Architectural Approaches for Wireless Computing in the Internet

Son Vuong, Laurent Mathy and Maria Törő[†]

Department of Computer Science
The University of British Columbia, Vancouver, Canada.
[†]KFKI Research Institute for Measurement and Computing Techniques, Budapest, Hungary.
Email: {vuong, mathy}@cs.ubc.ca, tmaria@sun60.mszki.kfki.hu

Abstract

We discuss two common architectures intended to support the internetworking of wireless networks and the Internet. The "proxy architecture" uses a TCP gateway to provide a straightforward and easy solution. On the other hand, the "end-to-end architecture" implies more design issues but is deemed to provide a more elegant solution in the long-term. The advantages and drawbacks of both approaches are discussed.

1 Introduction

Due to tremendous advances in digital signal processing as well as ever more compact and powerful computers, we have been recently witnessing a proliferation of wireless networks. Radio links, typically found in such networks, are usually characterized by deep fades whose lenght of time can be in the range of hundreds of milliseconds. Those fades, mostly caused by interferences resulting from signal reflections, can lead to rather high error rates or even to intermittent connectivity. Moreover, performance in wireless networks is worsened by the errors and temporary lost of connectivity due to handovers – which are encountered when a mobile host changes cell, that is "connects" to a new base station. This all results in rather low performance characteristics: bit error rates (BER) in the range 10^{-4} to 10^{-1} are common cases as well as bandwidth value in the range from a few kbps to a few Mbps at best.

On the other hand, wired networks can now offer a bandwidth ranging from several Mbps to several hundreds of Mbps, as well as very low error rates (BER down to 10^{-12}).

Interconnecting these rather different types of networks, not only results in routing problems [1][2], but can also lead to an overall performance which is even worse than the lowest performance achievable in the involved networks – namely, the low performance of the wireless network. This is due to "overreacting" end-to-end protocols (i.e. transport protocols) that have not been designed to deal properly with the type of errors encountered in wireless networks and therefore misinterpret them. A typical example is a transport protocol reacting to a handover as if it was a case of network congestion. Of course, the "obvious" solution of re-designing those end-to-end protocols has to be ruled out, since this would imply software upgrades in hundreds of thousands – or even millions – of computers, whether those computers be connected to a wireless network or not.

In this paper, we discuss two different approaches that are intended to solve the wireless-wired internetworking problem in the TCP/IP environment – that is, the Internet [12]. We will not discuss the problem of efficiently routing the information between the two types of network (for detail on this topic see [1] and [2]), but we will focus on the end-to-end performance problem. Needless to say that the fundamental constraint imposed on such solutions is to guarantee compatibility and seamless internetworking in the existing Internet.

2 Field of Action

In the sequel, we will encounter three different types of "entities":

- Mobile hosts: these are computers equipped with wireless modems or WirelessLAN interfaces and which can "roam" in a wireless network.
- Correspondent hosts: these are the computers involved in communications whose other end is a mobile host. A Correspondent host can either be a mobile host or any host in the Internet.
- Mobile gateways: these are computers realizing the interconnection of the wireless networks and the Internet. Such computers have (at least) one interface to the wireless network and one interface to the Internet.

Those entities are depicted in figure 1.

When we want to overcome mobility problems in the Internet, the following question has to be answered: "Where can we take appropriate actions and still be guaranteed compatibility within the Internet?".

Obviously, only two of our "entities" presented above are "aware" of the wireless environment: namely, the mobile hosts and the mobile gateways. Those "entities" are therefore the only ones where modifications to existing protocols can be made or where new protocols may be implented. As long as the protocols used in those "wireless-aware entities" are compatible with the "untouched" protocols in the correspond hosts (i.e. the TCP/IP protocol suite for a fixed host in the Internet), global compatibility should be insured.

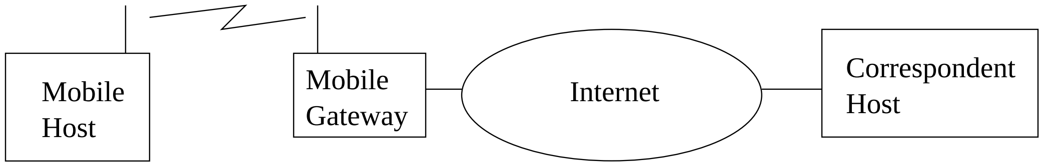

Figure 1: Wireless-Internet Interconnection.

3 The Proxy Approach

The idea behind the proxy approach is fairly simple: since the low performance encountered in the wireless networks forces TCP to "stick" to its slow-start algorithm – which produces very poor performance, splitting the TCP connections at the mobile gateway could improve performance.

Indeed, the mobile gateway operates now as a transport gateway, passing the data from a "wired TCP connection" onto a "wireless transport connection", each of those connections being provided by protocols suited to their respective environments. As long as the mobile gateway has statistically enough buffering capacity to "absorb" the difference in bandwidth between the two networks, the major performance degradation may be avoided.

Moreover, since wireless links are often characterized by small value of the Maximum Transmission Unit (MTU), that is the size of the biggest packet that can be transmitted over that link, that splitting of the transport connections will avoid the presence of small packets in the wired network, reducing the overhead and therefore enhancing the achieved performance.

This approach has been proposed by several authors [4][5][6].

It should be noted that the transport protocol used in the wireless network may either be TCP or a specially designed "wireless transport" protocol.

The main advantage of this approach is its straightforwardeness: it can easily and quickly be implemented and doesn't require any modification in the existing TCP/IP protocol suite since two "different" protocol suites are used. Also, in case of multimedia communications, a proxy acting as an application gateway would be a privilege place where service adaptation could be performed through filters.

However, this approach also suffers from an important flaw. The correspondent host and the mobile host, which are supposed to communicate with each other, are no longer in "direct contact", since they are interacting through a "relay" (i.e. the mobile gateway). This removes the end-to-end semantics supposedly associated with a transport protocol, with at least one of the host (the correspondent host) being even not aware of it! In other words, TCP is now being used as a connection-oriented network protocol. This situation may jeopardize the good working of application layer protocols like ftp or rlogin which assume an end-to-end reliability from TCP.

Moreover, unless authoring mechanisms are introduced, a proxy poses a security issue since it cannot be trusted.

Also, buffer managament in the proxy is a very complicated issue.

Finally, it should be noticed that what we called the "wireless link" between the mobile host and the mobile gateway is considered a single hop at the IP level (see [1]). Therefore, whether that link is physically realized through intra-net routing between several base stations or as a physical direct radio link between the two hosts, is not relevant at/and above the IP level. This raises the question of the usefulness of a dedicated transport connection on that link.

4 The End-to-End approach

The second approach we present here avoids the lost of the end-to-end semantics described in the previous section. Here, the mobile gateway acts as a classical router (i.e at the network layer level).

This approach has been proposed by several authors [8][7][11].

The most natural place for improvement in this approach is the link layer used for the control of the wireless link. The idea is to design a "Wireless Logical Link Control (WLLC)" protocol to deal with the specificities of the wireless link. That WLLC resides directly below IP. One of the main task performed by such a WLLC is to enhanced the error rates mesured on the wireless link. FEC, ARQ and other hydrid error control and recovery mechanisms are used at this effect [8][9].

An old argument in the scientific community is about the simultaneous use of error recovery mechanisms in the link layer (here, WLLC) and the transport layer (here, TCP). However, it has been shown [3] that the interaction of those mechanisms in the two layers becomes beneficial if the error rate of the network exceeds a given threshold (depending on the network characteristics), which is below the current and expected error rates shown by wireless networks. Moreover, it has also been shown [10] that error detection codes can be implemented in a very efficient way.

Those error control mechanisms implies that the WLLC protocol be connection- oriented. The resulting connection establishment phase will provide an opportunity for the mobile host and the mobile router to negotiate optimized protocol parameters. Such a connection-oriented WLLC will allow TCP/IP header compression.

Futhermore, the retransmission timers used by the WLLC protocol could be easily tuned to accomodate the characteristics of the wireless link and therefore provide a better response time. Also, by choosing retransmission timer values smaller than the ones in use at the transport level, the WLLC might take advantage of "transport communication pauses" to "get rid of" outstanding frames (by passing them to the mobile host. Those timer values are preferably dynamic. This should enhanced the throughput observed by the applications.

The IP processes on the wireless side of the mobile router and in the mobile host may also be configured to "cheat" on the actual MTU size supported on the wireless link. Indeed, having those two IP processes advertising a bigger MTU size and having the WLLC to take care of any segmentation/reassembly, would reduce the overhead observed in the wired network and therefore enhanced the achieved throughput.

Although the TCP process in the mobile host is in direct interaction with the TCP process in the correspondent host, some otpimizations are still possible at that level.

For instance, it has been shown [7] that if the mobile host TCP process is given the information that a handover just took place, it could then recover the resulting missing TCP segments fatser by forcing the correspondent host TCP to invoke the fast retransmit and fast recovery algorithm [12](by sending three consecutives acknowledgments) instead of the slow-start algorithm [12] (which is normally invoked). Although this requires a modification of TCP in the mobile host, compatibility is guaranteed with all existing TCP implementation.

The advantage of the end-to-end approach is that it does not imply any change in semantics of any of the protocols in the TCP/IP protocol suite.

However, it must also be recognized that the use of error correction in the link layer will introduce variable delays which will be observable by the applications.

Finally, it may appear a bit "old-fashion" to use error control in the link layer. However, it has been shown that the implementations can be kept simple [9]. Moreover, WLLC could also be designed in a modular way, so that object-oriented techniques can be applied.

5 Conclusion

We presented two different architectural approaches to wireless computing in the Internet.

The proxy approach is a straightforward solution that does not require any changes in the TCP/IP protocol suite. However, it results in a lost of TCP's end-to-end semantics.

The end-to-end approach keeps TCP's semantics but at the cost of more modifications in the wireless network protocol stack and "haevier" implementations.

Nevertherless, we do believe that this latter approach is more suited for the internetworking of wireless network with the current Internet, since it provides better long-term perspectives.

Finally, Multimedia communications were barely considered, since it is not clear at all to us whether the Internet will ever be able to support them in a satisfactory way, mainly in a wireless context.

References

[1] C. Perkins, editor: IP Mobility Support, *Internet Draft (Work in Progress)*,1996.

[2] A. Myles, D. B. Johnson, C. Perkins: A Mobile Host Protocol Supporting Route Optimization and Authentication, *IEEE Journal on Selected Areas in Communications, Vol. 13, No. 5*, June 1995.

[3] A. DeSimone, M. C. Chuah, O-C. Yue: Throughput Performance of Transport-Layer Protocols over Wireless LANs, *Proc. of IEEE GlobeCom*, 1993.

[4] J. Kiiskinen et al.: Data Channel Service for Wireless Telephone Links, *Proc. of Second Intl. Mobile Computing Conf.*, 1996.

[5] R. Yavatkar, N. Bhagawat: Improving End-to-End Performance of TCP over Mobile Internetworks, *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, December 1994.

[6] A. Bakre, B.R. Badrinath: I-TCP: Indirect TCP for Mobile Hosts, *Proc. of 15th Intl. Conf. on Distributed Computing Systems*, 1995.

[7] R. Cáceres, L. Iftode: Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments, *IEEE Journal on Selected Area in Communications, Vol. 13, No. 5*, June 1995.

[8] S. Nanda, R. Ejzak, B. T. Doshi: A Retransmission Scheme for Circuit-Mode Data on Wireless Links, *IEEE Journal on Selected Area in Communications, Vol. 12, No. 8*, October 1994.

[9] J. F. Whitehead, R. Krishnamoorthy: FEC, ARQ, and Throughput of Radio LANs, *Proc. of Conf. on Vehicular Technology*, 1994.

[10] D. C. Feldmeier: Fast Software Implementation of Error Detection Codes, *IEEE/ACM Transactions on Networking, Vol. 3, No. 6*, December 1995.

[11] D. Raychaudhuri, N. D. Wilson: ATM-Based Transport Architecture for Multiservices Wireless Personal Communication Networks, *IEEE Journal on Selected Area in Communications, Vol 12, No. 8*, October 1992.

[12] R. Stevens: TCP/IP Illustrated, Vol.1, The Protocols, Addison-Wesley, 576p, 1994.