♦ Measures of Wireless Data Performance

Qi Bi, Ching-Yao Huang, Patrick Li, and Mark E. Newbury

There is considerable interest within the industry in specifying the performance of wireless data. In this letter, we review and contrast various performance metrics such as peak rates, average throughput, and user experience. We discuss the relevance of these measures to varying applications such as marketing, planning, and technology comparisons.

Introduction

There is considerable interest within the wireless community in specifying the performance of wireless data. The availability of accurate, quantitative measures can clearly facilitate a service provider's choice of technology, specification of system performance requirements, and marketing information.

Historically, wireless mobile service providers have relied upon general measures of cell coverage and capacity to assess wireless voice systems. For example, comparisons across different air interface strategies (e.g., code division multiple access [CDMA], Global System for Mobile Communications [GSM]) are based on cell coverage and the number of channels that can be delivered per unit of radio spectrum. Characterizations of specific market performance have been based on measures such as dropped call rate, handoff success rate, voice quality, and network blocking. These measures are largely technology-independent and correlate well with network performance and voice user experience.

Although these measures are useful in assessing voice performance, they are based on some assumptions that do not apply to data. In particular, the *circuit-switched* concept is inherent in the assessment of voice networks, i.e., each user is assigned a *dedicated* channel with constraints of *fixed* data rate and *real-time* error correction. In contrast, data applications do not require a dedicated channel since data interactions need not be in real time. Accordingly, users can *time-share* a single or a few high-speed data channel(s). In addition, the rate of this channel can be varied in a fashion that exploits user conditions. For example, higher rate data bursts can be delivered to users that are closer to the cell. Finally, any error correction need not be real time. Unlike voice, data frames in error can be retransmitted later in the data stream, and assembled into the final package.

These differences prevent the ready use of voice-based measures in a data system. As a simple example, specification of the (fixed) channel rate in voice is straightforward. In contrast, specification of a single channel rate in data is not possible since the rate varies. Moreover, any statement involving a single data rate (e.g., peak) is hardly meaningful without some indication of how frequently this data rate might occur or whether it is restricted to some subset (e.g., inner) of the cell coverage area. Finally, a rate alone is not necessarily indicative of the pace at which data can be delivered to the end user. The channel may exist only at high error rate, necessitating many retransmissions and an overall slowdown in the speed of actual data (payload) delivery.

This letter suggests that measures indicating *actual* rate of data (payload) delivery *across* cell area are a

Bell Labs Technical Journal 7(3), 219–223 (2003) © 2003 Lucent Technologies Inc. Published by Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com). • DOI: 10.1002/bltj.10028

more accurate means of specifying system performance. These measures are more useful but may need to be supplemented by metrics that directly address user experience. We briefly address data measures below.

Data Measures

A discussion of data measures is best held in the context of data layers or protocols. For clarity, we briefly review the basic architecture (shown in **Figure 1**). The generic architecture (left column) is mapped into layers (right column) specific to this letter.

Briefly, the user interfaces with the network through a number of layers, beginning with his observation point (application layer) and terminating at the physical layer (radio channel). Each layer treats the information received from the immediately overlying layer as payload, adding its own overhead and passing the data to the layer below. The data transmitted at the physical layer thus consists of the original payload from the application layer plus a number of embedded overheads from the overlying layers.

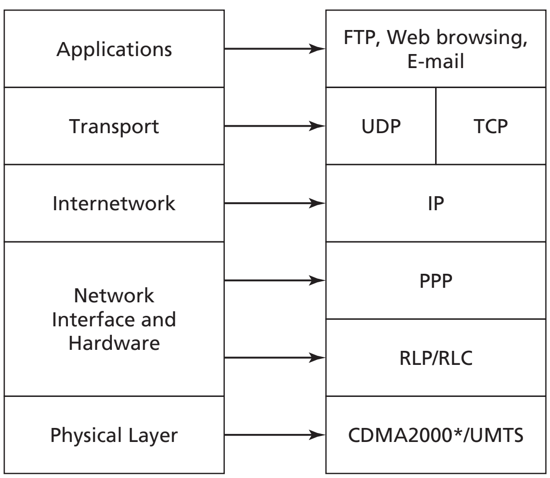
For wireless data, the focus is generally on the physical layer (radio channel) or the layers immediately above (radio link protocol [RLP]). The latter manages transmission across the former. In particular, the RLP retransmits data frames that are reported by the distant receiver as arriving in error.

The *peak data rate* represents the highest channel rate that can be achieved at the physical layer by a user. This single value is often used to market a given technology, without regard to frequency of occurrence, area of availability, or associated channel error rate. Therefore, it alone is a poor indicator of user experience or of relative performance across technologies. For example, CDMA2000* allows a maximum channel rate of 307 kb/s, which in most cases is restricted to an inner subset of a CDMA voice coverage footprint. Enhanced Data Rates for GSM Evolution (EDGE) provides for a maximum rate of 474 kb/s, which cannot be achieved uniformly across the cell footprint *and* requires more complex mobiles that support data transmission simultaneously on eight slots (the maximum value per time slot is 474/8 or 59 kb/s).

Another value often used is the *average user data throughput*, as measured by averaging the ratios of total user data downloaded to *total* user session time across

Panel 1. Abbreviations, Acronyms, and Terms

2G—second generation
3G—third generation
3G1X—CDMA2000* first evolution
1xEV-DO—CDMA evolution-only data
CDMA—code division multiple access
CDMA2000—3G evolution of IS-95 standard
EDGE—Enhanced Data Rates for GSM Evolution
GSM—Global System for Mobile Communications
IS-95—2G CDMA standard
RF—radio frequency
UMTS—Universal Mobile Telecommunications System



FTP - File transfer protocol
IP - Internet protocol
PPP - Point-to-point protocol
RLC - Radio link control
RLP - Radio link protocol
TCP - Transmission control protocol
UDP - User datagram protocol
UMTS - Universal Mobile Telecommunications System

* Trademark of the Telecommunications Industry Association.

Figure 1.
Layers in communication protocol.

the cell coverage area. This value is highly dependent on the data traffic characteristics since the total session time usually includes large intervals of reading or thinking time. To minimize the effects of the think time on this measure, this value can be calculated using the "full-buffer" traffic model, with which the think time is set to zero. For data applications in which the data sources are bursty, the *perceived user data throughput* is often an alternative. This value is defined as the ratio of the download size to the elapsed time, where the latter is measured from the moment of user request (e.g., mouse click) to the completion of a fixed-size data transmission. The perceived throughput is distinct from the *average user data throughput* in that it excludes the data rate in the think time. This is desirable since most users are insensitive to the download speed during the think time.

In contrast to the above values that measure the performance of a single user, the *forward link average aggregate throughput* can measure the network performance, as it captures the average *total* rate of *payload* delivery across the cell coverage area to the hosted data users. While this value is sometimes measured at the physical layer, it is preferred that this value be measured at the RLP, in order to properly account for any slowing of payload delivery caused by a high channel error rate (i.e., caused by many retransmits). As with the single-user measures discussed above, the average aggregate throughput can be measured by using either a full-buffer traffic model or a bursty traffic model. Note that the throughput using the full-buffer model is usually higher than measured network throughputs, since the latter does not include idle time when the network runs out of data to transmit. Nonetheless, it is a useful measure since it can be easily validated in drive test scenarios, where multiple simultaneous large-file downloads can be set up to eliminate the idle time gap.

With the full-buffer traffic model, the average *aggregate* throughput can be roughly related to the average *user* data throughput by multiplying the latter with the number of active users n. This very approximate relation becomes exact under the assumption that the average user data throughput is identical for all users. However, it should be noted that most wireless data systems utilize resource schedulers that try to maximize the network throughput, and in so doing may violate the equal user assumption. In these situations, user throughputs can vary because strategies that maximize the network aggregate throughput often lead to the assignment of higher data rates for users with better radio frequency (RF) conditions. Although other factors (e.g., eventual movement of the data user into a good area or the initiation of service because the user has been ignored for some time) may smooth out the user data rate in such scenarios, individual user throughput differences can remain. Aspects of such aggregate throughput maximization are employed by technologies such as 1xEV-DO and Universal Mobile Telecommunications System (UMTS) high-speed downlink packet access, which serve users by devoting their full downlink power to a single, well-selected subscriber at a time.

To account for this user throughput difference among users, the *minimum user throughput* is often considered in addition to the average user throughput. It is defined as the average user throughput for a user under the worst RF condition for a given unit of time. The difference between the former and the latter can provide indications as to how wide the spread is between rates of different users. It should be noted that the minimum user throughput is not necessarily equal to the minimum user rate defined by the standard, since the resource scheduler may assign the minimum user rate for only a small fraction of time. The minimum user throughput can, therefore, be much less than the minimum data rate defined by the standard.

Examples

We briefly illustrate the concepts above in **Figure 2**, which depicts a Web-browsing session supported by a UMTS wireless data network. At first, the session is in a dormant state where the user has been disconnected from the network due to inactivity. The user initiates a request for data (e.g., a new Web page) via a mouse click and is reconnected to the network after a short access delay. The requested data is downloaded in a series of data bursts with rates that vary as user conditions vary. After download, the user reads

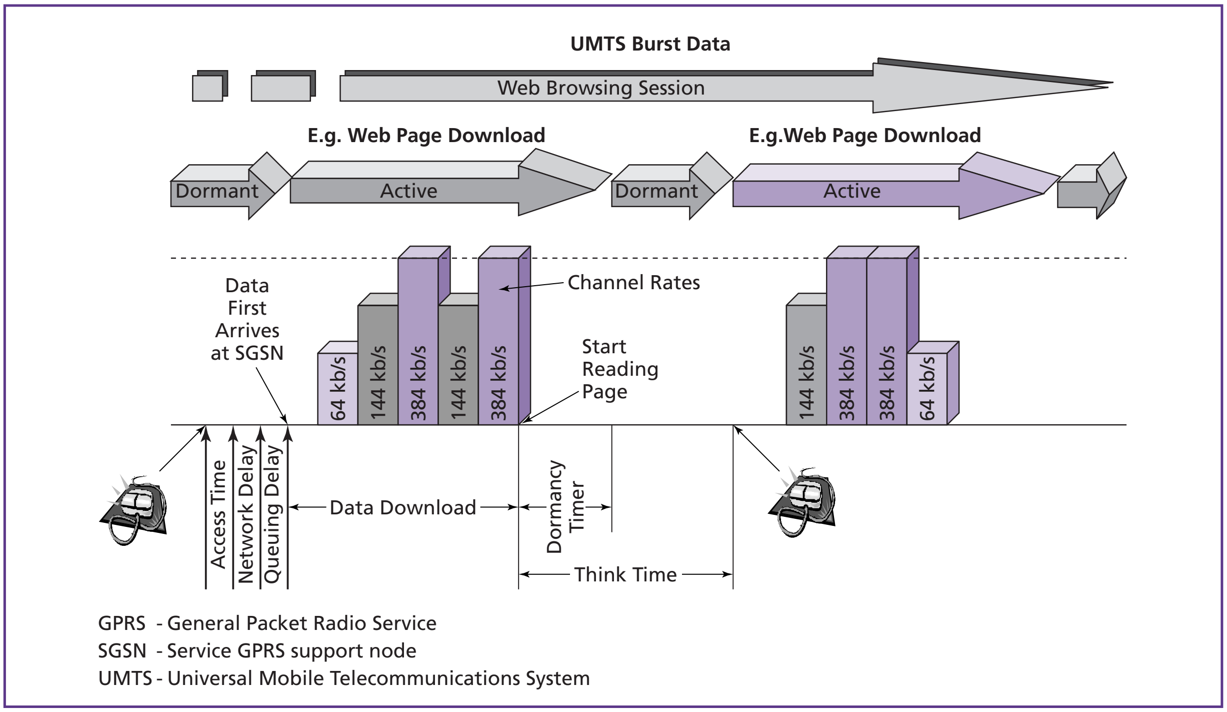


Figure 2.
A sample Web-browsing session.

the page for a period that exceeds the dormancy time, triggering channel disconnect in order to conserve resources. After some period, the user requests another download via a mouse click, and the cycle repeats.

In this example, the UMTS release modeled offers a peak data rate of 384 kb/s, which is experienced, albeit briefly, by the user. The presence of this peak only roughly correlates with user experience, which is better indicated by the *perceived user data throughput*. Presuming channel errors are negligible, the perceived throughput is essentially the average burst rate experienced by the user during the download (~200 kb/s). The value of perceived throughput is much greater than that of average user data throughput, which is the total downloaded files divided by the total session (*including think*) time. Since the think time is typically large compared to the download time, the average user data throughput can be quite modest (e.g., a few kilobits per second).

The experience of a single user does not capture the system performance as a whole. In this example, the average performance of the users scattered across the cell site can be characterized by *the forward link average aggregate throughput*. This measure provides a total average rate of data delivery to all users, each engaged in a browsing session analogous to that shown in Figure 2.

Conclusions

In this letter, we addressed possible data metrics including peak data rate, perceived user data throughput, average user throughput, minimum user throughput, forward link average aggregate throughput, and the number of active users. Although frequently used in marketing, the peak data rate alone is not meaningful without some indication of how often this rate is assigned. The forward link average aggregate throughput is a better measure for the network performance because it captures the total average rate of data (payload) delivery to multiple users scattered across the cell site. Since forward throughput can be enhanced at the possible cost of user experience, additional measures reflecting user experience are useful. The perceived user data throughput captures user experience by characterizing the average data rate experienced by the user during download. This value is distinct from the much lower value of average user throughput, which specifies the rate of data delivery across a total session time that typically includes long think periods between downloads. Accordingly, the average user throughput is useful in planning but is clearly not indicative of the user experience.

***Trademark**
CDMA2000 is a trademark of the Telecommunications Industry Association.

(Manuscript approved October 2002)

QI BI is a technical manager in the AMPS/PCS Base Stations System Engineering Department in Mobility Solutions at Lucent Technologies in Whippany, New Jersey. He holds B.S. and M.S. degrees from Shanghai Jiao Tong University in China, and a Ph.D. degree from Pennsylvania State University in State College. Dr. Bi has led many aspects of the CDMA project at Lucent. He holds over 20 patents and has written numerous papers, the majority of which are in the area of CDMA. He is a fellow of Bell Laboratories and a recipient of Bell Labs President's Gold Award 2000. A senior member of IEEE, he is currently editor of IEEE Wireless Communications Transactions, *and will be serving as the technical vice chair for the IEEE Wireless Communications and Networks Conference in 2003.*

CHING-YAO HUANG is a member of the technical staff in the AMPS/PCS Base Stations System Engineering Department in Mobility Solutions at Lucent Technologies in Whippany, New Jersey. His interests include CDMA radio resource management in power control, call setup, handoff, and scheduler algorithms for data applications, end-to-end performance, and provisioning strategies. He holds a B.S. degree in physics from National Taiwan University in China, an M.S. degree from the New Jersey Institute of Technology in Newark, and a Ph.D. degree from Rutgers University in New Brunswick, New Jersey, both in electrical and computer engineering. Dr. Huang has published more than 50 journal and conference papers and technical memorandums, and he is a chapter author of the Handbook of CDMA System Design, Engineering, and Optimization. *He also holds three U.S. and international patents.*

PATRICK LI is the director of the AMPS/PCS Base Stations System Engineering Department in Mobility Solutions at Lucent Technologies in Whippany, New Jersey. His team is responsible for specifying requirements for Lucent's base station hardware, call processing, and OA&M, as well as characterizing the radio link performance with respect to coverage and capacity for voice and packet data. The team is currently involved in specifying requirements for the OneBTS platform, Flexnet® TDMA and CDMA product lines, and negotiating performance warrantees for UMTS, TDMA, and CDMA (2G, 3G1X, 3G1xEV-DO). Mr. Li holds B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology in Cambridge.

MARK E. NEWBURY is currently a technical manager in the AMPS/PCS Base Stations System Engineering Department in Mobility Solutions at Lucent Technologies in Whippany, New Jersey, where his responsibilities include engineering guidelines, performance analysis, and warranty negotiations for wireless technologies. His work has encompassed system design techniques, performance predictions, and warranty testing methodologies for a wide variety of second-generation and third-generation voice/data air interfaces. A Bell Labs Fellow, Mr. Newbury was a principal contributing author to the Handbook of CDMA System Design, Engineering, and Optimization. *He holds a B.S. degree in physics from Harvard University in Cambridge, Massachusetts, and an M.S.E.E. degree from Syracuse University in New York.* ◆