An Overview of Cellular IP

Andrew T. Campbell, Javier Gomez
Center for Telecommunications Research, Columbia University, New York
[campbell, javierg]@comet.columbia.edu

András G. Valkó*
Ericsson Research
andras.valko@lt.eth.ericsson.se

Abstract

Recent initiatives to add mobility to the Internet and packet data services to third generation cellular systems are being considered by emerging mobile service providers as possible candidate solutions for the delivery of IP data to mobile users. Both of these two candidates have a number of shortcomings, however. Mobile IP represents a simple and scalable global mobility solution but is not appropriate in support of fast and seamless handoff control. In contrast, third generation cellular systems offer smooth mobility support but are built on complex networking infrastructure that lacks the flexibility offered by IP-based solutions. In this paper we present an overview of Cellular IP which represents a 'third way' combining the strengths of both approaches without inheriting their weaknesses. Cellular IP combines the capability of cellular networks to provide smooth fast handoff and efficient location management of active and idle mobile users with the inherent flexibility, robustness and scalability found in IP networks.

1 Introduction

The development of commodity-based palmtop devices with built in high-speed packet radio access to the Internet will have a major impact on the mobile telecommunications industry and the way we communicate. The availability of cheap, ubiquitous and reliable wireless Internet access will shift the service base traditionally found in mobile telecommunication networks toward emerging wireless Internet Service Providers (ISPs). This will result in significant demands being placed on both existing and next-generation cellular and IP networks.

In this paper, we present an overview of Cellular IP [4], an Internet host mobility protocol [5] that takes an alternative approach to that found in mobile telecommunications (e.g., General Packet Radio Service [1]) and in IP networking (Mobile IP [2]). Cellular IP represents a new mobile host protocol that is optimized to provide access to a Mobile IP enabled Internet in support of fast moving wireless hosts. Cellular IP incorporates a number of important cellular principles but remains firmly based on IP design principles allowing Cellular IP to scale from pico to metropolitan area installations.

The paper is structured as follows. In Section 2, we discuss the concept of an IP-based Cellular Internet where Cellular IP provides micro-mobility support interworking with Mobile-IP which provides macro-mobility support between Cellular IP wireless access networks. Following this in Section 3 we present an overview of the Cellular IP routing, handoff and paging algorithms. The Cellular IP distributed location management and routing algorithms lend themselves to a simple, efficient and low cost implementation for host mobility requiring no new packet formats, encapsulation or address space allocation beyond what is present in IP. In Section 4 we provide some concluding remarks.

2 Cellular Internet

Recent initiatives to add mobility to the Internet mostly focus on the issue of address translation [3] through introduction of location directories and address translation agents. In these protocols (e.g., Mobile IP) packets addressed to a mobile host are delivered using regular IP routing to a temporary address assigned to the mobile host at its actual point of attachment. This approach results in simple and scalable schemes that offer global mobility. Mobile IP is not appropriate, however, for fast mobility and smooth handoff because after each migration a local address must be obtained and communicated to a possibly distant location directory or home agent (HA).

Cellular mobile telephony systems are based on a different concept from that of Mobile IP. Instead of aiming at global mobility support, cellular systems are optimized to provide fast and smooth handoff within restricted geographical areas. In the area of coverage, mobile users have wireless access to the mobility unaware global telephony network. A scalable forwarding protocol interconnects distinct cellular networks to support roaming between them.

Even in limited geographical areas, however, the number of users can grow to a point where using fast lookups for per user data bases is no longer viable. In addition, mobility management requires mobile hosts to send registration information after migration. The resulting signalling overhead has significant impact on the performance of wireless access networks. To overcome this problem, cellular telephony systems require mobile hosts to register after every migration only when they are engaged in 'active' calls. In contrast, 'idle' mobile hosts send registration messages less frequently and as a result can roam large areas without loading the network and the mobility management system. In this case, the location of idle mobile hosts is only approximately known to the network. To establish a call to an idle mobile host, the mobile host must be searched for in a limited set of cells. This feature of *passive connectivity* allows cellular networks to accomodate a very large number of users at any instance without overloading the network with large volumes of mobility management signalling information.

Cellular networks offer a number of desirable features which if applied correctly could enhance the performace of future wireless IP networks without loosing any of the important flexibility, scalability and robustness properties that characterise IP networks. However, there are fundamental architectural differences between cellular and IP networks that make the application of cellular techniques to IP very challenging. Cellular telephony systems rely on the restrictive "circuit" model that requires connection establishment prior to communication. In contrast, IP networks perform routing on a per packet basis. In addition, current cellular systems are strictly based on hierarchical networks based on costly mobile-aware nodes (e.g., MSC). We believe that a future "Cellular Internet" should be based on IP, inheriting its simplicity, flexibility and robustness. A Cellular Internet should leverage mobility management and handoff techniques found in cellular networks. A single scalable host mobility protocol should be capable of flexibly supporting pico, campus and metropolitan area networks based on a set of simple and cheap network nodes that can be easily interconnected to form arbitrary topologies and operate without prior configuration.

3 Protocol Overview

In what follows, we provide an overview of the Cellular IP features and algorithms; that is, the Cellular IP routing, handoff and paging algorithms. For a full discussion, specification and evaluation of Cellular IP see [4], [5], [6], respectively.

3.1 Features

Cellular IP inherits cellular systems principles for mobility management, passive connectivity and handoff control, but is designed based on the IP paradigm. The universal component of a Cellular IP network is the *base station* which serves as a wireless access point but at the same time routes IP packets and integrates cellular control functionality traditionally found in Mobile Switching Centers (MSC) and Base Station Controllers (BSC). The base stations are built on regular IP forwarding engines, but IP routing is replaced by Cellular IP routing and location management. The Cellular IP network is connected to the Internet via a *gateway* router. Mobility between gateways (i.e., Cellular IP access networks) is managed by Mobile IP while mobility within access networks is handled by Cellular IP. Mobile hosts attached to the network use the IP address of the gateway as their Mobile IP care-of address. Figure 1 illustrates the path of the packets addressed to a mobile host. Assuming Mobile IPv4 [2] and no route optimization [7], packets will be first routed to the host's home agent and then tunnelled to the gateway. The gateway "detunnels" packets and forwards them toward base stations. Inside the Cellular IP network, mobile hosts are identified by their home addresses and data packets are routed without tunnelling or address conversion. As discussed later, the Cellular IP routing protocol ensures that packets are delivered to the host's actual location. Packets transmitted by mobile hosts are first routed to the gateway and from there on to the Internet.

In Cellular IP, location management and handoff support are integrated with routing. To minimize control messaging, regular data packets transmitted by mobile hosts are used to establish host location information. *Uplink* packets are routed from mobile to the gateway on a hop-by-hop basis. The path taken by these packets is cached in base stations. To route *downlink* packets addressed to a mobile host the path

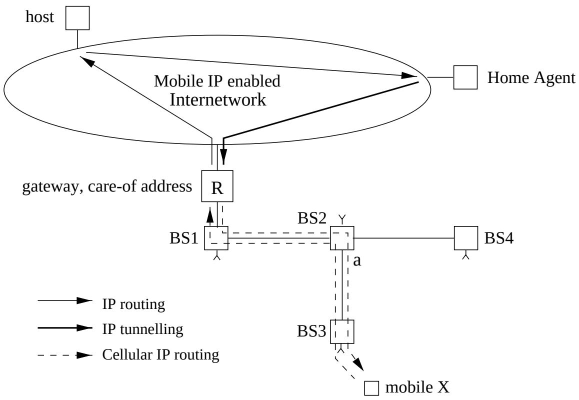

Figure 1: Cellular IP Access Network

used by recent packets transmitted by the host is reversed. When the mobile host has no data to transmit then it periodically sends empty IP packets to the gateway to maintain its downlink routing state. Following the principle of passive connectivity mobile hosts that have not received packets for a certain period of time allow their downlink soft-state routes to timeout and be cleared from the routing cache. In order to route packets to idle hosts a Cellular IP mechanism called *paging* is used.

3.2 Routing

The Cellular IP gateway periodically broadcasts a beacon packet that is flooded in the access network. Base stations record the interface they last received this beacon through and use it to route packets toward the gateway. All packets transmitted by mobile hosts regardless of their destination address are routed to the gateway using these routes.

As these packets pass each node on route to the gateway their route information is recorded as follows. Each base station maintains a *routing cache*. When a data packet originated by a mobile host enters a base station the local routing cache stores the IP address of the source mobile host and the interface over which the packet entered the node. In the scenario illustrated in Figure 1 data packets are transmitted by a mobile host with IP address **X** and enter **BS2** through its interface **a**. In the routing cache of **BS2** this is indicated by a mapping **(X,a)**. This mapping remains valid for a system specific time *route-timeout* and its validity is renewed by each data packet that traverses the same interface coming from the same mobile. As long as the mobile host is regularly sending data packets, base stations along the path between the mobile host's actual location and the gateway maintain valid entries in their routing cache forming a soft-state route between the mobile host and gateway nodes. Packets addressed to the same mobile host are routed on a hop-by-hop basis using the established routing cache.

A mobile host may sometimes wish to maintain its routing cache mappings even though it is not regularly transmitting data packets. A typical example for this is when the host is the receiver of a stream of UDP packets and has no data to transmit. To keep its routing cache mappings valid the mobile host transmits *route-update packets* at regular intervals called *route-update time*. These packets are empty data packets addressed to the gateway. Route-update packets have the same effect on routing cache as normal data packets; however, they do not leave Cellular IP access networks.

3.3 Handoff

The Cellular IP *hard handoff* algorithm is based on a simplistic approach to mobility management that supports fast and simple handoff at the price of potentially some packet loss. Handoff is initiated by mobile hosts. Hosts listen to beacons transmitted by base stations and initiate handoff based on signal strength measurements. To perform a handoff a mobile host has to tune its radio to the new base station and send a route-update packet. This creates routing cache mappings on route to the gateway hence configuring the downlink route to the new base station. Handoff latency is the time that elapses between the handoff and the arrival of the first packet through the new route. For hard handoff this equals the round-trip time between the mobile host and the cross-over point which is the gateway in the worst case. During this time, downlink packets may be lost. The mappings associated with the old base station are not cleared at handoff, rather, they timeout as the associated soft-state timers expire.

Before the mappings timeout, a period exists when both the old and new downlink routes are valid and packets are delivered through both base stations. This feature is used in the Cellular IP *semisoft handoff* algorithm that improves handoff performance but still suits the lightweight nature of the base protocol providing probabilistic guarantees instead of fully eliminating packet loss. Semisoft handoff adds one additional state variable to the existing mobile state maintained at mobile hosts and base stations. The semisoft handoff procedure has two components. First, in order to reduce handoff latency, the routing cache mappings associated with the new base station must be created before the actual handoff takes place. When the mobile host initiates a handoff, it sends a *semisoft packet* to the new base station and immediately returns to listening to the old base station. While the host is still in contact with the old base station, the semisoft packet configure routing cache mappings associated with the new base station. After a *semisoft delay*, the host can perform a regular handoff. The semisoft delay can be an arbitrary value between the mobile-gateway round-trip time and the route-timeout. The delay ensures that by the time the host tunes its radio to the new base station, its downlink packets are delivered through both the old and new base stations.

While the semisoft packet ensures that the mobile host continues to receive packets immediately after handoff, it does not, however, fully assure smooth handoff. Depending on the network topology and traffic conditions, the time to transmit packets from the cross-over point to the old and new base stations may be different and the packet streams transmitted through the two base stations will typically not be synchronized at the mobile host. If the new base station "lags behind" the old base station, the mobile host may receive duplicate packets. Reception of duplicate packets in this case is not disruptive to application operations. If, however, the new base station "gets ahead" then packets will be deemed to be missing from the data stream observed at the receiving mobile host. The second component of the semisoft handoff procedure is based on the observation that perfect synchronization of the two streams is not necessary. The condition can be eliminated by temporarily introducing into the new path a constant delay sufficient to compensate, with high probability, the time difference between the two streams. This can be best achieved at the cross-over switch that understands that a semisoft handoff is in progress due to the fact that a semisoft packet has arrived from a mobile host that has a mapping to another interface. The mapping created by the semisoft packet has a flag to indicate that downlink packets routed by this mapping must pass a "delay device" before transmission. After handoff, the mobile host will send data or route-update packets along the new path which will clear this flag and cause all packets in the delay device to be forwarded to the mobile host.

3.4 Paging

Cellular IP defines an *idle mobile host* as one that has not received data packets for a system specific time *active-state-timeout*. In this respect, idle mobile hosts allow their respective soft-state routing cache mappings to time out. These hosts transmit *paging-update packets* at regular intervals defined by *paging-update-time*. The paging-update packet is an empty IP packet addressed to the gateway that is distinguished from a route-update packet by its IP type parameter. Paging-update packets are sent to the base station that offers the best signal quality. Similar to data and route-update packets, paging-update packets are routed on a hop-by-hop basis to the gateway. Base stations may optionally maintain *paging cache*. A paging cache has the same format and operation as a routing cache except for two differences. First, paging cache mappings have a longer timeout period called *paging-timeout*. Second, paging cache mappings are updated by any packet sent by mobile hosts including paging-update packets. In contrast, routing cache mappings are updated by data and route-update packets sent by mobile hosts. This results in idle mobile hosts having mappings in paging caches but not in routing caches. In addition, active mobile hosts will have mappings in both types of cache. Packets addressed to a mobile host are normally routed by routing cache mappings. Paging occurs when a packet is addressed to an idle mobile host and the gateway or base stations find no valid routing cache mapping for the destination. If the base station has no paging cache, it will forward the packet to all its interfaces except for the one the packet came through. Paging cache is used to avoid broadcast search procedures found in cellular systems. Base stations that have paging cache will only forward the paging packet if the destination has a valid paging cache mapping and only to the mapped interface(s). Without any paging cache the first packet addressed to an idle mobile host is broadcast in the access network. While the packet does not experience extra delay it does, however, load the access network. Using paging caches, the network operator can restrict the paging load in exchange for memory and processing cost.

Idle mobile hosts that receive a packet move from idle to active state, start their active-state-timer and immediately transmit a route-update packet. This ensures that routing cache mappings are established quickly potentially limiting any further flooding of messages to the mobile host.

4 Conclusion

This paper has provided an overview of the Cellular IP protocol which we have submitted [5] to the IETF IP Routing for Wireless/Mobile Hosts Working Group for discussion. Cellular IP represents a new approach to IP host mobility that incorporates a number of important cellular system features but remains firmly rooted in IP. A number of issues remain to be studied. While the current protocol supports best effort traffic only, an extension with simple quality of service provisioning (e.g., based on differentiated services [8]) is being investigated. Another important issue that we are currently working on within the Cellular IP Project [9] relates to authentication and security. User authentication information can be provided in paging-update and route-update packets.

Evaluation of the baseline protocol operating in a pico-cellular wireless testbed is underway. Early results look promising in terms of the performance of hard and semisoft handoff, the network overhead associated with supporting mobility management and the scalability of the protocol to be configured to meet the requirements of widely different environments [6].

References

[1] G. Brasche, B. Walke, "Concepts, Services and Protocols of the New GSM Phase 2+ General Packet Radio Service," IEEE Communications Magazine, pp. 94-104, August 1997.

[2] Charles Perkins, editor, "IP Mobility Support," Internet RFC 2002, October 1996.

[3] Pravin Bhagwat, Charles Perkins, Satish Tripathi, "Network Layer Mobility: an Architecture and Survey," IEEE Personal Communications Magazine, Vol. 3, No. 3, pp. 54-64, June 1996.

[4] András G. Valkó, "Cellular IP: A New Approach to Internet Host Mobility," ACM Computer Communication Review, January 1999.

[5] A. Valkó, A. Campbell, J. Gomez, "Cellular IP," Internet Draft, draft-valko-cellularip-00.txt, November 1998, Work in Progress.

[6] A. Valkó, J. Gomez, S. Kim, A. Campbell, "Performance of Cellular IP Access Networks," Proc. of 6th IFIP International Workshop on Protocols for High Speed Networks (PfHSN'99), Salem, August 1999.

[7] David B. Johnson, Charles Perkins, "Route Optimization in Mobile IP," Internet Draft, draft-ietf-mobileip-optim-07.txt, November 1998, Work in Progress.

[8] D. Clarc, J. Wroclawski, "An Approach to Service Allocation in the Internet," Internet Draft, draft-clarc-diff-svc-alloc.00.txt, July 1997, Work in Progress.

[9] The Cellular IP Project at Columbia University http://comet.columbia.edu/cellularip