CHAPTER 1

CONTENT NETWORKING IN THE MOBILE INTERNET

SUDHIR DIXIT and TAO WU
Nokia Research Center
Burlington, Massachusetts

1.1 INTRODUCTION

The mobile wireless and Internet are clearly driving the need for more content that is varied, customizable, and available anywhere, anytime, and at low cost. With the penetration of mobile devices reaching 60% or more in many countries, it is only natural that an increasing number of people will access the Internet and invoke services and applications from such tetherless devices. From a business perspective, an interconnected population of over 6 billion worldwide presents vast opportunities at a time where distance is rapidly losing its significance. The universal availability of content and related services, including delivery, distribution, adaptation, management, and charging, will be critical for the success of the mobile Internet.

Human factors studies suggest that a delay of more than 8 seconds in the delivery of the content can easily result in either lost sale or permanent abandonment of the site and/or content by the user. In fixed-access networks, various content delivery techniques are already in use to enhance user experience; however, when the access is through the air interface with mobility, new issues and challenges of performance need to be dealt with. In addition, the mobility provides the dimensions of location and context that can be leveraged to enable enhanced performance and more meaningful services to the user. Although the mobile Internet is in its nascent stage, the explosive growth is yet to come when real-time multimedia applications become pervasive and as the need to access information anywhere anytime grows. Nonetheless, the air interface will continue to be a limited and unreliable resource, and, when combined with the slowdown in the telecom (telecommunications) market and deferred infrastructure upgrades, it is even more important to develop tools and techniques that will enhance the user experience with no or

Content Networking in the Mobile Internet, Edited by Sudhir Dixit and Tao Wu
ISBN 0-471-46618-2

minimum requirements on upgrading network protocols and infrastructure. These solutions will target the manner in which the content is managed (in terms of storage, value-added services, search, distribution, consistency, adaptation, digital rights management, and charging) and delivered to enhance performance (latency, reliability, efficiency) and usability. In a nutshell, the development of future mobile networks will be driven, in large part, by content and Web-based services, and user experience. These are essentially the topics that are covered in this book.

1.2 CONTENT NETWORKING IN THE MOBILE INTERNET

Before answering the question of what is content networking in the mobile Internet, we must first define the keywords: content, content networking, and the mobile Internet. *Content* typically refers to any sharable object in any of its manifestations that a user is interested in. *Networking* as a general term is well understood, and refers to the Open Systems Interconnection (OSI) or the transmission control protocol/Internet protocol (TCP/IP) layered models. Layers 1–3 (physical → data data link → network) provide the communication infrastructure of conduits and routing methodologies to deliver packets from one end to the other end regardless of the contents of those packets. *Content networking* generally refers to the tools and techniques that operate at the layer above the networking layer where the networking decisions are based on the content contained in the communication pipes and packets to satisfy the expectations of the users, operators, and the content providers. Such complex tools and techniques reside at the middleware that provides the important glue between the network infrastructure and the services and applications. The content networking middleware provides the necessary hooks in a commonly understood and documented manner to enable application and service developers and content providers to focus on their respective domains and optimize their goals and objectives to satisfy their users. To end users, content networking offers an environment where they have optimized and seamless experience in creating, sharing, managing, and consuming multimedia content. The main focus of this book is on content networking, as illustrated through a simple diagram in Figure 1.1. Nevertheless, there is sufficient background information provided at the lower layers (TCP/UDP/IP, link layer, and physical layer) to the extent that it is relevant to understanding the content networking in the mobile Internet.

In the mobile Internet, mobility can be regarded as another window to the Internet where the access is provided over the air via an access point, and mobility is afforded by locating a user's position and handovers from one access point to the other depending on the coverage. For any technology to be successful, the user experience has to be positive, and this is indeed the case with the mobile Internet today. The inherent unreliability and low bandwidth in the wireless access call for even more innovations at all layers of the protocol stack. In the wireless world of the future, since the mobile devices must be small and lightweight, there are additional challenges of local content storage, delivery, discovery, adaptation, and content presentation. Looking beyond these challenges, however, mobile terminals are

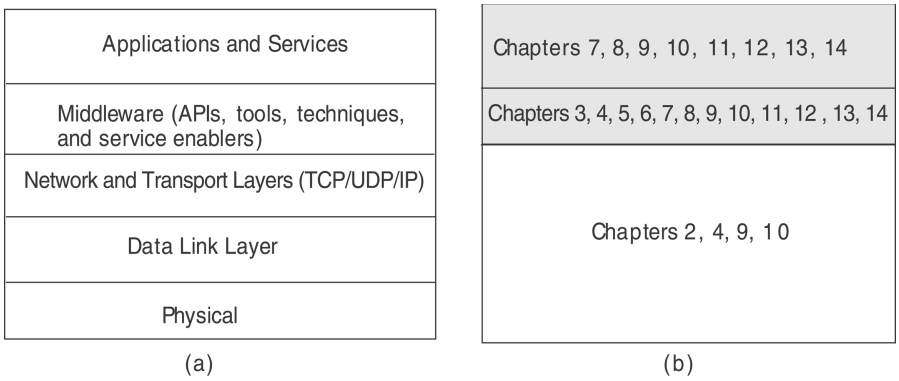

Figure 1.1 (a) Generic TCP/IP protocol stack; (b) shaded area shows the layers where content networking concepts and the main focus of the book reside.

personal devices that users carry almost all the time, and hold the promise of becoming the interface of choice for users to interact with the rich content environment, as shown in Figure 1.2, which depicts devices surrounding a user's immediate proximity (shown by the inner circle), which when interconnected, constitute a personal area network (PAN) or body area network (BAN). These devices could share content and/or services. The outer circle shows the wide area network (WAN) that interconnects the world of personal communications with the rest of the world (cyberworld) in a seamless manner. The content network is an overlay network over the existing IP infrastructure that provides the logical connectivity between the users' devices and the network elements. Any collaboration level infrastructure is the subset of the content network and largely leverages the capabilities and features of the middleware and application layer (e.g., multiplayer games, collaborative business applications).

Content delivery networks (CDNs) are often misunderstood for *content networking*. They are in fact a subset of content networking, and their main purpose is to distribute the content across the global Internet so as to reduce network latency and maximize availability, scalability, and flexibility. CDNs offer the desired feature of providing quality of service (QoS) guarantees at the application layer. The CDNs provide different cost-performance tradeoffs and are targeted for different market segments. The content delivery network service providers made their foray around 1999 by offering—which was then a revolutionary step—to move content across the Internet and serve it from the network edge. This technology is expected to become more intelligent and complex as the Internet becomes more mobile- and Web-enabled.

Because many of the enabling technologies to enhance the user experience in mobile content networking are strongly intertwined, including the Web as one of the dominant communication platforms and user interfaces, it is difficult to focus in depth on each and every aspect. Rather, we focus on the key technologies that provide the foundations to build a content networking infrastructure to suit

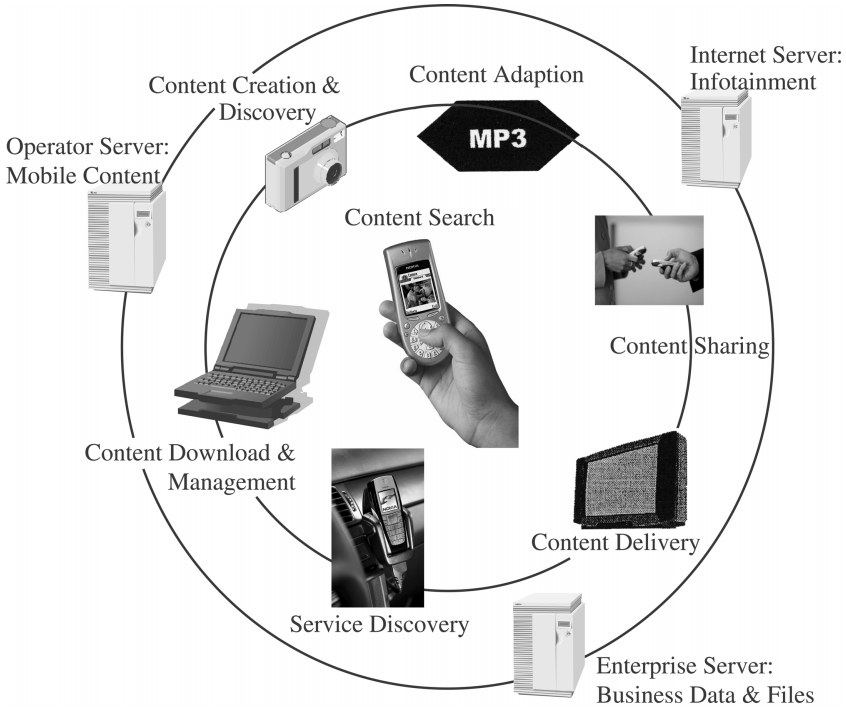

Figure 1.2 The mobile terminal serves as the interface of choice and the center of user experience for end users to interact with the content-rich environment.

the individual requirements and expectations. Next, we provide a brief overview of the various topics covered in this book.

1.3 BOOK OVERVIEW

1.3.1 Chapter 2: Mobile Internet Architecture Overview

This chapter provides a comprehensive introduction to third-generation mobile communications from the perspectives of radio-related issues and standardization. While the first and second generations of mobile communications systems are designed and optimized for voice communications, IP-based data networking is an essential feature of 3G mobile networks. This chapter not only discusses wideband code-division multiple access (WCDMA) technology and the global system for mobile communications (GSM) evolution in detail but also overviews IS-95 (Interim Standard, 1995) and its evolution toward 3G. In particular, the authors cover important recent developments in 3G technologies, such as high-speed downlink packet access (HSDPA), which increases the downlink bandwidth well beyond 2 Mbps (megabits per second).

1.3.2 Chapter 3: Protocols for the Web and the Mobile Internet

The Web, the Internet, and mobility all combined have dramatically impacted the entire world in a short period of about 10 years. This would not have been possible without the appropriate protocols. The Internet Protocol (TCP/IP) was well established when the Web came into existence. Therefore, protocols for the Web were built on top of the IP protocols that formed the lowest common denominator across a multitude of transport and access networks. This chapter presents a history of the World Wide Web (WWW) and describes the major Web protocols and their adjoining layers, including the Hypertext Transfer Protocol (HTTP), Wireless Access Protocol (WAP), Handheld Device Markup Language (HDML), and Wireless Markup Language (WML). The chapter describes how these Web protocols function from one end to the other end.

1.3.3 Chapter 4: Content Caching and Multicast

This chapter describes the nature of Web-based applications, in which the Web is the dominant carrier of information. The Web is used to either *disseminate* or *exchange information*. The chapter discusses the various types of Web content (e.g., static, dynamic, streaming media) and outlines what requirements they pose for the underlying network. The rest of the chapter is devoted to scalable content delivery via proactive multicast and on-demand caching with particular focus on the transport layer to support one-to-many data transport. IP multicast and reliable multicast are described in detail to support the multicast service model. Finally, the chapter presents a novel idea of application layer multicast. Since it is not practical to change all the routers in the Internet that "look" inside the packets, it is indeed possible to add intelligent nodes at strategic locations of the network to build a connection-oriented reliable overlay network, where both receiving and transmitting are done reliably using TCP by the overlay nodes.

1.3.4 Chapter 5: Characterizing Web Workload of Mobile Clients

This chapter investigates the important problem of characterizing mobile Web workload. Understanding the characteristics of mobile Web access is critical in numerous tasks, including network provisioning, developing services that can be scaled to millions of users, and offering content that is optimized to mobile usage patterns. Using mobile Web access logs, the chapter evaluates key attributes of mobile Web workload and provides valuable insights in designing efficient and effective mobile content services.

1.3.5 Chapter 6: ACME: A New Mobile Content Delivery Architecture

A fundamental mechanism of CDNs is exploiting user interest in the same content and distributing content with popular demand to multiple content caches. This chapter extends this principle to accelerate content delivery over the radio link,

which is usually the bottleneck of mobile user experience. The chapter develops the key concept of user interest correlation that identifies the similarity of interest in content between any two users. By exploiting user interest correlation, the chapter develops an *a*rchitecture for *c*ontent delivery in the *m*obile *e*nvironment (ACME), which pushes content to targeted mobile users with high accuracy to improve user experience. This system is highly suitable for mobile content delivery, because it uses user interest correlation to achieve high bandwidth- and terminal power efficiency.

1.3.6 Chapter 7: Content Adaptation for the Mobile Internet

Mobile content networking assumes the formidable task of delivering content of multiple formats and modalities (text, video, audio, etc.) to heterogeneous devices with different capabilities. In addition, mobile users may require that content be delivered in different ways and in different contexts. This chapter first discusses the mobile content and articulates the need for content adaptation for mobile terminals. This is followed by a presentation of several adaptation methods such as transcoding and content selection, and how terminals can signal their capabilities to the server. Two adaptation scenarios essential to mobile content services, mobile browsing and multimedia messaging, are discussed in detail.

1.3.7 Chapter 8: Content Synchronization

Data synchronization is an essential functionality for mobile phones, as users may store and update critical information such as address book, calendar items and emails in various devices, personal computers, or backend servers. This chapter gives a high level overview of content synchronization and specifications developed at the Open Mobile Alliance (OMA).

1.3.8 Chapter 9: Multimedia Streaming in Mobile Wireless Networks

Streaming multimedia is an appealing application for many mobile users. However, it also encounters significant technical challenges because of its complexity and stringent QoS requirements. This chapter presents the QoS requirements and solutions for streaming media services. Furthermore, it presents basic mobile streaming media system architecture and related protocols. Media encoding and decoding methods that are widely used in mobile networks are also discussed.

1.3.9 Chapter 10: Multicast Content Delivery for Mobiles

Envisioned mobile multimedia services consume large amount of radio bandwidth, and multicast is a key enabling technology to support simultaneous multimedia transmissions to multiple mobile receivers. This chapter presents an overview of the motivations, requirements, and mechanisms of providing multicast services in future mobile networks. In particular, this chapter reviews two promising multicast

technologies: IP datacast (IPDC), which originated from digital television broadcasting, and multimedia broadcast/multicast services (MBMS), which is rooted in cellular telecommunications. Combined, they promise to deliver mobile multimedia content to the mass market in a bandwidth-efficient and cost-effective manner.

1.3.10 Chapter 11: Security and Digital Rights Management for Mobile Content

Digital rights management (DRM) techniques protect the intellectual property of the content owner, and are thus critical in offering enforceable, profitable, and easy-to-use mobile content services. This chapter first presents a general overview of information protection and DRM, followed by specific considerations for mobile DRM systems, from both technical and business perspectives. State-of-the-art mobile DRM systems are discussed to give the reader a clear picture of the current status and future trends for mobile DRM.

1.3.11 Chapter 12: Charging for Mobile Content

It is obvious that no service is viable without a proper billing model, and this is critical to the success of the mobile content business. Charging for the mobile content is continuously evolving and is being refined all the time. The charging models vary widely from very simple models where the customer directly pays the content provider over the Web without involvement of the operator to very complex revenue-sharing models where the billing is done to a single entity (mostly the operator) and then is shared among the various participants in the value chain on the basis of some prior agreements. This chapter describes the various mechanisms that enable the mobile content providers to charge for their services. After an overview of charging, accounting, billing, and payment in the context of current fixed and mobile telephony services, the chapter presents several subscription and online billing models on how mobile content could be charged. The chapter concludes with the assertion that (1) for the mass market to develop, charging should be fully differentiated to make it understandable, predictable, and acceptable and (2) charging should not be thrown in at the end of the system and service design; rather, it should be designed in from the very beginning with appropriate flexibility to adapt the billing system to the realities of the marketplace.

1.3.12 Chapter 13: Algorithms and Infrastructures for Location-Based Services

Although providing mobile services that are as convenient and powerful as the wireline Internet is often challenging because of device and radio bandwidth limitations, such mobile services also offer unique values to the end user. Mobile services are usually *situational* or *context-related*, where location is the best understood and most widely used context. Location-based services provide the user with feature

and flexibility that is difficult to achieve in the wireline Internet. This chapter discusses the essential technology components in location-based services, including various location estimation methods as well as service architectures and platforms.

1.3.13 Chapter 14: Fixed and Mobile Web Services

As the Web becomes a valuable tool for people to obtain information, it is also developing into a unifying platform for automated services between computers. Web services facilitate service creation, integration, aggregation, and discovery by standardizing relevant interfaces. This chapter is a tutorial for Web services, covering Web services architecture and components, with focus on simple object access protocol (SOAP) and Web services description language (WSDL). It also discusses important issues such as over-the-air performance when applying Web services architecture in mobile services.

1.4 CONCLUDING REMARKS

Today, content plays a critical role in the success of the Internet, and will be even more important when the access goes broadband wireless with mobility. The ability to access the Internet anytime, anywhere, anyhow (from a multitude of devices) will give impetus to new services and applications, thus increasing the traffic volume enormously and especially burdening the wireless links. The wireless links are inherently limited in bandwidth capacity and are weakly connected, which puts an additional burden on operators and service providers to ensure that the user experience is positive. Coupled with the Internet today being best-effort, something has to be done, short of a forklift upgrade of the whole infrastructure, to deliver acceptable QoS, including reduced latencies in the delivery of the content. Content delivery technologies do exactly that. Additionally, to enhance the user experience even more, tools and techniques can be provided at the middleware and applications layer for managing and manipulating the content. Content delivery, management, and manipulation, typically referred to as *content networking*, also deal with service discovery, content synchronization, Web services, content adaptation, caching, streaming, multicasting, and so on. The mobile Internet cannot be successful unless there are appropriate safeguards for copyrights and charging models to enable all actors in the value chain to make money and pay for the fixed and recurring costs. This book covers the technologies discussed above in depth, and provides relevant pointers to other technologies that are needed or are prerequisite to building cost-effective, interoperable, efficient, and user-friendly mobility-enabled content networks.