DESCRIPTION OF A NARROWBAND PCS NETWORK

Allan Angus PE, Weblink Wireless, Inc., Dallas TX

Introduction

ReFLEX protocol based packet data networks are becoming the industry standard for Narrowband PCS (NPCS) wireless data applications due to their economic performance advantages and the fact the protocol is designed for long term migration to higher speeds as demand requires it. At present, there are three independent nationwide NPCS networks in the US using ReFLEX. Taken together, they support about 1 million mobile stations.

This tutorial paper provides an overall description of a modern NPCS network in the context of its logical network reference model and its practical implementation. Topics discussed include the "soft-handoff" simulcast/cellular architecture of NPCS sub-zones; network retry, resend and search procedures; IP-multicast via satellite; Broadcast services; Information-on-Demand; and protocol adaptation for email, Instant Messaging, and binary bearer services.

Network Reference Model

A ReFLEX NPCS network is best described in the context of a reference model. This model may not capture all of the possible detail, but represents a high level view from which some general observations can be made. It is worth noting that the dominant service delivered is the transport of alphanumeric messages of less than 500 characters between "2-way" capable mobile stations (MS). However, NPCS networks also support email to MS and vice versa, binary content to and from units, voice mail reception and indication, and many other traditional "paging" services.

Our reference model includes:

- a mobile device
- the ReFLEX air interface
- a base station with transmitter(s) and receiver(s)
- a sub-zone
- a zone
- a communication back-bone between sites and the MS-O
- an MS-O (output messaging switch)
- an MS-H (home messaging switch)
- a communication back-bone between MS-O and MS-H
- an MS-I (input messaging switch)
- a communication back-bone between MS-H and MS-I
- the PSTN
- the ISDN
- the Internet
- a PSTN terminal
- an ISDN terminal
- a host on the Internet
- communication pathways between PSTN, ISDN, and Internet, and an MS-I

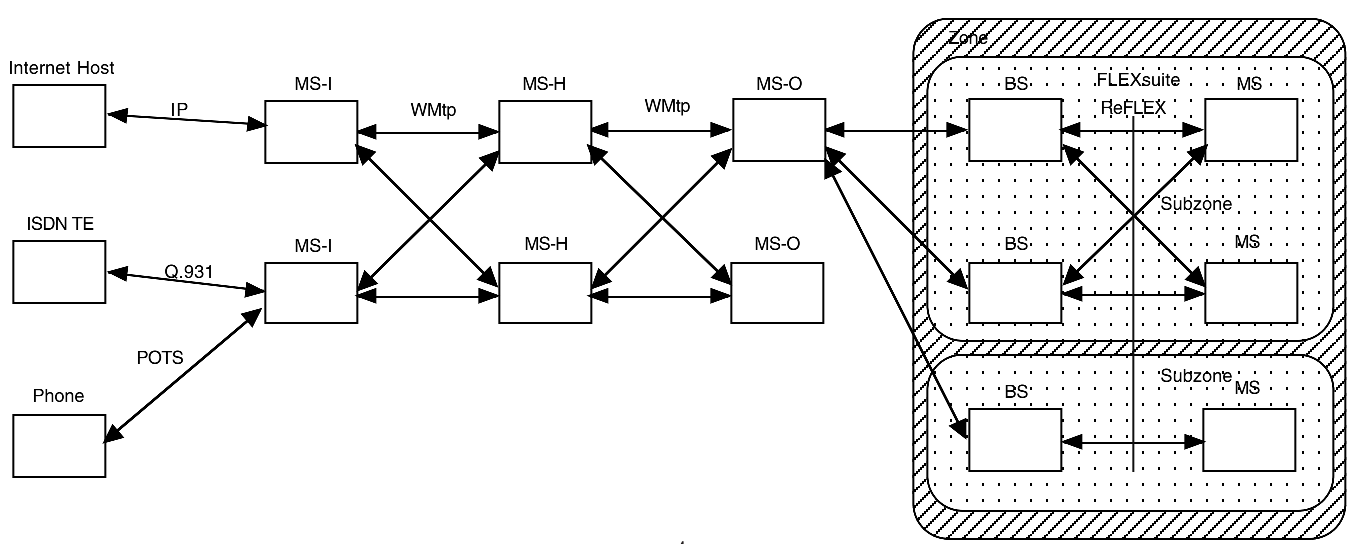
This reference model implies that in addition to using ReFLEX at the air interface, the Wireless Messaging Transfer Protocol (WMtp) is used within the NCPS network.

In the "real world" these elements can be replicated any reasonable number of times. So, one or more mobile devices communicate with fixed site transmitters and receivers via the ReFLEX protocol at the air interface.

Zones & Sub-zones

Fixed site transmitters and receivers are grouped into clusters called sub-zones. One or more sub-zones are grouped into a zone.

NPCS Network Reference Model



A zone, which includes all of the activity of fixed and mobile equipment within it, is managed by an Output Messaging Switch (MS-O). There exists some communication back-bone between all of the fixed sites and the MS-O that manages them. This could be a VSAT [1] (satellite), frame relay, fiber, ATM, LAN, WAN, Internet, or other means.

Mobile stations "register" in zones. Under most conditions, they also "notify" the network about mobility between sub-zones.

WMtp Network Elements

The MS-O is the logical network element that represents the zone at the WMtp application layer. It is functionaly similar to a cross between an MSC-VLR and a Base Station Controller in the cellular reference model. Mobile stations must register in an MS-O in order to be operational, but there is no "home" MS-O for any device. Said another way, every MS-O is a "visited" MS-O. The MS-O maintains a subscriber registration data base that includes every device assumed to be active within the MS-O, and its availability status. A device is assumed to be available if messages to it have succeeded. Generally, unavailable devices are subject to some search procedure.

The MS-O and the other two kinds of messaging switches are WMtp network elements. WMtp is an ASN.1 encoded application layer protocol. It is implemented on top of the Remote Operations Service Element (ROSE) protocol, which is adapted in this case to work over TCP and UDP on IP.

There will be one or more Home Messaging Switches. The MS-H plays a role similar to the MSC-HLR of the cellular reference model. There will be a communication back-bone between the MS-H and MS-O. Any number of means would suffice. The MS-H maintains a subscriber account data base, which includes a queue of inbound and outbound messages, the current logical MS-O in which the subscriber is registered, and the availability state of the subscriber.

There will be one or more Input Messaging Switches (MS-I) that manages traffic into (and out of) the MS-H. The MS-I will be inter-connected with the MS-H by some means. Also, the MS-I will inter-operate with one or more public or private networks, such as the PSTN, ISDN, and Internet. Various terminals or hosts using one or more protocols on these external networks will transport messages intended for ReFLEX mobile devices towards the MS-I as a proxy for, or gateway to, the intended mobile device[2].

Depending upon the design intent of any specific service provider, these network elements may be organized in a variety of ways. For example, one may co-locate all of the MS-Is, MS-Hs, and MS-Os in a nation-wide service offering in one common location. In that case, there is a requirement for back-haul of any telecommunications from all remote markets and the re-distribution of ReFLEX traffic to sites from the central location. To this provider, it is likely that the network out to sites; the "MS-Is", "MS-Hs" and "MS-Os" and their LAN; and the gateways to the "MS-Is" represent distinct aspects to the network, each with its own operational concerns.

Another service provider may choose to place the MS-I, MS-H and MS-O for each market in that location. To support roaming within such a network, the provider will need an inter-connection mechanism among all market elements, perhaps a mesh of frame relay or ATM, or a fiber ring. In any case, to this provider, the natural elements are likely to be the set of "MS-Is", "MS-Hs" , "MS-Os" and sites in each market; the network that inter-connects the market elements together; and the (presumably) common inter-connect of all elements to national or international networks, such as the Internet.

Yet a third service provider may cluster MS-Os and a VSAT net-work hub at one common location, MS-Is and MS-Hs at another switching hub, and back-haul all telecommunications and Internet traffic to the MS-I/H hub. To this provider, the market sites, the VSAT network, the MS-O/satellite hub, the MS-I/H hub, the inter-connect between the hubs, and the nation-wide telecommunications back-haul means are the most important network elements.

The point of this short exposition is just that the network reference model is broad enough to allow a significant variation in physical implementations, and that each such implementation carries with it its own natural grouping of network elements.

Together, ReFLEX and WMtp support a wide range of network implementations, satisfying the needs of almost any service pro-vider, big or small.

ReFLEX networks provide the largest wireless packet data network coverage of any kind in the United States, and arguably, the world. Experience has shown that coverage is perhaps the most important requirement for broad-based deployment of wireless data applica-tions to both businesses and consumers. Wireless service must cover users where they live, work, travel and vacation. Already constructed ReFLEX networks cover more square miles, including suburbs and recreation areas, than any other terrestrial wireless data network of any kind in the USA. Coverage includes many small cities such as Durango, Colorado and include most of the country's fastest growing suburbs important to potential wireless data users.

Interworking with other networks

Messages intended for ReFLEX NPCS mobile stations may be generated on hosts or terminals in any number of networks: the Internet, the PSTN, cellular networks, the ISDN, to name a few. Messages may contain alphanumeric encoded text, voice, DTMF or pulse key strokes, facsimile, or arbitrary binary content. Access protocols include smtp, http, snpp, wctp, dial-up modems, POTS lines, and so on. NPCS carriers have endeavoured to allow for virtually any feasible kind of internetwork operation.

Messages may also traverse between NPCS networks, and NPCS MSs may roam from one carrier's network to another's. Networks of WMtp elements are identified by their Service Provider Code (SPC). ReFLEX network elements (BSs and MSs) are identified by their Service Provider Identifiers (SPIDs) and Zone Identifiers (ZODs). The unique home switch (MS-H) of a device may be globally identified at registration time by mapping of its SPID and Home Index (HIX) to the corresponding WMtp Node Identifier and IP address of the MS-H. Roaming is supported merely be advertising for and allowing registrations from devices associated with other WMtp networks.

ReFLEX Air Interface

The ReFLEX protocol is uniquely engineered for reliable delivery of wireless packet data. ReFLEX has been selected by more carriers as the industry standard platform for light and medium wireless data load applications (e-mail, instant messaging, e-commerce, GPS locationing, information on demand, etc.). Every commercially operating NPCS licensee has committed to supporting the common version 2.7 standard ReFLEX wireless data protocol. At version 2.7, there is no longer a distinction between the earlier ReFLEX50 and the newer ReFLEX25, a single standard subsumes both.

Bit rates & Framing

ReFLEX operates with a 1.875 s frame rate on the forward and reverse channels. While a variety of bit rates are allowed in both directions, it is reasonably typical to use 6400 bit/s in both. The modulation is 4-level FSK, with a peak frequency deviation, f_d, of ±2400 Hz. The modulation index, $h = 2f_dT_s = 1.5$, where T_s is the symbol time.

On the forward channel at 6400 bit/s, each frame contains a synch burst and then 11 blocks of 32 code-words. Each code-word carries 32 bits which are interleaved across a 5 ms interval to provide the usual mitigation to fading. These code-words use a (31,21) BCH code with an additional parity bit for 2 bit error correction. On the reverse channel at 6400 bit/s, each frame carries 77 slots of 154 bits. On a full channel in either direction, there are no idle or guard bits.

There are also *no* requirements for adaptive power control or differential time-alignment of mobile stations.

The reverse channel slots in each frame are separated into a group available for random slotted Aloha access and a group for scheduled allocation. Aloha slots are last in a frame, after the "Aloha boundary", which may indicate some 15-20% of the slots. Aloha access is always on a single slot basis.

Reverse channel messages may be either short Aloha messages or long, scheduled, multi-slot messages. In all cases, messages are interpreted at the MS-I. In the former case, messages are frequently tokenized[3]. In the latter case, messages span multiple slots in a sequence consisting of 3 slot "Data Units" that employ a (31,23) Reed Solomon code capable of up to 20 bit error correction in each slot. This corresponds to a 12% raw BER. DU transmission is subject to an ARQ protocol.

Addressing

Any modern communications network must be designed in the light of the ever-present back-drop of the Internet, and its protocols. This is certainly true of any wireless data network, since most of the "off-net" traffic will flow either from the Internet or from a private network using Internet protocols. The architects of wireless networks invariably are forced to come to grips with two aspects of the Internet Protocol (IP.) Its address plan is verbose and does not accommodate mobility. Second, its protocols are verbose and depend upon its address plan.

The designers of ReFLEX have adopted a far simpler address model, which does not embody any implicit or explicit notion of network or sub-network. This implies the complete ability of the mobile device to roam between serving areas of one or more service providers without modification to the mobile unit's address. That is, the ReFLEX address of a mobile unit is a global and intrinsic attribute, in contrast to the IP address of a networked host, which must be adapted to its current network location.

IP is a "balanced" peer-to-peer protocol. In contrast, ReFLEX supports an "unbalanced", host-to-terminal model. Said another way, ReFLEX devices do not have the inherent capacity to communicate with other peer units; rather, they must always inter-work through the infrastructure of the messaging network, which provides store-and-forward message delivery functions that would be the onus of the sending host in a "balanced" protocol. One clear observation on this point is that any battery-powered mobile unit should not unnecessarily bear the burden of message delivery, as it would under of the Transport Control Protocol (TCP).

In any case, ReFLEX MSs have one "personal" address and zero or more "information service" addresses. Both addresses are 30 bit binary numbers. The personal address is used to uniquely identify the MS to the network, is used to register the device at the MS-O, and is mapped to the subscriber's account at the MS-H. Messages uniquely routed to the MS use the personal address.

However, multiple MSs may share one or more common information service addresses. A single info service message can thus carry the same message content to multiple mobiles in a very efficient way[4].

Tokenized messsage transfer

ReFLEX supports a number of methods to compress message traffic using tokens. Perhaps the most compact of these are based upon the use of pre-defined "canned text" or "multiple choice" replies. In either case, the reply message consists simply of single short Aloha burst that references a previous message, and either a table of pre-defined possible replies or a table of responses that are specific to the message in question.

In a variation on this model, the user may instead transmit a short message burst that references a pre-defined "action" list. A typical use of this model is "Information on Demand" in which the action list consists of specific information retrieval commands. Other uses are feasible as well. Telemetry or telematic units might define specific commands or alarms. And in general, this messaging option allows for a specific signalling alphabet to be defined on a customer by customer basis.

Diversity

One key benefit of ReFLEX NPCS technology is its implementation of wireless diversity at the base station level that permits mobile devices to communicate bi-directionally with multiple base stations simultaneously. This delivers unprecedented always connected service and coverage reliability for mobile devices compared to the single base station, single link implementation of all other packet data networks.

On the forward channel, diversity is achieved by virtue of having all transmitters within a subzone synchronized via GPS, and all transmitting precisely the same signalling stream. On the reverse channel, diversity is achieved by allowing every base station receiver to detect the transmission of each mobile station. As well, each receiver is typically a 2-branch equal-gain combining FFT-base channel bank detector.

Geographically, diversity operates within the scope of a subzone. However, it is feasible to add receivers outside the nominal boundaries of a subzone and control their traffic in such a way as to accept all of their flow or only utilize a subset; for example, Aloha. Similar comments may be made about the location of transmitters.

Within a subzone then, mobility management is essentially "free". As the MS moves within the subzone, it is effectively transferred from one BS to another without any action at either side of the link. Truly a "soft" handover. When a MS moves to another zone or subzone, it notifies the network and moves to the new frequencies after receiving an acknowledgment. At the version 2.7 level of the protocol, scanning for new subzone or zone alternatives is a background task.

Store & Forward Messaging

In an IP network, using TCP as the method for assuring message delivery between two mobile units, intermediate routing systems would provide no inherent "store and forward" functionality. So, if the sender is in good coverage and the receiver is in poor coverage, TCP would place the onus for retries on the sending mobile. Said another way, common physical, link, and network layer protocols (say, Ethernet, IEEE 802, and IP) provide no mechanisms to compensate for the vagaries of a UHF mobile radio interface.

The introduction of "store and forward" agents in the network creates a fundamental imbalance in the communication pathway between communicating end units. Almost everyone will be familiar with this effect in the context of simple mail transfer protocol (smtp) and Post Office Protocol (POP) servers. While it would be perfectly feasible for two Internet hosts to send messages directly using smtp, this is almost never done. Instead, mail servers are established and the end hosts communicate with the mail servers using a post office protocol. This has the distinct advantage of allowing mail to be relayed through various post offices to the one "closest" to the end host without that final host even having to be active on the network. It has the disadvantage of increasing message latency by virtue of the delays through the relay network and in the need for the receiving host to poll its POP server.

In a low latency mobile messaging network, in which communicating end units are in good coverage and active on the network, it would be undesirable to force mobile units to poll servers for messages. This implies a set of network functions that has little, if any, comparison in the Internet.

However, it is absolutely necessary to inter-operate effectively with Internet protocols such as smtp for email, http for web traffic, and so on. Again, the designers of NPCS networks have accounted for this requirement. ReFLEX supports a recursive stack model, which is in some respects even more sophisticated than the linear stack models of TCP/IP and ISO OSI. The protocols of the ReFLEX stack are collectively referred to as "FLEXsuite". For most of the dominant Internet protocols, there exists at least one corresponding FLEXsuite protocol type[5]. FLEXsuite also supports the common set of MIME types as well as the Wireless Applications Protocol (WAP) extensions.

All of this taken together implies a capability to transport arbitrary binary content from an Internet host to an arbitrary application running on a ReFLEX mobile device, and to have that application recognize how to process the data based upon content identifiers. This is the essence of the present success of the Internet: the ability for applications on one host to transport arbitrary content to applications on another host with absolute disregard for the intervening network elements.

Three broad models of interconnect to the Internet are available: a bearer service, a teleservice, and a teleservice supplementary. In the bearer service model, it is assumed that the Internet host has in inherent capability to encode and decode FLEXsuite. In this case, pure binary content is provided to the ReFLEX network, which transports the content untouched to the mobile device[6]. In the teleservice case, an Internet protocol, say smtp, is mapped onto a FLEXsuite protocol, say mailto. In the teleservice supplementary case, an Internet protocol, again say smtp, is mapped onto a FLEXsuite protocol, say Wireless Email (WEM) this time, and in addition, the ReFLEX network provides supplementary services such as mailbox filtering, store-and-forward guarantees, message compression, mail re-routing, attachment stripping, and the like.

These functions of store & forward, and protocol adaptation are distributed across the WMtp network elements, MS-I, MS-H, and MS-O.

NPCS optimized to short bursty data

Since ReFLEX does not start with a connection-oriented voice component, it could be designed for high performance, bursty, short data messaging from the start. The contrasting assumption in NPCS is that the mobile device is either available to the system, or not. If not, as proven by the failure to deliver a message, then the system begins to search for the device. Full details of the search process are beyond the scope of this paper, but suffice it to say that the mobile device is either recovered by the network or autonomously registers again. In either case, any pending messages are subsequently delivered at full speed.

It bears mentioning that in ReFLEX there is only a loose relationship between the base sites that are responsible for the delivery of forward channel messages to the mobile and those that are responsible for receiving reverse channel traffic from it. As an extension, there is no necessary relationship between receiver site locations, antenna patterns, antenna heights, and so on, and those same attributes of transmitter sites. In fact, it is generally desirable to offer a somewhat larger coverage footprint on the reverse channel than on the forward channel, to guarantee the mobile unit's ability to contact the system whenever it sees forward channel coverage.

In short, the RF design of the forward channel and reverse channel can be quite distinct in ReFLEX. Based upon design choices made by early implementers of the protocol, there has come to be an erroneous view that ReFLEX networks must have a much higher density of receive sites than transmit sites on the ground. This is not true, and recent implementations have been constructed at a 1:1 ratio of receive to transmit sites.

ReFLEX networks are capable of supporting increased offered load using an approach similar to cell splitting in cellular systems. This involves dividing a region that would otherwise be a simulcast zone in a traditional messaging network into distinct sub-zones, each with its own forward and reverse channel frequency assignments. This allows a ReFLEX network a similar capacity growth in both directions as a cellular system has, by reducing the effective area of coverage of a serving region. In cellular, this is a cell; in ReFLEX, this is a sub-zone. Likewise, ReFLEX sub-zones do not have to have the same capacity for load handling: one may have only a single forward and reverse channel, while a neighboring sub-zone may have several. As well, variations of the ReFLEX protocol supported in sub-zones do not all have to be alike. Therefore, dense regions of offered load can be served with high density, high capacity implementations of the protocol, while outlying regions of a major metropolis can be served with a capacity commensurate with the offered load.

Another aspect of the ReFLEX architecture is that the aggregate signaling rates on the forward and reverse channels are also highly de-coupled. ReFLEX (at version 2.7 implementation) allows forward channel signaling rates of 1600, 3200, and 6400 bit/s. Binary frequency shift keying (FSK) is used for 1600 bit/s and can be used for 3200 bit/s. Four-level FSK is used for 6400 bit/s and can be used for 3200 bit/s also. On the reverse channel, ReFLEX vs 2.7 allows rates of 800, 1600, 6400 and 9600 bit/s. Four-level FSK is used for all reverse channel transmissions.

This doesn't really describe the aggregate signaling rate in a market, however. In the forward direction, a number of transmitter sites will be synchronized for simulcast operation within a sub-zone. So all these transmitters are occupied with the same forward channel information stream; let us say at 6400 bit/s. Any given user might receive only part of that stream, but all devices would monitor the same shared information resource(s)[7]. On the reverse channel, mobile units that need to access the network compete with one another on a part of the available bandwidth dedicated for this purpose. Since the transmit range of a mobile unit may be limited to only a few fixed sites within a zone, the aggregate "raw" information rate for all sites in the market will be the sum of activity at all of them. The aggregate non-redundant information rate will be less than this because of duplication in information received at multiple sites and representing the same mobile unit's transmissions.

Once a mobile station has succeeded in contacting the network on the Aloha part of the reverse channel, it may, for example, register with the network, or request specific information services, or request bandwidth for a long inbound message. If we follow the last case through, the network will validate the subscriber account (at the corresponding MS-H) and then send a forward channel command to the mobile unit that tells when to transmit the message. This allocation will be made in a part of the reverse channel that is reserved for such scheduled activity, in contrast to the contention access part. Now, this raises several possible scenarios for the allocation of this scheduled bandwidth within a zone or sub-zone. In general, allocations within a sub-zone will potentially be "blocking;" that is, if two mobiles were scheduled at the same time, their messages could interfere with one another destructively at one or more receiving sites. Nonetheless, the network may "over-book" the scheduled reverse channel resource, driving up throughput at the cost of some likelihood of message re-tries. An alternative capacity enhancing approach is to reduce the number of sites in a sub-zone to the point where "over-booking" would add little benefit. Part of the flexibility of ReFLEX is that it allows either approach (or neither.)

Message delivery latency

An analysis of messaging latency in ReFLEX NPCS networks is a daunting task. Given the wide array of message delivery options, there are a corresponding array of contributions to latency. The most significant issue here with regard to latency is the aggregate of delays through the various forwarding agents and firewalls that are involved in Internet email. As mentioned earlier, the response of the ReFLEX industry to this issue has been the development of WCTP, a session-oriented protocol for messaging that employs an Extensible Markup Language (XML) protocol over http. Most Internet-sourced messages with mobile devices as destinations will likely employ smtp up to a gateway maintained by the wireless service provider. Once traffic is received by the service provider's network, the inherent latency to forward the message to the appropriate radio link location will undoubtedly be "small," in any well-maintained network.

A more interesting question is how will the network decide what the most appropriate delivery location will be? With what accuracy? And with what handling in the event of a delivery failure? None of these issues are resolvable in the air interface protocol itself. Rather, they must be addressed in the design of network elements, and in the protocols for their inter-operation. Although somewhat obvious, a message cannot be delivered to a mobile device that is off or out of the serving area or incapable of successful radio communication with the network for any other reason. In fact, the most extreme source of message delay is likely to be due to the link recovery processes in the event of a device being unavailable for communication.

In most ReFLEX NPCS networks, these recovery processes are managed by network elements in the Wireless Message Transport Protocol (WMtp) reference model. The two most important of these are the Output Messaging Switch (MS-O) and the Home Messaging Switch (MS-H). It is the role of the MS-O to manage the mobility of a device within the scope of a serving market, often called a zone. It is the role of the MS-H to manage the mobility of a device across all of the zones of a service provider, including roaming onto the zones of another service provider. This MS-H management role includes both the grant and denial, for any reason, of service rights to a mobile device attempting to operate within an MS-O.

To those familiar with cellular networking, the role of the MS-H is closest to the combination of the Mobile Switching Center (MSC) and its associated Home Location Register (HLR). Likewise, the MS-O is functionally closest to the Visitor Location Register (VLR) of cellular networks. One notable exception is that in the WMtp reference model, all mobile devices are visitors in every possible MS-O. In fact, roaming in a WMtp-based ReFLEX network has little, if anything, to do with location; rather, it has to do with the matching of service provider identifiers stored in the mobile unit, broadcast on the air interface, and advertised by WMtp elements.

A subscriber device is associated with only one MS-H, which is the anchor point for all forward and reverse channel messaging to it and from it. The MS-O acts as intermediary agent between the MS-H and the mobile unit. In the event of a communication failure at the level of an MS-O, the MS-H is informed of the failure. The original MS-O, and perhaps others, begins a search for the device. Recovery of the device by the original MS-O or by another one is signaled to the MS-H, and message delivery recommences. MS-O search procedures are adaptive, and inter-work with a variety of air interface parameters that are designed to ensure continuous and robust link availability.

Cellular and wide-band PCS networks do not support much in the way of search. Rather, the onus is placed upon the mobile device to monitor and update its location relative to the fixed network. This is accomplished through a variety of time-based or location-based registration procedures. A specific example may be instructive. Consider a mobile station that happens to be momentarily out of coverage because, say, the user is in the sixth sub-floor of an underground parking garage. At this point, the network forwards a message, and in spite of any re-tries, it receives no acknowledgement from the mobile. In a voice-based cellular network, this would represent a missed call; and that call attempt is lost. Any search for the mobile is pointless. From the point of view of the cellular mobile, it may have lost contact with the network for a period, but it has no recognition of any failed message event. It is within the same serving region when it regains contact as it was when it lost contact. It will not register with the network on that count. The clock will tick inside the mobile until its registration timer elapses, and then it will re-register. At this point, any pending traffic will be forwarded to it. Since the registration timers impact the behavior of all mobiles within the network, a reduction in the time to deliver any pending traffic to the mobile can only be achieved at the cost of increasing the "polling rate" of time-based registrations from all mobiles.

In contrast, in a ReFLEX network, there is a system parameter called the "incommunicado delay time." If a mobile unit loses contact with the network for longer than this time, it is forced to register. In the messaging scenario just described, the mobile is either out of contact for a period less than or greater than the incommunicado delay. If it is less than this time, the system begins the search process, recovers the device when it is back in coverage; and the pending message is immediately sent. If the loss of contact is greater than the incommunicado delay, then the mobile is forced to register, contact is established; and again, the pending message is immediately sent. Only devices that have been proven to be "unavailable" to the network by virtue of a failed messaging event or that have detected lost contact are subject to this search and registration process. In summary, ReFLEX makes the process of short message delivery event-driven instead of polled.

These observations are also true, for example, of many packet data networks. Examples include those based on the Mobitex model. Mobitex supports message oriented applications using a "connectionless" packet model. In such a scheme, (completely consistent with an adherence to the OSI protocol stack) any transport layer protocols that might provide functions to compensate for radio coverage issues are the responsibility of the communicating hosts. The layer 2 and 3 protocols at the Mobitex radio interface will provide 4 retries for any layer 3 protocol data unit. After this number of retries, failure of the layer 3 protocol data unit (often called a "packet" or "datagram") is reported.

In a sense, this forces responsibility for mobility management upon the sending host. Since the sending host has no access whatsoever to information concerning the radio link to the receiving mobile device, its only mechanism is to re-send the data. Users of such networks may enjoy very low latency messaging when the destination mobile unit is in good coverage. On the other hand, designers of fixed-host to mobile applications may learn more than they ever wanted or needed to know about mobility management. Finally, mobile-to-mobile communications places an untoward requirement for communication management on the sending mobile.

This architectural approach may be highly attractive to those application designers who want to have control over the exact details of the transport, session, presentation, and application layer protocols. On the other hand, most application developers and end-users will simply not have the desire or expertise to manage these issues.

MS battery conservation

In spite of the complexity of ReFLEX networks associated with store-and-forward and search functions, it is typical for ReFLEX devices to operate in coverage with normal use continuously for several weeks on a single AA battery charge. In some ways, they are truly "always on." In part, this is due to the efficiency of the ReFLEX air interface; that is, the low consumption of energy per bit to signal the network. It is also due to the inherent capabilities for battery conservation and "sleep cycles" embedded in ReFLEX.

The ReFLEX air interface is synchronous, and it supports "framed" forward and reverse channels. Frames are of a variety of types: control, system configuration information (SCI), information services (IS), and targeted. Mobile devices must receive each of the first three types from time to time, in order to remain in contact. ReFLEX supports a model in which the control, SCI, and IS frames that are necessary for particular devices can be transmitted on a cyclic schedule. For example, considering device address modulo 8 may identify 8 classes of device. Then all members of each device class wake up in sequence to receive control, SCI, or IS frames pertinent to it. Thus, each device needs to attend to the air interface, in this case, only 1/8 of the time.

This concept is called "collapse" in ReFLEX. The "collapse" of a frame sequence is a small number that is interpreted as a power of 2 to result in the number of distinct device families attending to that type of frame. For example, a collapse of 3 corresponds to the 1 in 8 case given in the previous paragraph. Control, SCI, and IS frames can all have different collapse values. Control frames include a variety of commands to specific devices. These all begin with the unique ReFLEX address of the device. Hence, it is possible for devices to awake on their control collapse schedule, test for the presence of their address, and not finding it, shut down their receivers.

If the device does sense its address, then it is further directed to the location of the information in the command being transmitted to it. One kind of command (or "vector" in "ReFLEX speak") is a further direction to the content of a message within a targeted frame. By increasing the control collapse, it is possible to decrease the amount of time that any device is active on the air interface. This trades battery life for latency. A ReFLEX frame is 1.875 seconds in duration, so 8 frames take 15 seconds. A collapse of 3 represents a significant benefit to the customer in terms of battery life, and yet a modest increase in message latency. Further increases in collapse can yield corresponding improvements in battery conservation.

It is worth noting that latency on the order of fractions of a second at the air interface demands constant electrical activity in the mobile unit. The design of ReFLEX allows the operator of a network to provide improvements in battery conservation to subscribers while trading that gain for only modest increases in message delivery time.

FLEXsuite

Almost all reverse channel messages are encoded in FLEXsuite, or more specifically, in either the "pageto" or "mailto" protocol types carried by the Uniform Addressing and Routing (UAR) FLEXsuite protocol. Generally, this FLEXsuite content is parsed in the MS-I network element.

Nonetheless, it is interesting to explore some of the results that occur from the support of the re-location of the FLEXsuite encoding point, or Internet proxy, from the service provider's network into a corporate or private environment. First, the Internet identity of the mobile unit is associated with the presence of the corporate or private network; and this remains "fixed" in a way that is transparent to mobility of the user, without having to deploy "Mobile IP". Hence, Internet protocol proxies for email, web browsing, and the like, will operate as if the mobile were an integral part of the private domain, with whatever access to private network resources would be present to any fixed host within the private network. With FLEXsuite's capability to define arbitrary new protocol identifiers in a manner similar to the use of sockets on IP, this allows for an equally rich environment for the development of special-purpose mobile applications.

Examples abound: Mobile sales staff may want to inspect an inventory data-base before booking orders. Mobile technical staff may want to check a trouble-ticket data-base to prioritize contacts on their daily routes. Mobile professionals may want to re-synchronize their personal calendars with a corporate or home-office calendar application. If the corporation wished to, it could provide whatever higher levels of authentication and privacy that it wished in order to support applications such as these.

Location services

NPCS providers are beginning to roll out "SnapTrak" GPS-based location services. This constitutes an interesting application of ReFLEX information service functionality. In the NPCS implementation, the GPS satellite ephemeris and almanac information is transmitted on the forward channel at regular intervals to specific information service addresses. Participating MSs are configured to decode this information, which is unique to each serving area. Since this information stream is usually the hardest part of synchronizing a GPS receiver to the satellite network, NPCS provides a significant "boot strap" to GPS-based MSs.

Also, the building penetration characteristics and battery conservation of NPCS GPS-based MSs are as good as any other ReFLEX MSs. Given access to ephemeris and almanic information by means of ReFLEX information services, the MS then detects the pseudo-range data from identified satellites. This information is then transmitted to a network server where location calculations are actually done. Position information can then be relayed back to the MS as a message or to a network-based client in need of the MS's location.

Endnotes

[1] Very Small Aperture Terminals: a kind of two-way satellite system.

[2] The addressing model for the mobile unit must be compatible with the addressing plan of the source network.

[3] Tokens can be substituted for specified text strings in order to compress over-the-air transmissions of commonly-occurring phrases.

[4] ReFLEX allows transmission to "information service addresses" which can be replicated in virtually every subscriber device. Thus, a single message transmission in each simulcast region can hit every device within coverage.

[5] For smtp email, there are two: "mailto" and "WEM" (Wireless Email). Mailto supports a simple interface. WEM supports full RFC822 headers and corresponding functionality.

[6] At present, the ReFLEX industry is moving to adopte common protocol for the dvery of binary content; namely, the Wireless Communications Transfer Protocol, WCTP.

[7] A ReFLEX sub-zone can support several outbound channels. However, once a device is "camped" on a channel, it remains there until some event occurs that would cause it to drop off the air interface or chose a better channel.