Why Cell Phones Will Dominate the Future Internet

S. Keshav
Canada Research Chair in Tetherless Computing
School of Computer Science, University of Waterloo
200 University Ave. W, Waterloo, ON, Canada N2L 3G1
keshav@cs.uwaterloo.ca

Categories and Subject Descriptors

C,2,1 [**Network Architecture and Design**]: Network Communications

General Terms

Management, Performance, Design, Economics, Reliability

Keywords

Internet architecture, Cellular communications

1. INTRODUCTION

Quick! What does the Internet look like? Chances are pretty good that you're thinking of desktop computers on a LAN connected to servers, middleboxes, and other LAN-based desktops by routers and local and long-distance point-to-point links. Implicit in this model is the assumption that all users access the Internet from a desktop computer. Although substantially true five years ago, I will argue that the picture is already wrong today and that in the near future most people will use cell phones rather than desktops to access the Internet. A good model for the future Internet would therefore be a very large number of cell-phone-like, mobile, wireless, lightweight, end-systems, connected using CDMA and GPRS (and potentially, IEEE802.11 as well), to well-managed cell-phone provider networks, that provide access to a highly-connected bandwidth-rich wired core and associated centralized servers. Desktops will not disappear, of course, but they will play an increasingly smaller role in the typical way an Internet user accesses the network; a large fraction of future Internet users may never use what we would think of as a desktop today.

As a surprising consequence, the holy grails of cost-effective, always-available network access, multimedia networking, and end-to-end quality of service may finally be achieved. A cell-phone dominated Internet may thus resolve the decades-old tension between the telephony and *laissez faire* packet-networking views of the world.

Section 2 outlines cell phone technology and demonstrates why today's mobiles are as much IP end points as desktop machines. In terms of Internet access, for simplicity, I'll only consider three alternatives to cell phones: wired desktops, wired or wireless laptops, and wireless Personal Digital Assistants (PDAs), where the wireless laptops and PDAs use 802.11 (WiFi) access points. Section 3 describes non-technical advantages of cell phones over these alternatives, and Section 4 presents their technical advantages. Section 5 discusses cell phone evolution trends, and presents some wild speculations on what this means for future Internet architecture.

2. OVERVIEW

A cell phone is essentially a battery-powered microprocessor with one or more wireless transmitters and receivers optimized for voice I/O. Even a bare-bones model provides a keyboard, an LCD screen, and a general-purpose computing platform, typically supporting Java2 Mobile Edition (J2ME) or .NET Compact APIs. More sophisticated models provide a camera, 1MB-5GB of local storage, a full-color screen, multiple wireless interfaces, and even a QWERTY keypad.

Importantly, a cell phone cannot be used without a globally unique, per-user, hard-to-forge identifier, called the International Mobile Subscriber Identifier or IMSI[1]. IMSIs are allocated by cell phone providers and allow them to track and bill for usage. A cell phone provider maintains a comprehensive database, called the Home Location Register (HLR), that keeps track of the current location of each IMSI, its usage, and associated subscriber information, such as a credit card number, or prepaid usage authorization. HLRs make it possible for cell phone providers to do very fine-grained billing.

Nearly all cell phones today provide voice and data I/O over either CDMA or GPRS networks[2]. In both networks, data access is over a channelized medium, where separate wireless frequency channels (or, equivalently, timeslots) are dedicated to data communication and signaling [BVE 99]. For example, in GPRS, cell phones contend for access to the data channel using Slotted-ALOHA on one of the control channels (called the PRACH channel). In response, the base station uses the packet grant channel (PAGCH) to explicitly grant it one or more time slots on the data channel (PDTCH). The cell phone uses these slots to send an IP packet, encapsulated in a convergence layer protocol (SNDCP), to the base station, which forwards it to a local packet router (the SGSN), which tunnels it to an IP *gateway* (the GGSN), where it enters the Internet. Symmetrically, data meant for a cell phone is routed through the Internet to the gateway (GGSN), which tunnels it to the SGSN, and thence to its base station. The base station uses a data channel to deliver the packet to the cell phone.

A cell phone that wants to send and receive IP packets starts by requesting a *packet data protocol context* from the cell phone provider. This context assigns it a packet data protocol (IPv4 or v6), a corresponding IP address, a quality of service specification, and, optionally, a DNS name. This process allows the cell phone

[1] Strictly speaking, a cell phone used for an emergency call (such as to 911 in North America) does not need an IMSI.

[2] This paper focuses on packetized data I/O, which includes Voice over IP, and ignores circuit-switched voice.

network (specifically, the GGSN) to associate the cell phone's unique ID (i.e. IMSI) with its IP address. Thereafter, standard cell phone locationing ensures that any packets sent to the cell phone's IP address can be routed to it no matter where it roams. Note that after receiving a packet data context, to all intents and purposes, the cell phone is on the Internet and can exchange IP packets with any other Internet host, making it a *bona fide* Internet host. In the next generation infrastructure, called IP Multimedia Subsystem (IMS), all mobiles will receive an IPv6 address, and the entire backbone will be IPv6-enabled.

In addition to basic data transport, cell phone networks provide two types of messaging. The Short Message Service (SMS) allows up to 160-character messages to be sent to a cell phone with little delay. Over 50 billion SMS messages were sent in 2004. Next-generation Multimedia Message Service (MMS) messages are compatible with SMTP and are unlimited length, thereby transforming a cell phone into standard Internet email endpoint.

3. NON-TECHNICAL ADVANTAGES

Having briefly surveyed cell phone technology, let us consider the non-technical advantages of cell phones over desktops, laptops, and PDAs.

3.1 Sheer numbers

By the end of 2003, there were 1.4 billion cell phones, serving about 25% of the world's population [ITU 05]. In comparison, there were only 607 million PCs (which includes both desktops and laptops), and negligibly small number of PDAs. In other words, in 2003 about a billion people had cell phones but not desktops/laptops or PDAs. Cell phones continue to maintain this lead because of a rapid rise in subscriber numbers in China, India, and Russia. For instance, in 2004, China reported 310 million users, about 25% of its total population, and India saw an increase of 11 million, or 25%, and reached a total of 44.5 million subscribers. In Russia, mobile phone subscriber numbers jumped 65% from 36.5 million in 2003 to 60 million by September 2004 [IT Facts 05].

Thus, simply in terms of numbers, cell phones are *already* the dominant platform for Internet access. Unfortunately, very few people use cell phones for data access today. However, this is likely to change. Here's why:

The large and a rapidly growing market makes cell phones attractive to handset vendors, operating system providers, software houses, and service providers, leading to fast-paced innovation. This is already apparent by considering the range of handset choices today – ranging from the Samsung SCH-V770 that has a 7-Megapixel camera, to Sanyo's HDR-B5GM that includes a 1-inch 5GB drive. Given that voice revenue has wafer-thin margins, there is a huge financial incentive to roll out innovative data services, such as those pioneered by NTT DoCoMo, rapidly [Imode 05]. For example, DoCoMo subscribers use data services at up to 384 Kbps to access video clips, upload camera images, and get street maps, stock quotes, restaurant menus, weather, and 'yellow pages' information. This sort of innovation will continue to transform cell phones from voice-only devices to integrated voice-data devices, to eventually becoming data-dominant devices.

3.2 Cost

A second effect of a large user population is that costs shrink dramatically, both for handsets and for service. Handsets are already manufactured in the hundreds of millions every year, which makes it possible to dedicate expensive chip fabs and ASIC developers to the task. These up-front capital investments greatly reduce the marginal cost per handset. Similarly, large volumes allow service providers to make healthy profits with small markups, reducing the price of service. Lower handset and service prices in turn increases the total addressable market, leading to positive feedback.

Similar positive feedback effects have led to decreasing prices for desktops and laptops, but PDA volumes are simply too small to ride the technology cost curve, and consequently several manufacturers (such as Sony and Sharp) have withdrawn from the market. Also, in contrast to wireless access prices, a lack of competitive pressure for wired local access has actually increased the service fees for monthly access in the United States! So, in terms of handset and service prices, cell phones have taken the lead, and are likely to maintain it for the near future.

3.3 Marketing model

Cell phone providers usually give away or highly subsidize handsets in an effort to gain market share. The handset cost is recouped over a two- or three-year period as part of a monthly service fee. Essentially, the provider acts like a bank to finance the cost of the handset. This reduction in the cost of the handset makes it very affordable. In contrast, few desktop or PDA vendors offer comparable terms, especially to consumers.

3.4 Well-established providers

Unlike WiFi hotspot providers and most ISPs, cell phone providers are well-capitalized, well-established, and have a large cash flow because they own significant shares of their home markets. Consequently, they are able to adequately provision their networks and, more importantly, enter into long-term settlement contracts with each other. This allows subscribers to seamlessly roam between coverage areas, receiving a single monthly bill. In contrast, one cannot imagine obtaining seamless roaming Internet access from either wired or WiFi service providers today: there are too many ISPs and they rarely trust each other! The ability to roam will greatly reinforce end user preferences to access the Internet using cell phone providers.

3.5 Form factor

Ideally, Internet access, like telephone access, should be high quality, cheap, and available everywhere. Cell phones have overtaken fixed-line phones for voice transport because people are willing to compromise on quality and cost to gain mobile access. One cannot carry the analogy too far: cell phones and fixed-line phones both provide roughly equivalent voice quality, but mobile cell phones will never have the screen size and ease of use of a desktop. So, there will always be a market for fixed, wired, powered desktops. Nevertheless, (a) some users may not need a desktop as we know it today and (b) cell phones are likely to supplant laptops and PDAs as mobile devices.

To begin with, as Moore's law allows increasingly more processing power to be crammed into a chip, cell phone processors will eventually be powerful enough to run common office productivity applications. Imagine that such a cell phone also comes with a dock that takes power in, and provides keyboard, video, and mouse out. This 'desktop' is well within reach by 2010, and may be adequate for most users who do not want to own another computing device. This 'docked' cell phone may be connected to the network on a fixed line, or may provide CDMA/GPRS and WiFi access.

Second, as a mobile computer, laptops are too heavy, too awkward to carry, and do not permit opportunistic data access, such as for reading email while waiting for an elevator. In contrast, cell phones and PDAs have the right form factor. However, PDAs that use WiFi for Internet access cannot benefit from the cost, subsidy and market size benefits that are available to cell phones. For these reasons, it is clear that, in the long term, cell phones, especially multi-NIC cell phones, will replace PDAs; and perhaps 'docked' cell phones may replace some laptops and desktops.

To sum up, we see that cell phones already numerically dominate the number of Internet end points, though data services accessed through cell phones today are relatively scarce. However, the cell phone market is likely to grow due to shrinking hardware and service costs and subsidized handsets. Moreover, with voice margins shrinking, service providers are sure to leverage their existing relationships and huge cash flows to fund innovation and provide data services and seamless worldwide roaming. These factors will make data services on cell phones rapidly gain popularity, making cell phones the dominant Internet access technology. 'Docked' cell phones of the future may also supplant laptops and desktops.

4. TECHNICAL ADVANTAGES

As good as this sounds, this still leaves a major question unanswered. Even if cell phones dominate future Internet access, are they the best technical solution to the problem? Are desktops, WiFi-enabled PDAs, or laptops better Internet access devices that are losing out to a technically inferior solution? In this section, I argue that cell phones not only will win out, but actually have several significant technical advantages over these competing solutions.

I will first outline technical advantages of cell phone handsets and wireless access over WiFi-based laptops and PDAs (Sections 4.1 and 4.2). I will then discuss why the cell phone-provider managed portion of the Internet is better than the 'general' Internet (Sections 4.3-4.7).

4.1 Power management

Power management is critical for battery-powered mobile devices. Laptop and PDA vendors tend to sacrifice power for backlit screens and processing speed, leading to short device lifetimes of two to eight hours. In contrast, cell phone vendors have always paid fanatical attention to power management, leading to much longer device lifetimes. Talk times are typically six to eight hours, and standby times are measured in days. For instance, unlike cell phones, laptops and PDAs do not support a standby mode where they can power off most services while still being available for incoming data. Therefore, from this perspective, cell phones are definitely a better technical solution than laptops and PDAs.

4.2 Channelization

IEEE 802.11 is not channelized, so control packets, such as RTS, CTS, and ACK, use the same channel as data packets. This leads to complex arbitration schemes and potentially unfair bandwidth allocation due to hidden and exposed terminals [BDSZ 94]. Cell phones use channelized media, which intrinsically share the wireless medium better and are immune to a variety of hidden terminal problems (See [BTV 04] and the references therein). Moreover, unlike a WiFi-based PDA, a cell phone cannot be blocked from accessing the channel because the data channel is hogged by another cell phone. For these reasons, it appears that a channelized cell phone may better use wireless spectrum than a WiFi-based laptop or PDA. Incidentally, cell phone spectrum is licensed, so it is also immune to interference from cordless phones and microwave ovens that occupy the unlicensed ISM bands.

4.3 Identity and location management

A major problem with the Internet today is that IP addresses are topologically significant. In contrast, a cell phone's identifier (IMSI) identifies its *user*, has no topological significance, and is bound dynamically to its location using a Home Location Register (HLR). As a cell phone moves, its location is always (more or less) known to the HLR. The 3G Partnership Project [AJM 04] defines how this is coordinated with Mobile IP. Essentially, packets destined to a phone are sent to its home network, and then forwarded using Mobile IP to the cell phone's current location, which is obtained from the HLR. No such locationing information is available for standard Internet endpoints: a PDA or laptop using a WiFi hotspot with a NAT'ted DHCP private address is simply unlocatable!

4.4 Quality of service

Packets to and from cell phones are transported (at least partially) over a provider's private IP network. Note that every cell phone is uniquely identified using hardware identifiers and has a billing relationship with the service provider. This makes it both technically and *economically* feasible to provide them quality of service guarantees, especially for multimedia applications and VoIP. Imagine that a cell phone user wants to view a video stream from a video server on its provider's private Internet. Because the source and destination endpoints are known, the path can be pinned down using MPLS, and quality of service guarantees can be provided using either IntServ or DiffServ. All of this makes economic sense because the provider can charge the cell phone user per-byte or per-video using an existing billing relationship. None of these pre-conditions for quality of service provision exist in the general Internet.

Note that GPRS-based cell phone providers are tied together using the private GRX network [GRX 04]. In principle, this allows global cell phone-to-cell phone provisioning of quality of service parameters, with built-in support for billing and settlement. This is an essential pre-condition to providing end-to-end quality of service, which is economically infeasible in the general Internet. and, by extension to laptops, PDAs, and even desktops on it.

4.5 Over-the-air software upgrades

Keeping application versions on endpoints up-to-date and consistent is a significant headache for every enterprise today. Cell phones, by design, do not suffer from this problem. When software on a cell phone has to be updated, it is automatically downloaded to the phone by the cell phone provider, using software such as that provided by Bitfone [Bitfone 05].

4.6 IPv6 support

The IETF has been struggling for about 15 years to migrate the general Internet from IPv4 to IPv6, and it might never happen. In contrast, the next-generation cell-phone provider IP network is specified to be IPv6. This is because cell phones are managed systems that are dynamically allocated IP addresses and whose software can be dynamically updated by a cell phone provider. Cutting over to IPv6 requires installation of the appropriate protocol stack on cell phones, and switching over to a parallel internal infrastructure. Neither the capability to download software to an end system nor the control over the network infrastructure to force a cut over to IPv6 exist in the general Internet.

4.7 Ease of maintenance

Both the wired and the wireless components of a cell phone provider's network are *managed*. This means that the provider is responsible for network provisioning, monitoring, and management. This delegation of responsibilities from the end system to the network makes the overall system stable: links can be provisioned on the basis of a measured traffic matrix, consistent routing tables can be centrally computed and installed on routers, routes can be pinned using MPLS, and link quality can be uniformly measured by a network-wide operations center, much like the telephone network. Because there are only a few hundred cell phone providers worldwide, in contrast to the tens of thousands of ISPs, worldwide coordination and management is feasible. Similarly, providers can roll out system-wide changes or improvements in infrastructure without having to impact the service seen by the end points.

5. IMPLICATIONS

With rapidly increasing coverage, data-enabled cell phones will soon deliver the long-held vision of 'anytime anywhere information access.' Moreover, based on their technical and non-technical advantages, I believe that cell phone handsets will quickly displace PDAs. Over the longer term, dockable cell phones, as described in Section 3.5, may even displace laptops and desktops.

Much like an iPod, such dockable devices with large local storage will allow users to carry all their data with them all the time. In other words, these devices will not only provide data access, but also *data* and *compute* mobility. This will make it desirable to allow updates of the local data store, either from CDMA/GPRS networks, or opportunistically from WiFi/Bluetooth networks, so that periods of disconnection can be hidden from the user.

User demand for higher bandwidths and lower operational costs will inevitably lead to the proliferation of WiFi- and WiMax-enabled cell phones that can switch to WiFi or WiMax when available. To provide good voice and data quality, especially for VoIP, this will require cell phone providers to either build out or partner with WiFi/WiMax providers, and provide the same degree of management on these networks and their backhaul as they provide on their own backbones.

Following this train of logic, it seems clear that the proliferation of cell phones, their use for data access, and the concomitant growth of cell phone-based Internet service providers will lead to an increasingly larger portion of the Internet being managed and provisioned by cell phone providers. As the fraction of users accessing the Internet from cell phones grows, there will be a strong financial incentive for Internet application service providers like Yahoo and Google to establish a presence on this provisioned network. It may well be that over time, most, if not all Internet service providers (VoIP providers and web site hosters included) have links both to the provisioned and the 'best effort' Internet, and eventually the provisioned Internet may make the best-effort Internet vestigial. If this happens, the provisioned Internet would integrate the best ideas of the last fifty years of telecommunications: temporal statistical multiplexing gains through packet switching, traffic management using provisioned MPLS paths, and end-to-end quality of service guarantees using effective scheduling disciplines. This very welcome state of affairs is achievable in the next ten years, and may perhaps be inevitable, given that cell phone users will dominate the future Internet.

6. ACKNOWLEDGMENTS

The ideas in this work owe much to ongoing discussions with my students, Aaditeshwar Seth, Nabeel Ahmed, and Majid Ghaderi at the University of Waterloo. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program. Thanks also to Intel Corp., Sprint ATL, Nortel Networks Institute, and Sun Microsystems for the generous financial support.

7. REFERENCES

[AJM 04] I. Akyildiz, X. Jiang, and S. Mohanty, "A Survey of Mobility Management in Next-Generation All-IP-Based Wireless Systems," *IEEE Wireless Communications,* Vol. 11, No. 4, Aug. 2004, pp16 - 28.

[BDSZ 94] V. Bharghavan, A. Demers, S. Shenker , and L. Zhang. "MACAW: A Media Access Protocol for Wireless LANs," *Proc. ACM SIGCOMM*, September 1994, pp. 212-225.

[Bitfone 05] http://www.bitfone.com/usa/product_mprove_architecture.html

[BTV 04] A. Baiocchi, A. Todini, A. Valletta, "Why a Multichannel Protocol can Boost IEEE 802.11 Performance," *Proc. MSWiM '04,* October 2004.

[BVE 99] C. Bettstetter, H-J. Vogel, and J. Eberspeher, "GSM Phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface," *IEEE Communication Survey*, 1999.

[GRX 04] GSM Association, "Inter-PLMN Backbone Guidelines." Official Publication IR.34, August 2004.

[Imode 05] http://www.nttdocomo.com/corebiz/services/imode

[IT Facts 05] http://www.itfacts.biz/index.php?id=C0_4_1

[ITU 05] http://www.itu.int/ITU-D/ict/statistics/at_glance/