Redefining Mobile Applications: Mobility Services and the Unified Network

A Farpoint Group White Paper

Document FPG 2008-199.1
May 2008

Farpoint Group

Farpoint Group has for some time used the analogy of a circulatory system to describe modern enterprise networks. But while certainly apt, such a comparison is at best incomplete today – the issue is no longer simply the circulation of IP-based traffic, but rather, *and fundamentally*, the support of a broad range of applications - which is, after all, the reason the network (and IT itself) exist in the first place. As we will discuss in this White Paper, a clear need to move more intelligence into the network has materialized over the past few years, as applications have become both more network-centric and more complex, and now need to run across not just wired infrastructure, but on multiple classes of wireless networks (from WLAN to WWAN) as well. We thus see an obvious requirement for the network to provide a range of services that to date have been implemented in (or very close to) applications directly. This shifting of functionality between layers is not uncommon, and, indeed, has been the norm throughout the entire evolution of IT. A good example can be seen in modern operating systems, which today are a collection of many common services, such as graphics and networking, originally left to the application itself.

Indeed, it is this requirement for support of a broad range of applications, ever-increasing numbers of users, growing demand for throughput (including time-bounded traffic), and continual expansion (in terms of both coverage and capacity) of the network itself that are providing the motivation for a trend just now emerging – the definition of a *mobility services layer* that extends across multiple networks. While it is perhaps a bit early to re-draw the famous seven-layer diagram to include the functionality we'll discuss in this document, suffice to say for the moment that such a re-definition is one of the most important developments in the history of networking, and one with far-reaching ramifications for network planners, application developers, and IT management alike.

The Contemporary Enterprise Network

To date, we've thought of the network as a collection of physical layers (PHYs) universally united by IP at Layer 3. A good example here is the wired Ethernet network which is today often augmented by a wireless LAN. Wireless LANs can, of course, perform all of the functions of the wired LAN, with some potential compromise in throughput, but with corresponding benefits in mobility, convenience, cost (both capital and operational), and productivity.

Modern enterprise networks are often built from many additional PHYs beyond these – such elements as RFID, Bluetooth and other wireless personal-area networks (WPANs), and cellular technologies grouped under the general heading of wireless wide-area networks (WWANs), but including such diverse technologies as EDGE, EvDO, HSPA, and, soon, mobile WiMAX and LTE. The fact that these WWANs are under the control of carriers can make their inclusion in an enterprise network solution all the more complex. Regardless, modern network-centric IT solutions can and will, then, have components of personal, private, and public services as core elements.

We thus need to think of (and build and operate) network solutions built from a variety of wired and wireless components as a single *unified* structure. There is an essential interdependence between the two, as wireless almost always requires the services of wire for interconnect and backhaul. And it thus behooves us to think in terms of a common *platform* for the provisioning of services across any network, especially in terms of a common management/ operational-support environment, *and a common platform for applications* − what we are defining here as the mobility services layer. This extends the concept of unified networks even further, potentially far beyond what we have thought of as the LAN, but all the while providing a uniform LAN-like environment for the diverse set of applications that are IT today.

But while almost all modern network solutions are based on IP, applications are often written with certain expectations as to the behavior of the network that supports the application during execution. Variability in network behavior, however, is the rule here, not the exception. Among the core challenges are:

- *Throughput* − Traffic performance will vary based on the fundamental capabilities and limitations of a given network itself, along with instantaneous loading and traffic-service demands, and the ability of the network to support, via appropriate protocols and sufficient headroom, time-bounded traffic. RF interference can play a role in the behavior of wireless LANs and other services operating in unlicensed bands, but, even in licensed spectrum, radio behavior can and will be highly variable based on the range between the endpoints, antenna type and orientation, and the presence of obstructions in the path between transmitter and receiver, among other factors and radio artifacts.

- *Capacity* − Throughput alone is an insufficient consideration in successful application deployment. As most networks will be required to support multiple users and multiple applications simultaneously, it's critical that network planners, in the case of both private and carrier-based networks, provision sufficient capacity to support successful application execution, and that services be defined to manage this variable as well.

- *Availability* − Not all networks will be present in all locations. As we noted earlier, it is not an unusual requirement that modern applications run across a mix of enterprise and public networks, as well as in some cases residential networks. What's more, the exact mix of networks cannot always be determined in advance, thus requiring that applications be written with this need for flexibility in mind. Again, while essentially all networks today will provide IP-based transport, otherwise widely-varying behavior must be expected.

- *Cost* − And finally, the costs associated with using different networks can also vary dramatically. This introduces the requirement for least-cost routing in terms of financial cost – but such could be much too complex to allow reasonable implementation in applications directly.

All of the above combine to provide the final motivation we need to define the mobility services layer. Following the example of extending operating system functionality over time that we noted above, such is now emerging in networking – extending the functionality of the network allows services to be *implemented as part of the network*, and not the application. This strategy will simplify application design, broaden the availability of new applications, lower costs, and simplify the management of networks in the bargain.

Re-Drawing the Network Map

To date, there has been no consistent way to access network services, or any form of commonality in network services, across all networks and applications. The first step in this direction, initially popular during the mid-1990s, was the use of *mobile middleware* technology, which at one point was available (in proprietary implementations, of course) from more than 200 software suppliers. Mobile middleware attempted to isolate network behavior from applications, and allowed dynamic binding of networks to applications at runtime. Protocol differences (few mobile networks supported IP at the time) were made transparent, and application development was at least minimally simplified. The IT world, however, was left highly fragmented by this approach. Software development was still expensive, and IT solutions could still be easily described as silos of isolation, in terms of network capabilities, network and operational management, and the applications themselves. There was little real commonality across networks, and solutions remained complex, expensive, proprietary, and minimally extensible at best. And some functionality, like location and tracking, was not thought of as a network service at all, necessitating more custom software development, and more proprietary APIs.

All of this changes with the introduction of an appliance-based mobility services layer. The first product in this space, the recently-introduced Cisco 3300 Series Mobility Services Engine (MSE) and related software products, moves services previously implemented in applications into a single hardware-based network resource. Note we are not talking in this case about a new router or switch; indeed, core network functionality is enhanced by the MSE, but basic network operations are in no way modified or otherwise affected by its introduction. A diagram of the overall solution is presented in Figure 1; technical details are presented in the sidebar below.

For the first time, we have a high degree of intelligence applied to services delivery, and with this intelligence now available as *network* functions. The benefits of such a strategy are many:

- *Centralization of functionality* − First and foremost, a service-based strategy centralizes common network services in a single engine, and allows applications to access this functionality in a defined, consistent, and uniform fashion across otherwise disparate networks. Services are shared among applications, simplifying application implementation and allowing the capabilities of applications to be easily extended as required. A single interface point and API makes this possible.

- *Transparency* - Applications can get access to network intelligence provisioned from a central point without interfering with the operation of the production network itself. While it might be tempting to add application services to other already-present network elements, such makes little practical sense. For example, a router, switch, or even wireless-LAN controller, depending upon implementation, might indeed have sufficient processing power to implement common services for applications, but we believe any inclinations in this direction should be immediately discarded. While the nature of the upper layers in the OSI model are modified by the introduction of application services (think Layer 6.5 or 7.1 or some such), lower-layer functionality is optimized for data *transport* and this functionality should in no way be disturbed or otherwise potentially compromised by the addition of new network services functionality. A services engine should be optimized for the delivery of services to applications and should, similarly, play no role in lower-level network functionality.

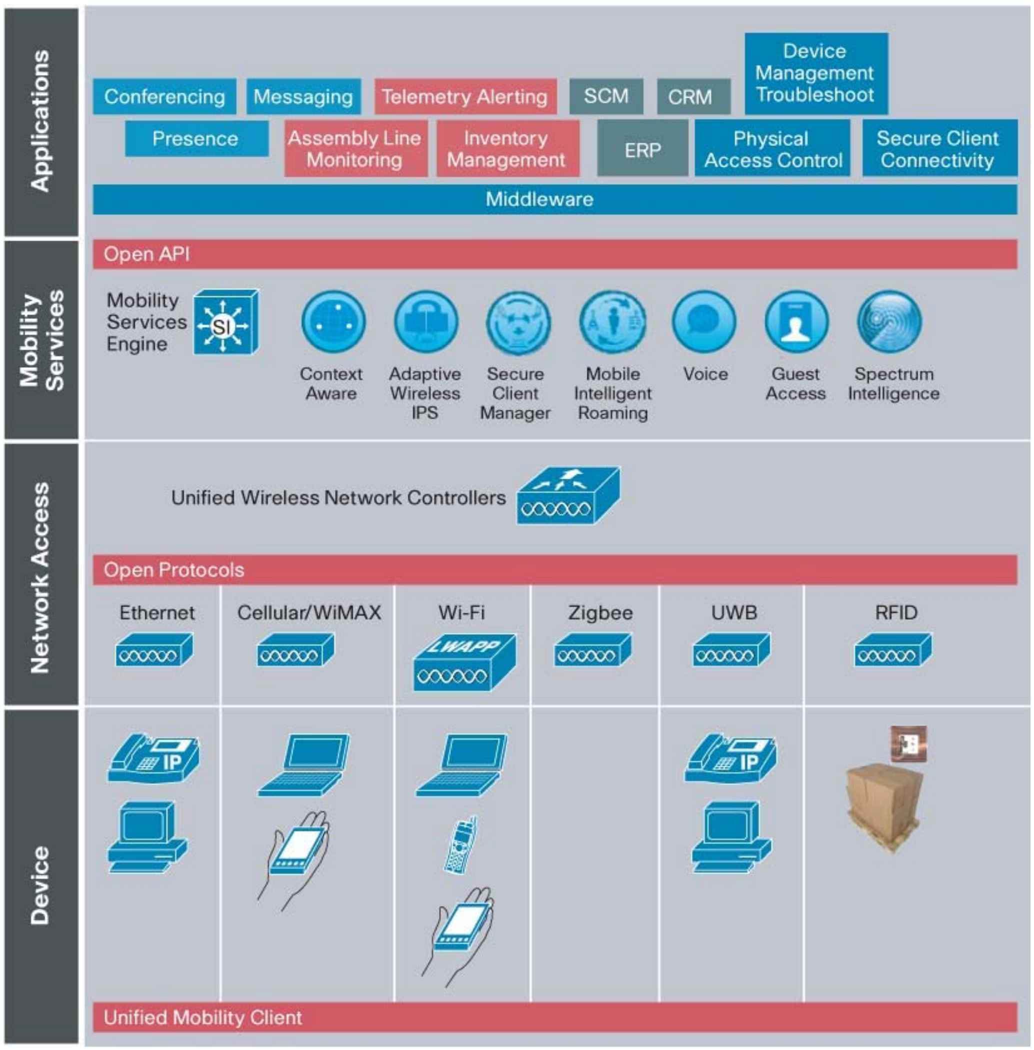

Figure 1 - A block diagram of the services model established by Cisco's MSE. Note how services at the top are mapped through an open, extensible API across multiple disparate networks and mobile devices. *Source:* Cisco Systems

- *Scalability* − While we have noted that network services are logically centralized in the model introduced by Cisco, in reality the capabilities provisioned can scale over time (by adding more engines) to cover larger numbers of users and applications with high demands for services across potentially very large networks. This means that applications can continue to perform at very high efficiency even as the inevitable network growth that we noted above continues, and with no loss of responsiveness or compromise in other measures of application performance.

- *Openness and extensibility* − The services model must be implemented as a platform with an API that allows third parties to extract and integrate intelligence from the network without compromising the performance and operations of the control network itself. This means that the services engine can rely upon the transport capabilities already in place, and can serve as a single point where additional (perhaps undefined or unknown today) functionality can be added and made available to applications in the simplest possible fashion. Again, the services engine is independent of core network operations, and no changes to key network elements are required to adopt the services model.

And, while ultimately extensible to meet current and future applications demands, Cisco has defined an initial set of services currently available in their implementation, as follows:

- *Context-aware mobility* − This includes the export of environmental variables (such as geographic location, or even such quantities as temperature or device condition) to make decisions about services provisioned in a given location or other context – a form of the location-based/aware/enabled services rapidly becoming essential to many mobile applications.

- *Adaptive Wireless Intrusion Prevention* − This includes the ability to monitor and mitigate threats that may emerge over the wireless network. Additionally, this service provides correlation of threats that may emerge over other wired and wireless networks.

- *Mobile Intelligent Roaming* − This is a form of mobile/mobile convergence, enabling the dynamic handoff of connections across private and public networks.

The Cisco 3300 Series Mobility Services Engine

As we noted elsewhere in this document, the Cisco 3300 Series Mobility Services Engine (see Figure 2) is the first product of its type, designed to provide a single point of interface to an extensible set of network services. Architecturally, the product is perhaps best thought of as a standalone framework engine that allows the integration of a broad array of functionality into a single entity on the network, functioning independently of the underlying physical networks that provide transport and lower-level network services. The 3300 Series thus provides a degree of isolation and commonality that would otherwise need to be implemented as application software – clearly a complex and costly alternative.

Figure 2 - *Source:* Cisco Systems

The 3300 Series implements a classic service-oriented architecture that uses a common SOAP/XML-based API to provide interfaces to an extensible set of services. These services can also communicate among themselves, via what Cisco calls a "Services Communication Bus", and all can be centrally managed, authenticated, authorized, and monitored as well. The hardware itself is quite powerful, featuring dual quad-core Xeon processors, eight GB of memory, dual hot-swappable power supplies, and dual hot-swappable disk drives. Dual gigabit-Ethernet ports are also included, and the entire device fits in a single rack space. Redundant configurations for fault-tolerance are also supported.

- *Secure Client Manager* − This functionality includes provisioning, security, management, and service personalization across a broad range of mobile devices to help ease the task of device management.

And, as we previously noted, we believe that all successful implementations of this concept will be fundamentally designed to facilitate third-party extensibility. This means that we can expect to see a very rich set of network services defined over the next few years as developers discover the power and flexibility of the mobility services layer approach.

Key Benefits and Application Examples

Perhaps the most obvious benefits of the mobility services approach is in savings on development and operational expenses. Granted, there is a small amount of capital expense associated with adding a new appliance to the network, but the cost savings on application development should allow for a very quick payback. Similarly, services functionality is concentrated in a single network element, making management, troubleshooting, and upgrades (via the open API and extensibility capabilities) much easier than has been the case in the past. All services are available to all authorized applications, reducing complexity in the applications themselves and enabling them to operate uniformly and seamlessly across all supported network PHYs. For the first time, software developers can build network-independent applications with consideration simply for the services offered by the network, and not the specifics of wired and wireless networks separately and independently.

It would be hard to imagine an application environment where the centralized-services approach would not be appropriate. Indeed, we believe that *all* mobile applications, including those in healthcare, education, retail, manufacturing, financial services, field service, and general enterprise applications and access will benefit from this strategy. Just by way of a couple of application examples, consider the following:

- *Collaboration* − A common service delivery platform allows network services to integrate with one another. Consider the integration of two services, location and presence. The combination of these two services provides a richer collaborative experience; for example, a doctor entering an operating room can automatically have any phone calls routed directly to voicemail, so as to not be disturbed. Students with common interests can receive real-time updates as to their whereabouts and meet (physically or virtually) to collaborate in an *ad hoc* fashion. We believe that presence and location will greatly enhance the utility of collaboration applications in the future.

- *Asset tracking and RFID* − RFID is used extensively in manufacturing, logistics, transportation, retail, and many other industries. Location and asset tracking is possible with wireless-LANs as well as cell phones and other mobile handsets. Centralizing these services provides a high degree of isolation between a given application and the peculiarities of any given location and tracking solution, and also provides support for features not always enabled without extensive custom coding, such as presence, environmental conditions, and the physical state of a given device – or user. Application features supporting exception conditions can be easily implemented, with the resulting application automatically capable of working with many different mobile devices and sensors.

- *General enterprise applications* − Enterprise solutions are increasingly being constructed so as to operate across several different networks with diverse capabilities. Roaming across these network boundaries has traditionally been difficult to implement, requiring extensive programming and long debug cycles. Centralized services can provide a previously-unavailable degree of transparency here, simplifying development and enhancing both reliability and functionality for the end user.

And, of course, all of these scenarios will benefit from centralized mobile device management, centralized security configuration, monitoring, and management, and extensibility making the addition of new applications and new devices much easier than has been the case to date.

Conclusions and Future Directions

It was perhaps only a matter of time, and regardless a very natural evolution, for the fundamental concepts embodied in mobile middleware to move into the network in the form of an appliance. But the introduction of the Cisco Mobility Services Engine and its associated software components represents more than a simple evolution. In fact, we expect this approach to become the *dominant* technique for implementing the interface between applications and the many services that networks today must support. We are especially fond of the inclusion of an open API and the fundamental extensibility of the offering – meaning that much of the expense fundamental to building and maintaining mobile applications is greatly reduced.

This implies, of course, than we can expect to see a very rich ecosystem of capabilities grow around this initial product and future products of this type. And this bodes very well for the future of mobility itself as the primary or default access for an increasingly large number of users, further narrowing the behavioral and performance differences between wired (which, lest we forget, are also supported here as well) and wireless networks, and thus significantly advancing mobility and the promise it holds. As we have previously noted, future network solutions must be conceived with unified networking mind. The mobility service layer builds on this trend, augmenting capabilities, lowering costs, and getting solutions into the hands of those who need them faster than ever before.

Farpoint Group
Ashland MA 01721
508-881-6467
www.farpointgroup.com
info@farpointgroup.com

The information and analysis contained in this document are based upon publicly-available information sources and are believed to be correct as of the date of publication. Farpoint Group assumes no liability for any inaccuracies which may be present herein. Revisions to this document may be issued, without notice, from time to time.