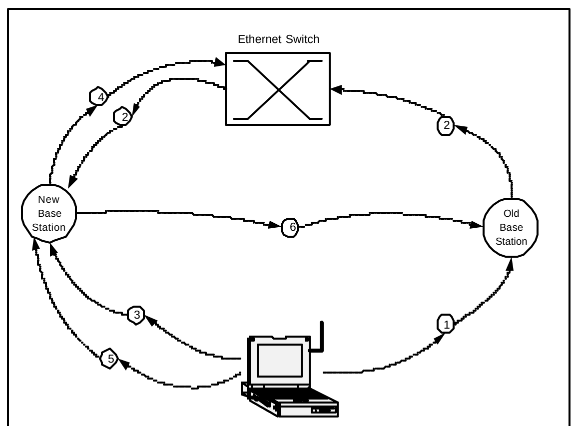

Figure 2. Soft handoff protocol

When the mobile host receives the beacons from the new base station, after a certain signal level, the mobile host sends a handover warning (l) to the current base station. Once the current base station receives the warning message, an invoke (2) message is sent to the MES and then is forwarded to the new base station. This invoke informs the MES that a handoff is beginning and to prepare to forward the incoming data to multiple ports. The invoke message also informs the new base station that a migrating host is entering its domain. Once the mobile host enters the critical region it commands the new base station to initiate the handover via the greet (3) message. The new base station in turn sends a redirect (4) message to the switch to allow for the connection to be established. At this moment in time identical data is being forwarded over two ports, however the mobile hosts only receives data from the new base station and the data from the old base station is being buffered at the switch. This buffer is implemented as a sliding window. The old base station would receive all the ACKS related to the data forwarded by the new base station in order to judge the motion of the window. If the mobile hosts suddenly decided to withdraw from the new cell, the old base station and port would know exactly where the data transmissions left off with the new base station. If this occurrs, the connection with the new base station will time out. Once the mobile host is totally out of the old base station's range, it sends a done (5) message to the new base station and the new base station sends a tear-down (6) message via the wired segment to the old base station.

V. CONCLUSION

Wireless LANs are being integrated with and, in many cases, replacing wired LANs at a surprising rate. At the same time, users are beginning to demand more from their networks than solely data traffic, but also delay-sensitive real-time streams to support key applications such as IP video and IP telephony. To effectively deliver these applications to the wireless user and provide the equivalent end-user experience to that of a wired user, the wireless network must support user mobility without adding delay. The proposed solution eliminates sources of delay caused by data being delivered to the wrong base station by providing a technique for a soft handoff from one base station to another.

REFERENCES

[1] Rockwell International Corporation, "Ethernet for Industrial Control: An Ethernet White Paper" [Online Document], Available HTTP: http://www.ab.com/networks/enetpaper.html

[2] Douglas Comer: Computer Networks and Internets. Upper Saddle River, NJ: Prentice-Hall, Inc., 1996.

[3] Anixter Inc., "Ethernet Switching: An Anixter Technology White Paper" [Online Document], Available HTTP: http://www.anixter.com/ind/pdf/Ether.pdf

[4] M. Hein and D. Griffiths, Switching Technology in the Local Network: From LAN to Switched LAN to Virtual LAN. Boston, MA: International Thomson, 1997.

[5] C-K Toh, Wireless ATM and Ad-Hoc Networks. Norwell, MA: Kluwer Academic, 1997.

[6] Cisco Systems. (2001). "Radio Technologies" [VOD stream]. Tomasheski, Mark.