CHAPTER 6

Beyond 3G: 4G IP-Based Mobile Networks

Donal O'Mahony and Linda Doyle

6.1 Introduction

The architecture of both the PSTN and also the first three generations of mobile telephone networks [1] were shaped by two main considerations. First, the networks were required to effectively provide a single service—circuit switched voice communications—across a network with low-capability edge nodes and a large amount of intelligence within the network. Any additional services that could be deployed on this infrastructure were looked upon as a bonus. Second, they required the operators to be able to charge for network usage based on the destination and duration of the call. Mobility between network operators was only worth doing if the home network operator stood to gain financially from this possibility.

The evolution of the Internet took a different path. In the early years, it was assumed that communities of users collaborated for the benefit of all. Since no organization wanted to act as a central network operator, the interior of the network was kept as simple as possible while allowing users to offer such services as they wished at the edges. The possibility of users moving around was catered for at the edges by Mobile IP and was implemented by agents external to the network itself, or at least located within edge routers.

In recent years, we have seen the beginning of convergence between the telecommunications and Internet communities. The PSTN has become the main Internet access network for residential users, and there is a huge industry focus on reengineering the channel structure of wireless telephony networks to accommodate data streams. In the other camp, the Internet telephony industry has proved that it is possible to build a production voice service on a packet-switched Internet-based network infrastructure, and the makers of large circuit switches have conceded that there are huge economic advantages to shifting over to packet switching.

The rapid adoption of mobile telephony, most notably in the Scandinavian countries, demonstrates another major trend that shows a demand on behalf of users to avail of the same communications on the move as were available through fixed lines.

In the midst of the great changes being brought about by the trends outlined above, the research community is considering what form the next (fourth) generation of fixed and mobile communications systems will take. A popular view in the telecommunications industry is that the 4G should evolve naturally from the 2G and 3G with incremental improvements being brought about without any fundamental architectural changes being necessary. We do not believe that this is the correct approach. In the following sections, we set out what we believe are the main drivers shaping the design of 4G systems. We briefly survey related work and then describe the progress we are making with our 4G testbed.

6.2 Drivers for the 4G Architecture

We adopt the view that the forthcoming 4G mobile systems offer us an opportunity to take a fresh approach in designing a network driven by services that people want now and in the future with less emphasis on the services that they wanted in the past. In the following sections, we examine five fundamental requirements underlying our concept of 4G mobile systems.

6.2.1 Support for IP-Based Traffic

It is clear from the rapid growth of the Internet in the late 1990s that IP forms an effective delivery mechanism for the kinds of network services that users are interested in availing of. The advent of VoIP has shown that person-to-person audio communication can easily be carried out across a packet-based IP network in spite of some difficulties in delivering consistently low end-to-end delays across the current infrastructure. This fact, when taken together with very low forecast growth rates for voice as compared with data, point to a future where voice makes up a very small proportion of total traffic. It is imperative, then, that the architecture of 4G networks is determined primarily by the need to deliver a quality IP service. The ability to handle voice and other streams with real-time constraints is a secondary (but essential) goal. The problem of delivering a predictable quality of service over wireless networks is a significant challenge in its own right and is fully addressed in the chapters of Part II of this book.

6.2.2 Excellent Mobility Support

The adoption of mobile telephony services offered by 1G and 2G voice systems has surpassed many people's expectations. In many countries, most notably in the Scandinavian region, more than 60% of the overall population makes use of mobile telephones, and the trend elsewhere is to follow the same path. It is forecast that there will be almost 1 billion subscribers to GSM, the most successful 2G system, by 2005.

The mobility facilitated by cellular telephones is often referred to as terminal mobility. An additional form of mobility envisaged for both the fixed and mobile phone networks is personal mobility, which allows a person to move from one terminal to another and have his calls and environment follow him as he registers on different mobile or fixed terminals. The Universal Personal Telephony service allows a user to be contactable via one personal number. Intelligence within the network, coupled with information on the users preferences and likely movement patterns can forward the call to the appropriate terminal or to an automated response system.

In 4G systems, we must assume that all users of the network are potentially mobile with a sizeable proportion of them communicating via wireless terminals. Users must be contactable via a single (albeit multifaceted) identity. A means must be found to map from this identity to an address to which packets can be routed. Control of this mapping must lie firmly with the user who can modify the destination of the mapping and regulate the access that callers may have to it. In an environment where the path from source to destination may cross many different network domains, it would be unwise to associate this mapping with a single network operator. It is imperative that 4G networks provide a single consistent means of identifying users and allow this identity to be controlled by the user and efficiently mapped to a mutable destination.

Where 4G nodes are in motion, they may need to change their point of attachment to the network while a flow of packets is in progress. The degree to which this changes the path from source to destination depends on the topology of the network and the route chosen. The 2G and 3G cellular systems allow handoff where a mobile node shifts its point of attachment but stays within the same network domain. Typically, a change in domain is referred to as roaming, and active connections cannot be maintained under these circumstances. This implicit two-level tracking of a user's location (i.e., current domain and current position with that domain) can be generalized into an N-level mobility tracking scheme that can allow handoff in any circumstances where the network topology renders it possible.

6.2.3 Support for Many Different Wireless Technologies

The 1G, 2G, and 3G mobile systems relied on the use of spectrum reserved for public land mobile use and licensed by a very small number of network operators in each country. Differences in allocation timing and strategy of the spectrum have led to the need for multimode phones capable of adapting to use the spectrum available in a particular region.

In 4G systems, it is likely that a large number of different radio technologies may be used for network access. One trend that is very apparent is the use of radio spectrum in the less regulated *Industrial Scientific and Medical* (ISM) band. We have seen the emergence of Bluetooth [2] radio offering very short-range radio links of below 1 Mbps. Although this radio was originally developed as an easy way to replace cables interconnecting adjacent equipment, it has recently been standardized by the IEEE as standard number 802.15.1 as a more general purpose *wireless personal area network* (WPAN) technology. Also operating in the same band is the IEEE 802.11b WLAN radio system. This offers a throughput of 11 Mbps with a range of around 100 feet and lesser speeds over longer distances. This technology has been embraced by the market and is being used to provide high-bandwidth services to private users in buildings and campuses and also to public users in built-up areas and in airports. More advanced versions [3] of the standard are in preparation, including 802.11a operating in the 5-GHz band (which will deliver higher speeds of 54 Mbps) and 802.11g (which uses the same frequency band as 802.11b, but allows the data rate to be increased to 34 Mbps and higher).

At the other end of the spectrum, areas of very low population density or users at sea or in the air may be better served by satellite systems using a completely different set of radio technologies.

Existing 2G and 3G cellular may be useful in between these two extremes. A 4G node should be capable of adapting its radio capabilities to take maximum advantage of available spectrum.

6.2.4 Free from Unnecessary Operator Linkage

The GSM system was developed by European telecommunications companies in the mid- to late 1980s primarily as a mobile extension to the PSTN. The GSM model envisaged that users would subscribe to an operator who would build a cellular network infrastructure that would track the user as he moved from one location to another, making every effort to maximize the availability of service. In common with the dominant PSTN billing model in Europe at the time, all usage of GSM services was to be metered, and since the only way to pay for this was via the home operator, every action carried out by the mobile handset is with reference to the home network operator. Even when a user roams into a new domain, contact is made with the home network to establish a link with a billable entity before any calls can be made.

Two GSM handsets cannot communicate with each other directly; rather, they must each first authenticate themselves to the network, be linked with their billing details, and thereafter, operator mediated communication can take place. This mode of operation is consistent with the fact that the operator in effect owns the spectrum and is entitled to individually regulate and meter each access to it.

Where spectrum ownership is much more open, such as is the case in the ISM band, such restrictions are completely unnecessary and highly inefficient. Much of the dynamism that has taken place in wireless communications in recent years is the result of the easy access to such spectrum, and it is likely that this trend will continue in the future.

Where access to spectrum is not an issue, pairs or groups of nodes can form ad hoc networks to allow direct communication between nodes, and if appropriate, nodes can collaborate, relaying each other's traffic. Naturally, issues such as the need for user authentication occur in this kind of any-to-any direct communication, but arguably, these need to be solved in an operator-independent way.

Once the special position of the operator is removed, there remains the problem that a wireless node wishing to communicate with a node outside its range cannot do so unless an intermediate node relays their packets to either another wireless node or on to the fixed network. If we could find a means of making real-time payments across a link, this would relieve us of the need to be associated with a billable entity and also allow us to motive individual nodes to cooperate with us to relay traffic.

This motivation (financial or otherwise) could be used in a sparsely populated area to allow an individual wireless node to act as a packet relay between two out-of-range nodes. The payment method would allow the relay to be compensated for the battery drain and the usage of bandwidth that might otherwise be available to it.

In a heavily populated area, the same motivation could be used to encourage organizations to erect networks of wireless network access points in places like university campuses and shopping malls. Organizations that undertook this effort to any great scale would become the network operators of the 4G, but the fact that no special status is required to become an operator should ensure healthy competition.

6.2.5 Support for End-to-End Security

The security features inherent in 2G and 3G mobile systems are focused on two main services. First, the mobile users must be authenticated to the network operator. This authentication is generally limited to associating the user with a billing relationship that is operating satisfactorily. Where accounts are prepaid, this billing relationship often has no stored details on the identity of the user. There is no end-to-end exchange of credentials between a mobile user and their peer at the other end of the link.

The second service supported by 2G and 3G mobiles is content encryption. While this does deter attacks using simple scanning devices, it is no substitute for genuine end-to-end confidentiality.

In the 4G, mobile and fixed nodes will interact with each other without reference to their relationship to an operator. It is imperative that protocols and procedures are devised to allow the users of these nodes to authenticate one or more facets of each other's identity to a degree necessary to achieve the communication they desire.

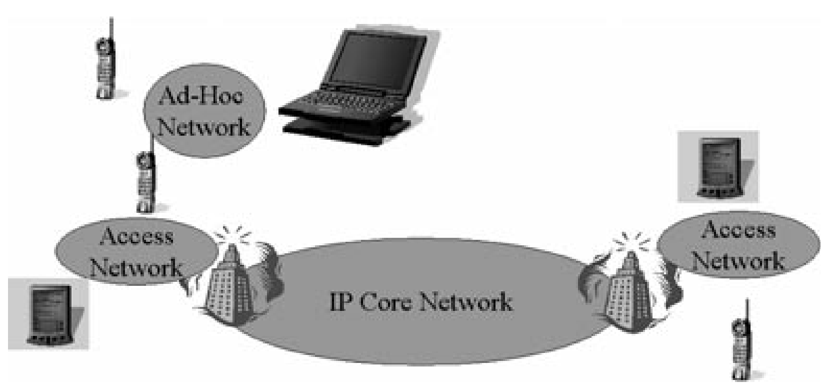
6.3 4G Architecture and Research Issues

We envisage that the fourth-generation mobile networks of the future will be based around an IP core network, which will, over time, completely displace both the fixed PSTN and also the 2G and 3G mobile voice networks. The architecture, as shown in Figure 6.1, will be based on delivering an IP transport service to a population where every user is potentially mobile and a large proportion of them make frequent use of wireless nodes to interact with people and services on the network.

Untethered users will gain access to the fixed network using a variety of different radio technologies. These will include short-range wireless systems such as Bluetooth and 802.11 as well as perhaps new access methods made available on the spectrum currently in use by 2G or 3G mobile voice systems. Individuals and organizations will install and operate wireless access points making use of a real-time payment system to provide such quasi-network operators with their financial incentive.

High-capability nodes will be equipped with software radios, which carry out as many radio functions as possible by executing signal processing

FIGURE 6.1
Components of a 4G network.



software on general purpose processors. Thus, the nodes will be able to communicate using whichever radio technology is expedient given its location and financial resources.

Ad hoc network protocols will be used to bind together groups of nodes for local communication and also to provide a link between each node and a range of geographically close fixed network access points. The motivation for this collaborative behavior may derive from the fact that the users have authenticated each other as belonging to the same group (e.g., workers in the same office or paramedics at the scene of an accident), or some kind of micro-payment method may be used to allow relaying nodes and network access points to profit from their activities.

Interactive links to other users will be established based on a user identity. In order to support both terminal and personal mobility, a directory service will be required to map between an identity or identity facet and an address to which packets can be routed. This service should be standalone and independent of any particular network infrastructure. The mapping may also be quite complex, allowing a user to build a profile of when and to whom he is available. In cases where confidentiality of location may be an issue, artificial relay points may be used to disguise the position of a node from the caller.

Before a call can be set up, each participant will need to authenticate themselves to the other parties within the call. Typically, users will have a multifaceted identity, and it may be inappropriate to make use of all of these facets for any given call. Facets may also depend on one another and progressive authentication may take place. For example, to gain entry to an office building, it may only be necessary to authenticate the fact that one is an employee. If an individual wants to open a safe containing financial documents, it may be necessary to also authenticate the fact that the individual's organizational role is financial director. Such identity facets would be inappropriate for use in a coffee shop where the same user wished to traverse a high-speed fixed network access point.

Once the parties have authenticated each other, end-to-end communication can take place with appropriate authentication and confidentiality measures being applied to the data traffic.

Where prebuilt access infrastructure does not exist, or where communication is taking place in a local region, nodes may resort to the formation of ad hoc networks. While this greatly enhances a node's ability to communicate, it also raises a number of additional research issues. Collaboration is the essence of ad hoc networking and will naturally take place where the members of the network belong to the same organization or have some other preestablished motivation to work together. In a public context—for instance, a number of wireless nodes encountering each other in a shopping mall—the motivation may be less obvious. There is a need for security protocols to allow nodes to selectively reveal information about different facets of their identity with an aim of maximizing the level of cooperation that is possible. If nodes are to offer relaying services to others, there will be a need for some kind of real-time payment method to allow a node to be compensated for the use of its resources (battery power and more).

Once an ad hoc network grows beyond a small number of nodes, there is a need to develop sophisticated routing protocols to allow nodes to relay for each other, thus maximizing their interconnectivity. It should also be possible for a wireless node that is within range of a fixed network access point to relay for other nodes within its ad hoc network.

Applications and services in the 4G will be built outside the network in keeping with the Internet tradition of keeping the core simple, fast, and efficient. Today, many of the Internet applications rely on entities contacting each other based on a network address. This is problematic for a number of reasons. First, it leads to a demand by end entities for the fixed assignment of addresses. If the address space is finite or inefficiently allocated (as is the ease with IP version 4), this leads to address shortages and also causes problems when these end entities move with respect to the network. We envisage that in the future, applications will evolve to where addresses are deemphasized in favor of the use of more abstract names. Mobile users will be contacted by name, with appropriate servers to perform the mapping between address and the necessary routing information. Clearly such a mapping must be done in such a way that the user has control over what kind of routing information is held and who may gain access to it. Where a user is in motion, appropriate mechanisms must be devised to maintain a connection to a "name" as the route to the node is changing.

6.4 4G Research Efforts

Broadly speaking, research into the form of 4G mobile systems is being carried out by two somewhat overlapping communities. The first of these is the mainstream mobile telecommunications industry. The second community has come from a data networking background and is driven by technological progress in the area of wireless data networks leading to quite a different perspective on the path ahead. We will outline the direction being taken by both communities before elaborating further on our own thoughts on the fourth generation.

At the time of this writing, the telecommunications community's main source of revenue is based on a huge population of users of 2G mobile telephones. Now they are faced with the impending launch of third-generation mobile systems. Most of the architectural decisions underlying the 3G were made with an eye to preserving the large investment in building the 2G network and retaining the customer relationship with the user population. It was thought that if the core network architecture was preserved, then new radio access methods (such as UTRAN) could be progressively introduced, and gradually, more support for data traffic and Internet access would be made available.

At the time of writing, prospects for the 3G look bleak. The mobile telephone market appears to be near saturation point, and the introduction of highe-rspeed data services has failed to produce the anticipated excitement and surge in user demand. Looking ahead to the data capabilities of 3G systems, it seems unlikely that these will match user expectations. In Europe, the auctioning off of spectrum at very high prices has placed a huge debt burden on aspiring 3G network operators that may be difficult to service.

Against this background, there is a reticence to contemplate revolutionary technologies that would divert attention away from the 3G, and consequently, new innovations are being discussed as add-ons to mature 3G networks rather than as technologies that would render it obsolete.

The two major proponents of 3G mobile systems—namely, the European UMTS Community and the U.S.-dominated cdma2000—have formed consortia (3GPP and 3GPP2, respectively) that are developing common standards for the provision of data services to users. In the case of 3GPP, the core network architecture is very much shaped by GSM with special purpose network entities added to support the transfer of IP datagrams between network endpoints or to the Internet. The 3GPP2, on the other hand, is much more Internet-like.

The Wireless World Research Forum, a consortium of many leading mobile telecommunications companies, has produced a "Book of Visions" [4], which summarizes the thinking of their member companies on what are important research topics for the future. This recognizes the need for greater support of the IP protocols and the importance of new radio access technologies. While there is recognition of the potential of ad hoc networking technology there is no mention of any architectural changes being made to accommodate this.

The rapid uptake of commodity WLAN products based on the IEEE 802.11b standard has given rise to many popular public wireless services that promise to deliver far superior data services to any currently envisaged in the 3G. In order to combat this threat, a number of organizations are attempting to integrate WLAN into 2.5G and 3G networks as simply a new radio access technique without any changes to architecture of the core network. This has led to a proposal to the 3GPP standardization effort, made in late 2001, that a study be commissioned to determine how WLAN technology such as IEEE 802.11 and HiperWLAN2 could be integrated into UMTS. Attention to date has focused on a loose coupling where access to the WLAN is regulated with reference to the UMTS subscriber database. End-to-end traffic would pass from the WLAN to the Internet without going through the UMTS network entities.

The research community has been more daring [5] and has proposed an architecture where many heterogeneous networks, among them WLAN and UMTS, form subnetworks of a larger 4G system whose architecture is very Internet-like with mobility being handled at the IP level.

The second community of researchers that is interested in this area come from a wireless Internet perspective and have focused on the efficient transport of IP traffic over wireless channels with provision of varying qualities of service. These themes are elaborated on in other chapters of this book. In terms of mobility handling, there appears to be broad support for the idea that Mobile IP can be used to handle user movement between domains (roaming), while some alternative system such as Cellular IP, HAWAII, or TIMIP be used to deal with movements within a wireless domain. While the concept of handover is very much a part of these efforts, this community does not appear to be attempting to build a system that could aspire to displace the current global fixed and mobile telephony infrastructure. In a similar way, their efforts in the security arena do not appear to be targeted towards replacing telephone numbers by a single globally acceptable identity that could be cryptographically asserted.

6.5 The NTRG 4G Testbed

At Trinity College in Dublin, Ireland, the *Networks and Telecommunications Research Group* (NTRG) has been investigating the form of 4G mobile systems since 1998. We have ongoing projects investigating many different aspects of 4G technology from applications though to physical layer issues. The individual projects are bound together by virtue of their individual contributions towards our 4G testbed.

Since we envisage that the 4G mobile nodes of the future will be constructed using commodity computing platforms, our target nodes are general purpose PC workstations, and where portability is important, laptop and palmtop variations of these. In order to keep a consistent operating environment across all platforms and to allow our work to integrate with popularly available applications, we have chosen to develop the Microsoft Win32 environment as supported by Windows 2000 on PC and laptop and make use of Windows CE on handheld and palmtop environment.

6.5.1 The Layered Architecture

Components of our 4G environment are implemented as standalone layers each realized by a single main thread. The interlayer interface is very simple, consisting of primitives to send information upwards or downwards through the stack and to attach a blackboard of parameters to each request that can be used by any layer through which the data passes. A sample layer stack is depicted in Figure 6.2.

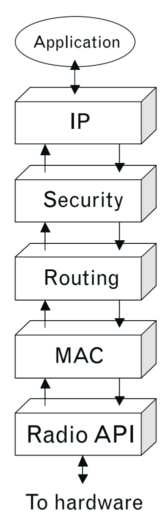
6.5.2 Wireless Alternatives

Different radio hardware can be accommodated by writing a layer that interacts directly with the hardware and presents the simple interlayer transfer interface to whatever is above it. In this way, we have been able to perform our wireless experiments with links as diverse as Infra-Red (IRDA), Bluetooth, IEEE 802.11, and also a very simple half-duplex radio we constructed in-house, which uses amateur RFs and allows us to experiment at a very low level.

6.5.3 Software Radio

Ultimately, we expect to be able to replace the real radios at the bottom of this protocol stack with a software radio operating in conjunction with a wideband front end, which would allow operation across a wide frequency band with the chosen form of modulation being performed in software. The individual building blocks of the software radio (such as channelization, coding, modulation, and demodulation) will be realized by individual layers with an appropriate stack being assembled to deliver the desired radio waveform. Thus far, we have receiver-only systems that implement a variety of forms of amplitude and frequency demodulation and can rapidly switch between the two if necessary.

FIGURE 6.2
The layered structure of the 4G testbed.



6.5.4 Routing Protocols

The subject of routing protocols enabling the formation of ad hoc networks of nodes has been under active study by the research community for some time now [6]. Although many different protocols have been simulated, very few implementations exist that can test these protocols across a real radio channel under different mobility scenarios.

Using our layered architecture, we initially implemented the *Dynamic Source Routing* (DSR) Protocol. This simple reactive protocol only begins to try to establish a route to a destination when there is data to be sent. This is in contrast to a proactive protocol, which attempts to maintain knowledge of the state of the network so that it is already in possession of sufficient routing routine information when data needs to be sent. We have also implemented a hybrid *Zone Routine Protocol* (ZRP), which is proactive for nodes that are in a nearby zone and reactive for those that are farther away. We anticipate that the use of these protocols across real wireless channels will give us a unique insight into their properties.

6.5.5 Emulation Facilities

When routing protocols or mobility aware applications are being developed and tested, one of the major problems is to expose the evolving software to particular mobility scenarios without conducting all debugging on the move and out of doors. Our initial approach to this challenge was to develop a layer (which we call the datagram layer) that emulates the radio broadcast environment across a collection of Internet links on the local area network.

Each of the emulated nodes is assigned an IP address, and the emulation layer in each node is told what other nodes are visible to it in radio terms. When a packet arrives to be sent on the emulated radio interface, it is encapsulated in an IP datagram and sent to each of the visible nodes.

This effectively allows us to construct ad hoc networks made up of processes running on any Internet-connected host. By constructing a relay layer that moves packets from one protocol stack to another, we can freely intermix nodes that are running on real wireless nodes and others that are sitting on the emulated radio layers. An example of this is shown in Figure 6.3.

While the above is an effective debug and test tool, emulated wireless nodes either see each other or they do not, and transmissions always get through without transmission errors. The ability of nodes to see each other or not is also statically configured.

In an attempt to improve our radio emulation environment, we have devised a system that we refer to as a reality emulator. In place of the radio layer in each node's protocol stack, we place a reality emulator client layer. Each of these clients connects via sockets to a server, as shown in Figure 6.4,

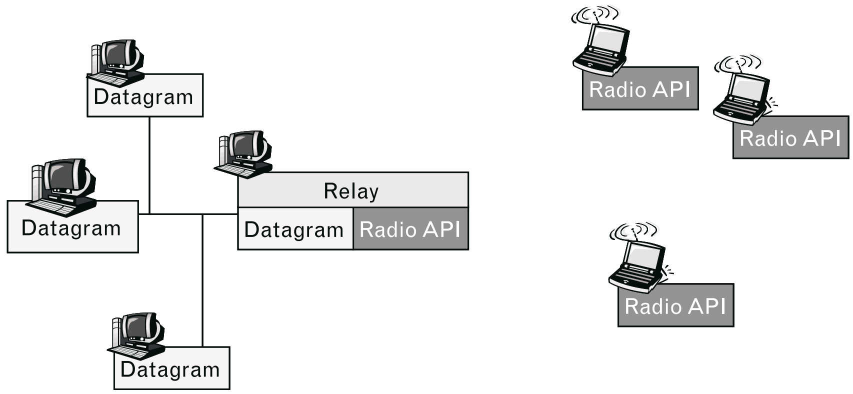


FIGURE 6.3
An ad hoc network emulated wireless nodes of genuine and emulated wireless nodes.

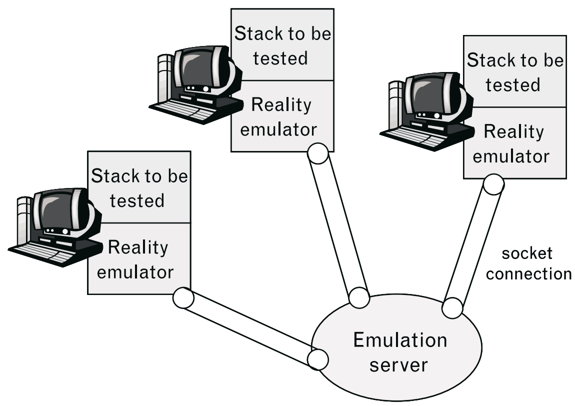


FIGURE 6.4
The reality emulator.

which emulates the way in which a radio channel will behave as nodes move with respect to one another and also encounter collisions in transmission.

When a node initializes itself, it connects to the emulation server where it is given an initial geographic position. By interacting with a graphical user interface on the server, a designer can control the transmission range of each radio and their movement with respect to one another. Figure 6.5 shows the server's view of a collection of emulated nodes moving around the Trinity College campus.

From the point of view of the client protocol stack, the behavior of the link is identical to that experienced by the node when running across a real radio link, and provided that the number of emulated nodes is kept reasonably small, the emulator can support substantial amounts of traffic including real-time voice streams. The system is particularly good at exposing the routing software to particular node movement scenarios that might otherwise be hard to achieve.

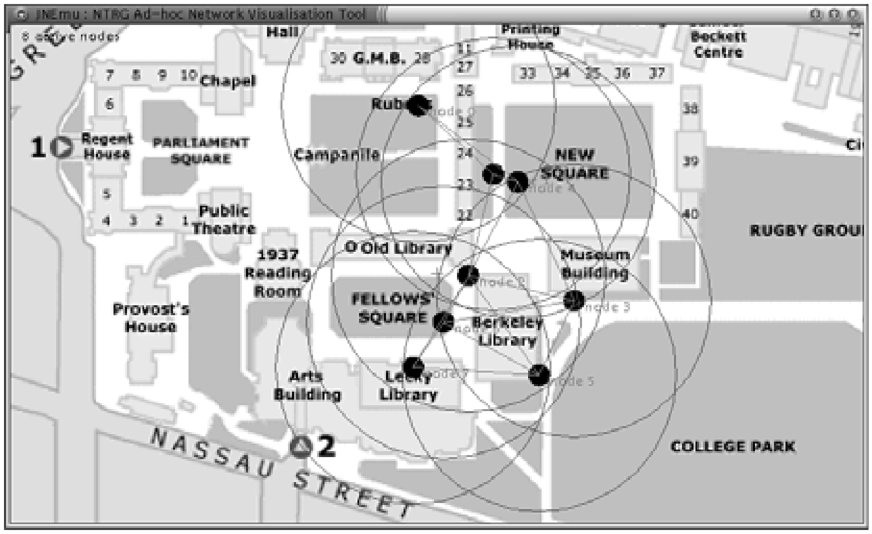


FIGURE 6.5 *The emulation server showing eight nodes and their radio ranges.*

6.5.6 The Security Architecture

We envisage that all users in the fourth generation are potentially mobile, and a large proportion of them will avail of wireless devices. This means that nodes will typically be interacting with peer nodes with little information on the identity represented by that peer.

Authentication in the physical world is typically achieved using a multiplicity of cards that people carry around in their purse or wallet. These may contain basic information about a person, such as might be present on a driving license. Other cards may contain information on an individual affiliation with a bank or perhaps more personal information such as a card detailing his or her blood type.

Individuals produce cards detailing different facets of their identity only as the need arises. For example, to make a cash withdrawal at a bank, they may need to produce both a driving license and a bank card. They might hand over details relating on their medical record only to someone who had already proved he was a medical professional.

The above exchanges are characterized by individuals entering into a negotiation dialog where credentials are exchanged in order of increasing importance as trust is progressively built between the individuals. This process concludes either when the shared trust has reached its highest level, or it has exceeded that required by the communication exchange.

At present, we are designing a system to do this kind of credential exchange between nodes in an ad hoc network. When a node wishes to avail of a service on another system, this will cause the credential exchange agents to enter into a dialog where details on different facets of a user's identity are exchanged according to a *credential release policy* (CRP).

Nodes may authenticate neighboring nodes up to a level at which they are happy to relay traffic through them. A considerably higher level of authentication may be needed before a pair of nodes may be willing to enter into a specified application dialog. Figure 6.6 depicts the security agent that carries out this authentication.

Based on the authenticated identity facets, which are shared between the two nodes involved in the trust negotiation, groups can be formed to share a common purpose. This will allow shared keys to be negotiated, which can be given to all members of the group to allow shared access to resources. Similar group formation systems have been developed for Internet multicast environments that can serve as a model for this [7]. This can be used for such applications as a walkie-talkie type system where all users hear all traffic or, indeed, any other form of data-based group collaboration.

6.5.7 Real-Time Payment

Most of the flaws in 2G and 3g mobile systems can be traced to the fact that the major technical decisions defining the architecture have been made based on one overriding concern, namely that of generating revenue for the network operators.

If the promise of the 4G is to be fulfilled, the focus must be on enabling connectivity between network users without tying this connectivity to a user-operator subscription. Clearly some other way needs to be found to allow providers of network infrastructure (even if this is just a single relay node) to be rewarded for making this available to the general population of nodes.

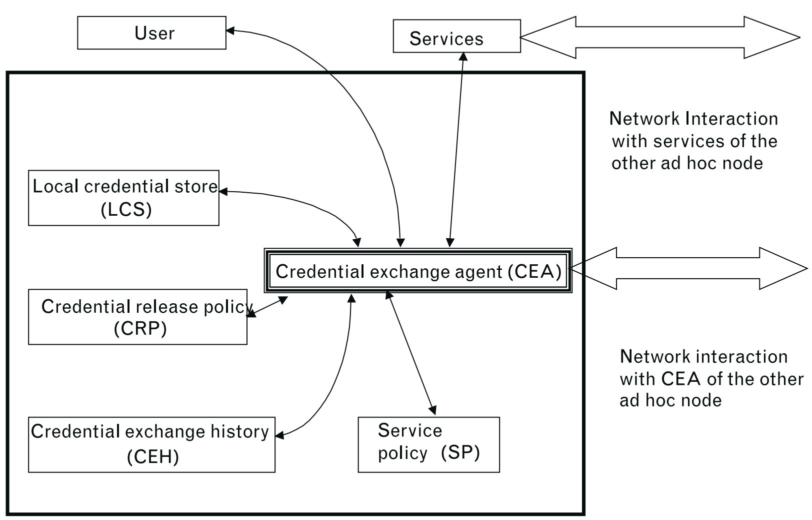


FIGURE 6.6
The 4G security agent.

Figure 6.7 depicts a multiparty micro-payment system [8] developed within our research group that allows a stream of cryptographic payment tokens to be interspersed with the normal data packets that make up the end-to-end flow. Before this can commence, a pricing phase was undertaken where each party (or network operator) along the path offered their services for a particular price. Once the contract is agreed to, the payment tokens can be captured by nodes along the path and redeemed for an agreed proportion of the overall payment for the traffic. We plan to adapt this system to support ad hoc operation where the end-to-end route may change over the lifetime of the communication.

6.5.8 Applications

The principle application we have used on our 4G network testbed has been simple Web access, where pages containing HTML as well as complex multimedia data have been delivered to handhelds moving in a wireless network. We have also performed some experiments using the wireless link for video image data. More recently, we have developed a simple point-to-point telephony application using the Session Initiation Protocol for signaling. In the future we will modify this to integrate elements of our security architecture, in particular, to incorporate the notion of multifacetted user identities. There are a whole host of issues to be resolved before we can support mobility thorough the mapping of this identity information to addressing information that works well with both fixed and ad hoc networks.

6.6 Concluding Remarks

We have outlined above some of the problems inherent in the architectural design of 2G and 3G mobile systems. Arising from this, we have enumerated some of what we believe are imperatives for the design of a new 4G architecture. These ideas are being progressed by the mobile Internet research community and to a lesser extent, by researchers in mobile telecommunications. In order to progress our ideas for 4G systems, our research group has embarked on a number of distinct projects dealing with different aspects of the overall system. Each of these projects contributes to the construction of a testbed based on the use of commodity PC and PDA hardware running common operating systems. The testbed supports wireless mobility via a range of technologies from infrared to software radio and allows the experimentation with real ad hoc networks that are engineered to integrate with populations of fixed nodes. Security concerns are also catered for both in terms of node authentication and also as a means of payment for resources consumed. The utility of the 4G systems will be demonstrated with

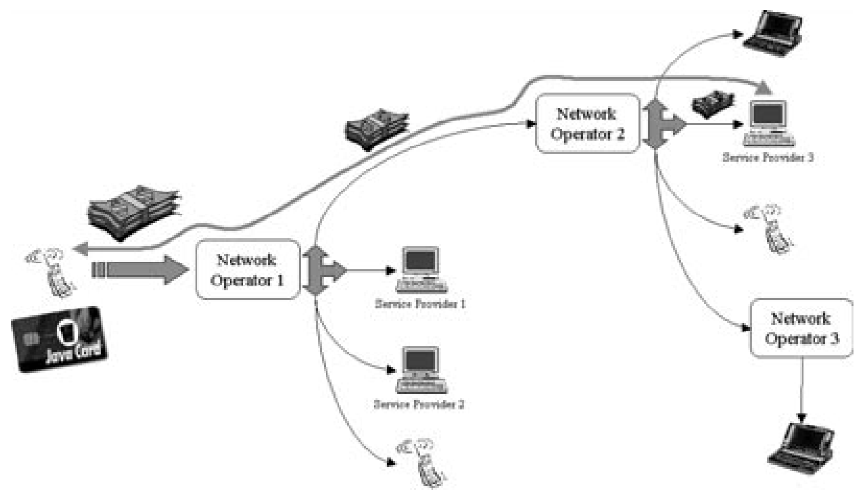


FIGURE 6.7
Real-time, multiparty micro-payment.

noninteractive data-based applications as well as those, such as telephony, that involve continuous multimedia streams.

REFERENCES

[1] O'Mahony, D., "Universal Mobile Telecommunications Systems: The Fusion of Fixed and Mobile Networks," *IEEE Internet Computing*, Vol. 2, No. 1, January/February 1998, pp. 49–56.

[2] Bisdikian, C., "An Overview of the Bluetooth Wireless Technology," *IEEE Communications*, Vol. 39, No. 12, December 2001, pp. 86–94.

[3] Leeper, D. G., "A Long-Term View of Short-Range Wireless," *IEEE Computer*, Vol. 34, No. 6, June 2001, pp. 39–44.

[4] Wireless World Research Forum, "Book of Visions," December 2001, http://www.wireless-world-research.org.

[5] Bria, A., et al., "Fourth-Generation Wireless Infrastructures: Scenarios and Research Challenges," *IEEE Personal Communications*, Vol. 8, No. 6, December 2001, pp. 25–31.

[6] Royer, E., and C. -K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications Magazine*, April 1999, pp. 46–55.

[7] Waldvogel, M., et al., "The VersaKey Framework: Versatile Group Key Management," *IEEE Journal on Selected Areas in Communications*, Special Issue on Middleware, Vol. 17, No. 8, August 1999.

[8] Peirce, M., and D. O'Mahony, "Flexible Real-Time Payment Methods for Mobile Communications," *IEEE Personal Communications*, Vol. 6, No. 6, December 1999, pp. 44–55.