To appear in Proceedings of IEEE GLOBECOM2000
Workshop on Service Portability, December 1, 2000

Wireless Open Service Networks

P. H. Hsiao, A. Hwang, H. T. Kung, and D. Vlah

Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138, U.S.A.

September 30, 2000

Abstract

We describe a networking infrastructure, called wireless open service networks, that allows service providers such as content providers to compete in open marketplaces. Using this open infrastructure, customers are empowered with dynamic choice of a variety of services at various levels, beyond just their traditional choices in network access services. In the meantime, service providers are given the opportunity to bundle services at multiple levels to maximize the performance of service delivery to end users.

For the implementation of these wireless open service networks, we present a Service Address Resolution Protocol (SARP). Using SARP requests customers can discover available services over a local wireless network. Using SARP replies service providers can make their services known to customers. SARP resides at the link-layer and is light-weight, to allow easy and ubiquitous deployment. SARP can be viewed as a natural extension of the widely used Address Resolution Protocol (ARP) which discovers Ethernet addresses.

We give application examples to illustrate these concepts. In addition, we discuss the design of SARP and our implementation experience.

Keywords: open service, wireless networks, mobile networking, service discovery, service bundling.

1 Introduction

A variety of Internet appliances equipped with low-cost and short-range radios such as IEEE 802.11 [3], Bluetooth [1] and HomeRF [6] are emerging. As prices of these radios continue dropping due to economy of scale, wireless Internet appliances are expected to be widespread. In fact, it is widely believed that they will surpass the wired population in the future.

With the majority of services likely to reside on the wired Internet, wireless devices will primarily desire to reach the wired network services from within stub wireless networks. In addition, with Mobile IP [8], such appliances themselves can also be servers capable of receiving connection requests to support a range of server-based applications such as Voice over IP.

Under these scenarios, we are interested in designing a wireless networking infrastructure to provide the most value for mobile end devices as well as service providers at all networking layers. Note that mobile users and end devices will have diverse tasks that require services from the network. It is already the case that today's users have choice in using remote network services. The Internet and the Web offer a broad range of possibilities: often selecting a remote service is a matter of visiting a web site. Therefore we suggest that a desirable wireless networking infrastructure would be a *wireless open service network* as depicted by Figure 1. In such a network, a mobile user or end device D can choose a service gateway G among a number of available candidate service gateways, to best fit the current application.

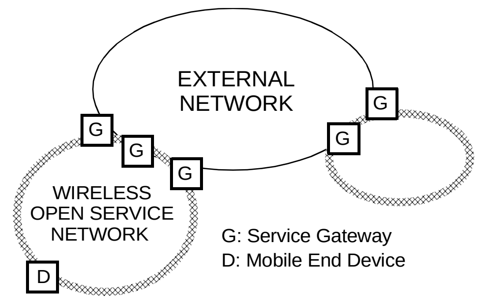

Figure 1. Open service network for stub wireless networks addressed in this paper, providing mobile device D with choice in service gateways G.

However, current mobile users do not have choice in wireless networks regarding services such as access to the wide-area network, configuration service like DHCP [4], data caching service, etc. Monolithic service models of cellular networks and data networks such as CDPD [10] and Metricom [7] prevail. A user cannot dynamically select between the access link types and services to suit his or her needs.

Wireless networks have greater potential than wired environments in providing service choice for consumers. While wired networks are constrained by the inflexible physical configuration such as wire connectivity, such matters do not constrain wireless devices. In addition, radios naturally share the airspace, allowing devices to be fully aware of services in a local region. Thus it is natural trying to allow devices or end users to make choices in the local network for services they utilize?

It is important to have dynamic choice of services in a local wireless network because the remote service qualities can be dominated by the properties of local services. For example, if the network access link is poor, the quality of overall service will likely be poor regardless of the quality of the remote end. Thus the cooperation of remote and local network services can ultimately optimize end-to-end service delivery.

This paper suggests use of wireless open service networks for the benefit of both end customers and network service providers. Such a service network is a collection of local wireless networks, via which an end user/device can connect to a set of service gateways. On each local wireless network, service providers at various networking layers compete with each other. Devices may switch service affiliation dynamically to obtain the best network services at each layer that are appropriate for the users and their present applications.

For efficient implementation of these wireless open service networks, we introduce a link-layer service address resolution protocol (SARP) for distributed service discovery. Using SARP, users discover wireless services, whereas service providers announce their services to potential users. As an extension of ARP [9], SARP is light-weight to allow easy and ubiquitous deployment. Using this service discovery protocol, the users can dictate how choice is made in selecting local network services—the user can control whether the data content source, local network provider, mobile device, application, or the user him/herself make the final selection. This is a capability beyond those provided by current wireless access networks.

This paper is organized as follows. In Section 2, we describe the motivation and structure of wireless open service networks, the Service Address Resolution Protocol, and the important concept of "service bundling." Section 3 details the architecture and protocols including usage examples of the protocols and our system implementation experience. We discuss related work in Section 4, and conclude in Section 5.

2 Open Service Networks

We give a working example to motivate open service networks and the Service Address Resolution Protocol. We will refer to this example throughout our discussion.

Suppose that inhabitants of a small but populous island called Big Orange drive around the grid-arranged roads in cars equipped with wireless mobile appliances. The roadsides are lined with access points for the appliances to connect. The Big Orange people often sit in heavy traffic so they browse the Internet or listen to digital music to spend their idle time. Fortunately, the people of Big Orange have mobile devices that use SARP for efficient access of networks and services.

2.1 Open Service

We would like the world to be different. Unlike current wireless networks, we would like to have substantial choice in using network services. Using our example of the citizens of Big Orange, suppose that all the citizens' mobile devices have IEEE 802.11 radios [3]. Citizen A likes to check stock quotes using a low-latency path to the wired Internet where the financial brokerage servers are. A's mobile device queries the local broadcast domain for a service gateway that provides low-latency access service. Service provider FastAccess specializes in low-bandwidth low-latency access connectivity while service provider WideLoad specializes in high-bandwidth connections. To A's service query, FastAccess' gateways reply with their local hardware network addresses (i.e., their Ethernet or MAC addresses) so that one of them can become the default gateway for A.

On the other hand, citizen B would like to view news video broadcasts. To B's query for a "high-bandwidth gateway," WideLoad's gateways, not FastAccess's, reply with their local hardware addresses.

In the wireless networks of Big Orange, multiple providers of network access can coexist giving the citizens a set of service gateways from which they can dynamically choose. In addition, the providers themselves such as FastAccess and WideLoad can choose access service for their short-range wireless gateways. That is, the service providers form hierarchical layers of multi-hop wireless access.

We formally define an open service network to be a collection of local wireless networks, where there can be one or more providers at each layer for any local network service. These services may be access service (routing packets out to the wide-area network), DNS (name resolution), DHCP (host configuration service), home/foreign agent (for Mobile IP), data caching, etc.

This is not the case with current wireless networks such as cellular networks and data networks like CDPD, 802.11 Ethernet, and Metricom. Currently, a single provider is responsible for all service between the user's device and the wired-wireless gateway.

In contrast, an open service network may have different access providers everywhere from the user to a service gateway. This is similar to the wired network where the user can choose from ISPs, and ISPs can choose from backbone providers.

Even so, wired networks do not completely fit our definition of open service networks because a single provider controls all services in its portion of the network. For example, users can select an ISP for routing packets into the wide area network but cannot change who that ISP is dynamically for each application. Open standard cellular networks such as GSM [11] also fall into this category.

In summary, with open service networks, consumers at all points in the wireless network are presented with services at various networking layers for them to choose. There may be layer-1 service providers to which end devices connect using some wireless channels, layer-2 service providers from which end devices receive MAC-layer services, layer-3 service providers which will route IP packets to the external world, etc. In this hierarchy of network services, the layer-1 provider gateways can be deployed in high density using short-range radios to support high-bandwidth end user access. In these networks, inexpensive radio devices such as Bluetooth can easily reach the wide-area network at high bandwidths, via multiple wireless hops.

However, how will customers in a local wireless network find available services at various layers? As described in the next section, we use a service address resolution protocol (SARP) as the mechanism.

2.2 Service Address Resolution

Returning to the citizens of Big Orange, it is clear that a mobile device may not reside in a particular access point's radio range for too long. The Big Orange residents prefer a simple and light-weight mechanism to discover local network services dynamically. Each device will simply query the local wireless network for a "gateway." The gateways in the area reply with their hardware network addresses so that the end device may begin using one of them for access or service.

It happens that they enjoy sharing digital music and offering digital radio services for each other. Citizen C may configure her mobile device to be a "classical genre" radio station for her neighbors while citizen D may specialize in "rap." The citizens can directly scan the local network for their favorite genre radio stations with little difficulties.

In a wireless environment, a distributed mechanism best achieves discovery of local services. Only each mobile device itself has the correct point of view of desired services or contents, and capability to discover in-range local network services because of overlapping but non-concentric radio ranges.

To avoid scanning all possible radio channels, we use a shared channel to broadcast service address discovery information. This facilitates an open access network because only the device itself controls the discovery process. That is not the case with centralized look-up services.

We introduce a mechanism for achieving an open service network that we term Service Address Resolution Protocol (SARP). SARP extends the traditional Ethernet ARP for discovery of various network services on the local network. For example, consider access service (routing) to the wide area network. An end device wishing to send data to a remote node by use of the default gateway entry in its routing table, will "discover" the default gateway's MAC address via ARP.

However, during subsequent communication, the end device does not utilize the actual IP address of the default gateway. It is only the MAC address that dictates where the data is forwarded. In this sense, the end host need not preconfigure or learn the default gateway address via bootstrap mechanisms. There can be a symbolic name, or a Service Address that identifies a node as the host for "default gateway" or access service on the local network. This Service Address can be well known, much like the default port for WWW queries, the value 80. Using a well-known Service Address thus saves the management overhead of configuring default IP gateways or running DHCP servers.

As described in detail in Section 3, SARP is designed to be light-weight and similar to Ethernet link layer protocols.

2.3 Service Bundling

Service bundling refers to the use of a set of compatible services at various layers, which together provide a streamlined delivery of end-to-end services. Service bundling is one of the major benefits offered by wireless open service networks suggested by this paper.

In service bundling, a service address need not correspond to an exact service type or provider identity. It can represent a certain set of bundled services to reflect an application's or even data source's preference as to what combinations of network services the mobile device should use.

Let us return to the island of Big Orange. Recall that citizen A is a financial aficionado that constantly checks stock quotes. Frustrated by the lack of low-latency paths to her brokerage, citizen A signs up for brokerage service with Faithful Investments. Faithful Investments has partnered with FastAccess to provide Faithful customers with dedicated low-latency access to its stock quote service. Instead of "low-latency gateway," Citizen A queries for Service Address "Faithful" on the local network. FastAccess' privileged service gateway replies with its hardware network address for A to use as default gateway. Thus citizen A enjoys a bundled service consisting both of content services for stock quotes and the desired low-latency access services offered by FastAccess.

Citizen B who keeps abreast of current events similarly benefits from using bundled services. News content providers place wireless caches near the highways so that the mobile devices on vehicles access the video broadcasts in a small number of hops. B's mobile device issues a SARP query for service address "news broadcast" and chooses among the many news content providers that send SARP replies. Citizen B actually prefers a particular content provider's service. B may elect to directly connect to his favorite news site Orangy News which happens to have placed their caches on WideLoad's access network. Upon receiving a SARP query for "Orangy News broadcast," WideLoad gateways that are in range of B reply with their hardware network address. Using the bundled WideLoad service, B enjoys a high-quality delivery of Orangy News' broadcasts resulting from a high-bandwidth connection to a locally cached content.

In general, bundling services with the data source can expect to yield increased service quality. The data content source may have exclusive agreements with faster routing and better cache providers for its content. The end users utilize the co-branded local cache instead of a slower farther cache provided by a general service provider with no agreements with the content provider. Thus, with our open service networks the service providers have the opportunity to bundle their services to optimize end-to-end service quality for the end users.

With SARP the user can specify how choice is to be made in selecting a local network service. The content source itself can persuade the end device to use a local caching service. Similar to current monolithic providers, the access service provider can try to dictate use of services for the users. Even the user's mobile device or the current application may participate in selection of underlying local services. This is user empowerment.

3 Architecture and Protocols

In this section we describe the role of different nodes in the network and the SARP protocol. In addition, we give examples which demonstrate gateway selection based on SARP messages.

3.1 Architecture

As shown in Figure 2, there are three types of nodes present in the network: external gateways, internal gateways, and end nodes. Gateway nodes are nodes at the edge of a

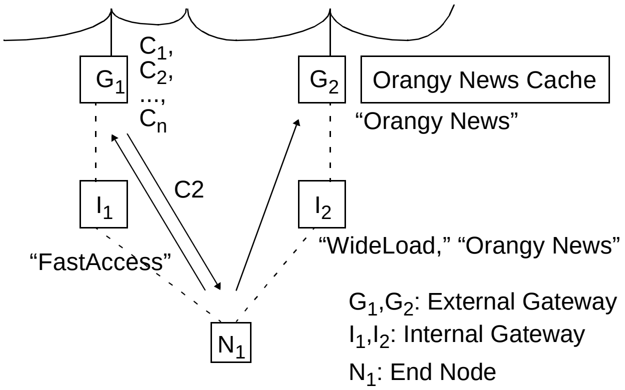

Figure 2. Example use of Address and Content SARP Requests, and a list of different types of nodes.

wireless local area network. In the upstream direction, an external gateway has a wired connection to the external wired Internet, while an internal gateway may have a wireless connection to an external gateway or another internal gateway. In the downstream direction, an external or internal gateway may have a wireless connection to another internal gateway or an end node, which is a mobile end device.

Internal gateways listen for messages from external gateway or other internal gateways above. They combine received information and periodically send it downward over another interface.

The end nodes communicate to gateways. They use the received SARP reply messages to make appropriate link selections. In Figure 2, node N_1 can select either gateway I_1 or I_2.

An important feature of gateways is a well-known control channel over which all SARP messages are sent. Additional interfaces on other channels are dedicated for data traffic. In Figure 2, end node N_1 would use the control channel to reach both I_1 and I_2, learn their data channels, and make a choice. Once it switches to a data channel, it only communicates with a single internal gateway.

3.2 Description of SARP

Three types of messages are exchanged in SARP: Address, Value, and Content messages. The formats of these messages are shown in Figure 3. Each message can be either a Request or Reply, determined by the Op field. In a Request, the Service MAC and Channel fields are empty; in a Reply, they are filled out with the hardware network address and the data channel of the answering node.

SARP is designed to resolve addresses based on two kinds of information: globally disseminated low-volume, and locally kept high-volume information. Address and Value Requests are designed for the first kind; they are broadcast in the local area, and answered by direct neighbors. Content Requests are designed for the second kind; they are flooded from end nodes toward gateway nodes, and answered by any node matching the Request.

An Address Request carries one or more symbolic names in fields denoted SA_1 (Service Address 1), SA_2, ..., SA_N. The length of each name is stored in fields Len_1, Len_2, ..., Len_N. There is a set of names assigned to each node, such as the service type or content provider; when an Address Request matches any of these names, the node returns an Address Reply with filled Service MAC and Channel fields.

A Value Request carries a set of attributes. The number of attributes varies depending on Type. Attributes in Value Requests encode quantifiable characteristics such as bandwidth, latency, cost, jitter, etc. Internal gateways receive attributes from one or more sources, combine them using well-known composition rules [13], and propagate the result downward.

A Content Request contains a single textual query, stored in the Query field. The length of the query is stored in the Len field. The query is substring-matched against a list of contents at each node that receives a Request. In contrast to Address and Value Requests, the Content Request is flooded toward external gateways over multiple hops, and the source route is recorded. Whenever a node matches the query, it changes the Operation field to a Reply, and returns the packet along the recorded source route. At each return hop, the hardware network address and data channel of the hop is written into the Service MAC and Channel fields. End nodes receiving the reply get just the last hop's hardware network address and data channel; this is sufficient to select the correct internal gateway.

3.3 Examples

In this section we show examples of use for each type of SARP message.

3.3.1 Service Bundling

Figure 2 shows how service bundling is achieved using SARP, continuing the example of our Big Orange citizens. Some of the nodes in the figure have been assigned symbolic names by their operators: "FastAccess" and "WideLoad" are assigned to I_1 and I_2, while "Orangy News" is assigned to G_2. In addition, I_2 has the "Orangy News" name because it has been disseminated on periodic Address Replies from G_2.

When end node N_1 sends an Address Request for "Orangy News," only internal gateway I_2 replies because it is the only

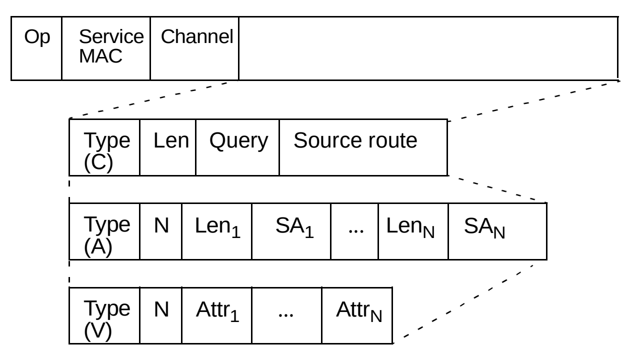

Figure 3. Three SARP packet types: Content, Address, and Value Request/Reply.

node with a matching name. The reply contains the MAC address and data channel of I_2, which N_1 uses to connect to I_2.

Users' service improves after this selection because of the transparent Web cache (such as Squid [12]) that Orangy News had deployed at G_2. When users send requests for news that are cached at G_2, G_2 can reply on behalf of the Orangy News's server. This application of service bundling requires no user-side configuration, so users can automatically benefit from all similar services deployed in the network.

3.3.2 Content Matching

In Figure 2, node G_1 is shown to be carrying a number of contents described with labels C_1, C_2, ..., C_n. For example, these labels could be radio genres available at G_1. In contrast to names, labels are not disseminated in the network because the number of labels is potentially very large. To find labels, end nodes send Content Requests toward the external gateways. In the figure, N_1 is shown sending a Content Request asking for C_2. The request is flooded over I_1 and I_2 to G_1 and G_2; at G_2 the request does not match and is dropped (it never leaves the wireless network). At G_1 the request matches, and a Content Reply is returned along the source route over I_1 to N_1; at this time N_1 selects I_1 as its access router.

The content labels at G_1 deserve a closer look. In particular, where do the labels come from? The labels are stored in a table, as a list of strings. The application running on that node is responsible for populating the table with appropriate labels; in the example of radio genres, the application could be an MP3 music server which inserts into the table the full name of each song. On an end node, an MP3 music player would send a Content Request message by using an application programming interface. Substring matching would be used at G_1 to search the labels, and a Content Reply returned if the matching succeeded. The Content Reply would arrive to the end node, providing it enough information to select the internal gateway leading to G_1. At that time, transfer of music would commence using application-specific protocols.

3.3.3 QoS Routing

In Figure 4 we show an example of Value Request/Reply SARP messages propagating QoS metrics, such as bandwidth and latency. Gateways G_1 and G_2 announce available service by sending Value Reply packets with two attributes, corresponding to the two metrics. For bandwidth, the concave composition rule is used. For latency, the additive composition rule is used [13]. The combined attributes sent by I_1 and I_2 denote the best service these two nodes can provide for each metric. This information allows N_1 to choose the correct internal gateway depending on which metric it wishes to optimize.

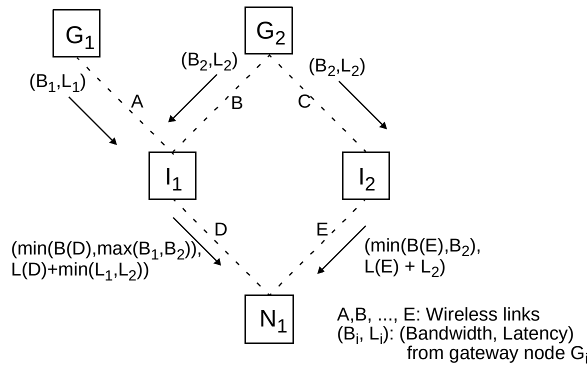

Figure 4. Propagation of QoS information using Value Request packets.

3.4 Implementation Experience

We have implemented a testbed in our lab at Harvard for the open service network described in this paper. In this section, we describe the testbed setup and design choices.

3.4.1 Testbed setup

Our testbed has three layers, each layer consisting of nodes of one of the three types: external gateways, internal gateways, and end nodes. External and internal gateways are Intel-based PCs running NetBSD 1.4.2, while end nodes are IBM Thinkpad laptops running FreeBSD 3.4.

End nodes are connected to internal gateways via Lucent Orinoco IEEE 802.11 wireless Ethernet adapter. Since one layer of wireless links is sufficient for testing our use of wireless interfaces, internal and external nodes are connected via wired Ethernet for simplicity.

3.4.2 Design issues and choices

During our implementation of the testbed, we came across a number of interesting design issues. Here we elaborate on some of these issues and discuss our design choices.

SARP using a well-known control channel

As we mentioned earlier, we use a separate control channel for SARP. When faced with wireless network interfaces such as those that conform to the IEEE 802.11 standard or the emerging Bluetooth standard, we note that the underlying radio hardware supports multiple frequencies or channels of communication. In addition, we note that the hardware design specification generally does not include efficient channel switching mechanisms accessible outside the device firmware. For example, we coarsely measured an Orinoco's channel switching latency to be tens of milliseconds.

We find that in supporting discovery of services that potentially exist on several different channels, using a well-known control channel is significantly faster than alternatives based on channel scanning.

Routing Framework

We use source routing for packets going in both directions. The reason is that maintaining routing tables on gateway and internal nodes is expensive and may lead to inaccurate information. We insert source routes at gateway nodes for each packet destined for end nodes.

To deliver packets from the wireless network to the wired Internet, we also use source routes. When a mobile device makes selection of a path, subsequent packets follow the same path. The internal gateways do not maintain routing tables, so the end nodes specify the path explicitly.

4 Related Work

There are two sets of related work. The first set focuses on mobile devices roaming between channels or link types. The second set consists of service discovery protocols.

Mobile devices can choose services from different channels as is the case with Global System for Mobile Communications (GSM) [11]. In some areas, GSM networks operate on 900MHz and 1800MHz radio bands. Dual-band GSM handsets can select the best network according to signal quality or system capacity. It is convenient to have these handsets if the provider operates in both frequency bands in the area. However, if that is not the case and the provider does not have roaming agreements in the other band, the user is out of luck. Our open service networks also have multiple service providers in the airspace for consumers to select. However, it is feasible for users to switch local service providers dynamically even on packet-granularity.

The BARWAN project [2], where mobile devices can roam seamlessly in heterogeneous networks using multiple link types, is an example of choosing services on different link types. The BARWAN project does not assume that devices can roam between channels, and there are different services on different channels. We provide a solution for services to coexist in different channels while a mobile device can freely roam in between, and complement BARWAN. Moreover, we support service bundling.

Traditionally, a local service can be located using one of the network interfaces that a device has. For example, a device can use DHCP on a network interface to retrieve the assigned address for the network interface. However, when a network interface can have access to different channels for different services, the notion of network interface is not enough to locate all the services. For instance, if a device has an IEEE 802.11 radio interface, and there are four DHCP servers that exist on four different channels, how will the device decide from which DHCP server to retrieve the assigned address?

On the other hand, hardware network address is associated with network interface. A hardware network address can identify a unique network interface of a device. However, when the network interface can access to different channels, the association becomes unclear — which channel the network interface should use?

Service discovery protocols [5] that only operate at network-layer or above will not be sufficient to discover all the local services a network interface has, and IP address is not enough to identify a service in wireless networks. For this reason, we have developed SARP, an service discovery protocol that operates at link-layer, to discover all local services, even they exist in multiple channels.

5 Conclusion

We have proposed an open service network infrastructure for wireless IP networks. Using the service address resolution protocol, customers of network services have choice of providers. The providers can compete in open wireless markets. SARP can be implemented on existing off-the-shelf hardware and network stacks with minimal retrofitting. It is simple enough for mobile appliances.

We have also presented example scenarios that benefit from open services and illustrates the use of SARP. In particular, data content providers can provide better end-to-end services by bundling their services with other service providers. Finally, we have described our implementation experiences with SARP.

Bundling of content and access services, as we have shown in an example, is a powerful mechanism for differentiating services based on the content that is being transported. Content providers can present options for users for better access to the content data through alternate paths in the network or using locale-specific caches. This trend is already occurring in the wired network by cooperation of content providers, cache service providers, and access service providers (ISPs). The bundling concept for wireless networks in this paper appears to be novel.

Using the light-weight SARP, it is easy to build an open service network that enables remote and local service coordination to optimize source-to-destination network service. A virtue of SARP is that modest modification to existing network stacks can convert current mobile devices for open service networks. Namely, network driver modifications that recognize SARP packets make devices compatible.

References

[1] The Bluetooth Special Interest Group, "Specification of the Bluetooth System," December 1999.
[2] E. Brewer, et al. "A network Architecture for Heterogeneous Mobile Computing," IEEE Personal Communications Magazine, October 1998.
[3] B. P. Crow, I. K. Widjaja, G. Jeong, and P. T. Sakai. "IEEE-802.11 Wireless local Area Networks," IEEE Communications Magazine, September 1997, vol. 35, No.9: pages 116–126.
[4] R. Droms, "Dynamic Host Configuration Protocol," RFC 2131, March 1997.
[5] E. Guttman, C. Perkins, J. Veizades and M. Day, "Service Location Protocol, Version 2," IETF RFC 2608, June, 1999.
[6] Home RF, "Technical Summary of the SWAP Specification," http://www.homerf.org/tech/index.html.
[7] Metricom Inc., "Ricochet Technology Overview," http://www.metricom.com/individuals/techover.htm.
[8] C. E. Perkins, "IP Mobility Support," RFC 2002, October 1996.
[9] D. C. Plummer, "An Ethernet Address Resolution Protocol—or—Converting Network Protocol Addresses to 48 bit Ethernet Address for Transmission on Ethernet Hardware," IETF RFC 862, November, 1982.
[10] R. F. Quick, Jr., "An Overview of the Cellular Digital Packet Data (CDPD) System," Proceedings of IEEE Personal Indoor And Mobile Radio Conference, pages 338–343, 1993.
[11] M. Rahnema., "Overview of the GSM System and Protocol Architecture," IEEE Communication Magazine, 31(4):92–100, April 1993.
[12] Team Squid, "Squid Web Proxy Cache," http://www.squid-cache.org/.
[13] Z. Wang and J. Crowcroft, "Quality-of-Service Routing for Supporting Multimedia Applications," IEEE JSAC, 14(7):1288-1234, September 1996.