ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/3436819

# An Empirical Study on the Capacity and Performance of 3G Networks

**Article** *in* IEEE Transactions on Mobile Computing · July 2008

DOI: 10.1109/TMC.2007.70788 · Source: IEEE Xplore

CITATIONS
78

READS
911

**3 authors:**

Wee Lum Tan
Griffith University
**39** PUBLICATIONS **1,071** CITATIONS

SEE PROFILE

Fung Lam
The University of Hong Kong
**4** PUBLICATIONS **147** CITATIONS

SEE PROFILE

Wing Cheong Lau
Chinese University of Hong Kong
**147** PUBLICATIONS **2,680** CITATIONS

SEE PROFILE

All content following this page was uploaded by Fung Lam on 23 February 2015.

The user has requested enhancement of the downloaded file.



SAMSUNG 1023

# An Empirical Study on the Capacity and Performance of 3G Networks

Wee Lum Tan, *Student Member*, *IEEE*, Fung Lam, and Wing Cheong Lau, *Senior Member*, *IEEE*

**Abstract**—This paper presents the findings of an extensive measurement study on multiple commercial third-generation (3G) networks. We have investigated the performance of those 3G networks in terms of their data throughput, latency, video and voice call handling capacities, and their ability to provide service guarantees to different traffic classes under saturated and lightly loaded network conditions. Our findings point to the diverse nature of the network resources allocation mechanisms and the call admission control policies adopted by different operators. It is also found that the 3G network operators seem to have extensively customized their network configurations in a cell-by-cell manner according to the individual site's local demographics, projected traffic demand, and the target coverage area of the cell. As such, the cell capacity varies widely not only across different operators but also across different measurement sites of the same operator. The results also show that it is practically impossible to predict the actual capacity of a cell based on known theoretical models and standard parameters, even when supplemented by key field measurements such as the received signal-to-noise ratio $E_c/N_0$.

**Index Terms**—3G, HSDPA, network capacity, performance, empirical measurements.

## 1 Introduction

UMTS-BASED third-generation (3G) cellular networks are now in operation in many countries around the world. Built upon the WCDMA [1] technology, these 3G networks offer a substantial increase in the capacity for data and voice communications compared to the previous 2G/2.5G networks. HSDPA [2] is an upgrade to the Release 99 versions of the 3G networks, offering higher downlink speed of up to 10 megabits per second (Mbps). In this work, we examine the 3G network capacity and the throughput and delay performances of IP-based applications over 3G and HSDPA networks.

Extensive efforts have been devoted to find the maximum number of simultaneous users that CDMA-based systems can support while maintaining the desired quality of service for each user [3], [4], [5], [6], [7], [8]. However, they are all theoretical studies, which are more useful for preliminary capacity approximation and network planning purposes. There have also been quite a number of field measurement studies on 3G networks, but they are mainly focused on the performance of pure data traffic under lightly loaded or controlled environments [9], [10], [11], [12], [13]. Another study [14] separately measured the voice and data call capacities of an experimental 3G network. However, the measurements were conducted in a fully controlled environment, where the network configuration consisted of only one cell, with the admission control disabled.

• *The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong.*
*E-mail: {wltan3, flam, wclau}@ie.cuhk.edu.hk.*

*Manuscript received 10 Apr. 2007; revised 14 Nov. 2007; accepted 3 Dec. 2007; published online 12 Dec. 2007.*
*For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2007-04-0099. Digital Object Identifier no. 10.1109/TMC.2007.70788.*

To the best of our knowledge, our work is the first large-scale study of its kind, which evaluates the performance of live 3G networks under saturated conditions by using a mixture of data, video, and voice traffic. We have conducted field measurements on three commercial 3G networks in Hong Kong, with the measurement sites spread out over multiple different regions with a large variety of characteristics, ranging from rural to urban sites and commercial to residential sites. The performance of the networks is studied in terms of their data throughput, video, and voice call capacities. Trading off the data saturation capacity against video and voice calls, we examined the behavior of the network resources allocation mechanisms and the call admission control policies in the 3G networks. We also investigated the fairness of the radio-link scheduler in allocating the bandwidth resources to multiple data calls in a data saturated network. In addition, we examined the throughput and latency performances of 3G data services in fully loaded and lightly loaded network conditions. We have also performed data throughput and latency measurements on a lightly loaded HSDPA network and compared the results with those obtained over the 3G networks. The term 3G used in this paper refers to the Release 99 version of the UMTS standards, and unless specified otherwise, the results reported in this paper are those of the 3G networks.

The road map for our paper is given as follows: We will first describe the measurement methodology that we employed in our tests in the next section. Section 3 documents our experimental observations and experiences, whereas Section 4 presents the measurement results from a qualitative and quantitative point of view. We conclude this paper by summarizing our key findings and results in Section 5.

1536-1233/08/$25.00 © 2008 IEEE Published by the IEEE CS, CASS, ComSoc, IES, & SPS

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

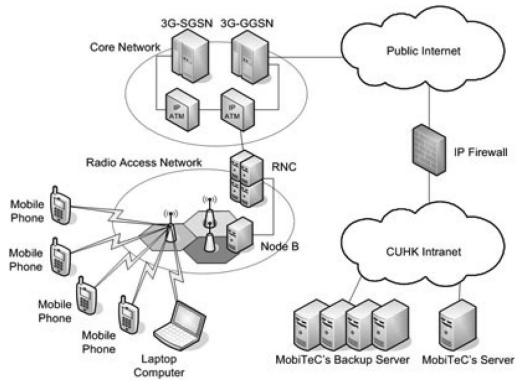



Fig. 1. Measurement setup.

Fig. 2. Measurement equipment.

## 2 Measurement Methodology

Our measurements are designed to investigate

1. the data, video, and voice capacities of 3G networks,
2. the trade-offs between data and video/voice traffic in 3G networks,
3. the trade-offs between video and voice traffic in 3G networks, and
4. the end-user's experience in lightly loaded 3G and HSDPA networks.

Fig. 1 depicts our measurement setup, whereas Fig. 2 shows the equipment used in our measurements. We used a set of client-server programs, of which the client programs were developed on the Java Platform Micro Edition (Java ME), whereas the server programs were developed on the Java Platform Standard Edition (Java SE). For clients, we used Nokia 6680 mobile phones, which are Java-enabled MIDP-2.0 compatible devices. For servers, we used Dell workstations loaded with Linux kernel 2.4.21, running TCP Reno, with the TCP SACK enabled. We also used laptop computers equipped with 3G/HSDPA data cards to remotely control the servers from the measurement sites and to conduct the user experience measurements. In addition, GPS devices and the NetMonitor[1] program were used during the measurement site selection process.

For the data measurements, we measured the 3G network's data capacity in both the downlink and uplink directions. Multiple TCP connections are set up between the mobile phones and the server in order to transfer bulk data in the downlink/uplink directions. During the data transfer, the current individual and aggregate data throughput values are periodically calculated and displayed by the client and server programs, respectively. This is useful for determining the data saturation point and keeping track of the aggregate data throughput value during the different stages of the measurements. In addition, we used Ethereal to capture the packet traces at the server. By analyzing the traces with the aid of tcptrace and a set of scripts that we developed, we are able to determine the throughput and delay performances of the data connections over the 3G network.

1. http://www.symbian-freak.com/quides/netmon.htm.

### 2.1 Measurement Tests

We defined nine measurement tests covering the whole range of services provided by the 3G networks. Table 1 shows the measurement tests and their definitions.

For the measurement tests involving trade-offs between data and video/voice, the network is saturated with data calls first before video/voice calls are made. This is to determine the maximum number of video/voice calls that can be made when the network is saturated with data calls. For more efficient data and airtime consumption, measurement tests involving data are performed consecutively. For example, video calls are added into the network (MT2) immediately after data saturation (MT1) and are then taken away so that the network is saturated with data calls again (MT1) before voice calls are added into the network (MT3).

For the MT7 test, video calls are made until no more video calls can be admitted into the network. Voice calls would then be made until the network is fully saturated. The MT8 test is performed immediately after that, where one video call is dropped, and voice calls are then made in place of the dropped video call. Through these two tests, we can determine the network's video call capacity and identify the number of voice calls that are equivalent to one video call. In all our tests involving video/voice calls, the calls are made in pairs by using the handsets at the measurement site so that we can better utilize our limited handsets to saturate the network.

### 2.2 Measurement Site Selection

Working on the information provided by the Office of Telecommunications Authority (OFTA), Hong Kong, we

TABLE 1
Measurement Tests

| Measurement Test | Definition |
|---|---|
| MT1 | Downlink – Pure data saturation |
| MT2 | Downlink – Data vs. video |
| MT3 | Downlink – Data vs. voice |
| MT4 | Uplink – Pure data saturation |
| MT5 | Uplink – Data vs. video |
| MT6 | Uplink – Data vs. voice |
| MT7 | Video – Pure video saturation |
| MT8 | Video – Video vs. voice |
| MT9 | Speed test & Ping test |

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

selected potential measurement sites based on the following criteria:

1. the potential of receiving a strong dominant signal from a single base station,
2. the diversity of site demographics and characteristics,
3. the accessibility to the site, and
4. the proximity of base stations from different operators.

We also conducted onsite assessments by using the GPS devices and the NetMonitor program in order to locate the physical location that has a strong dominant signal from a single base station with significantly weaker signals from other neighboring base stations. This is to avoid any complications in our measurement tests due to CDMA soft handovers, where a mobile phone could potentially be simultaneously connected to two or more base stations. During our onsite assessments, we have observed that a mobile phone simultaneously connected to two or more base stations usually has weak received signal strengths, which can cause our measurement calls to be easily disconnected. Furthermore, from the NetMonitor program, we observed that the entries on the base station list (showing the base stations that the mobile phone is connected to) are constantly vanishing and reappearing due to the weak signal strengths. Therefore, to avoid these problems while measuring the cell capacity, our chosen measurement sites are the sites that have a strong dominant signal from a single base station. In all our measurements, the signal strength $E_c/N_0$, base station ID, and the GPS location are recorded down before the measurements are started.

### 2.3 Additional Considerations

On the average, each measurement set (downlink measurement set, uplink measurement set, and video measurement set) took about 25 minutes to complete, and the three measurement sets were normally carried out consecutively. Due to its length, measurements were usually conducted during off-peak hours in order to minimize the interruption to normal 3G users due to the overloading of the network during our measurements and, more importantly, to minimize network capacity measurement errors due to other traffic users in the network. In addition, all our measurements were conducted at fixed locations so as to eliminate the effect of mobility on our results.

We have also performed tests to verify that the cause of the low measured throughput values ($<$ 384 kilobits per second (Kbps) per data call) in the 3G networks is not due to any bandwidth bottleneck at the server's network connection or a processing bottleneck at the server itself. In these verification tests, we used the mobile-phone emulator program (supplied with the Java ME Wireless Toolkit) running in two laptop computers to directly connect to the server through a 100-Mbps Ethernet LAN. We set up 20 concurrent TCP connections between the laptop clients and the server. The test results showed that the achievable throughput with the mobile-phone emulator can reach up to 4.5 Mbps (downlink) and 550 Kbps (uplink) at each of the 20 concurrent TCP connections, which is higher than the maximum downlink rate of 384 Kbps offered by the 3G networks. These results proved that the server is capable of serving multiple simultaneous TCP connections at a much higher rate than that seen in the measurements over the 3G networks and therefore verifies that the server's performance is not the cause of the low throughput values observed in the 3G measurements.

Additional tests have also been conducted to determine if the 3G networks performed any data payload compression on packets passing through the networks. In these tests, we used two types of data payloads: one consisting of all "1s" and another one consisting of some random data sequences. The test results show that the achievable throughput values are approximately the same for both types of data payload. This indicates that data payload compression is not performed on packets passing through the 3G networks.

## 3 Experimental Observations and Experiences

We have conducted measurements at 170 sites in a four-month period, covering the networks of three 3G operators in Hong Kong. For ease of reference, the three operators will henceforth be referred to as operators X, Y, and Z in this paper. During our measurements, numerous interesting observations were made regarding the behavior of the 3G networks during saturated condi tions.

### 3.1 Determination of Data Saturation Point

In the MT1 and MT4 tests, data calls are added one by one into the network until the network is saturated. It is seen that the aggregate data throughput value will keep on increasing as the data calls are added until the data saturation point, where it would be at the maximum. However, it is difficult to know in advance the optimal number of data calls needed to saturate the network. One observation that helped in determining this saturation point is that when an additional data call is added into a saturated network, the aggregate data throughput value would drop from its peak value. When this extra data call is dropped, it is observed that the aggregate data throughput value will rise back again. Using this "trial-and-error" method, we are able to find the optimal number of data calls to saturate the network. Another observation that helped in determining the data saturation point is that during the MT4 test in the network of operator Y, repeated attempts to add an additional data call after the data saturation point would fail. Therefore, using this observation, if an additional data call cannot be added into the network, then it means that the data saturation point has already been reached.

### 3.2 Trade-Offs between Data and Video/Voice Calls

During the MT2, MT3, MT5, and MT6 tests, attempts were made to add video/voice calls into the network after the data saturation point. We observed that in almost all of the MT5 tests and in some of the MT2 and MT6 tests in operator Y's network, these attempts would fail, even when more than 30 tries were attempted using more than 10 different pairs of handsets. The implication that video/voice calls may not necessarily have higher priority compared to data calls could be due to the admission control policy employed in operator Y's network, which tries to ensure some form of minimum service for the already-admitted data calls (to be discussed further in Section 4.1.2).

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

Nevertheless, it is generally seen that in the MT2 and MT3 tests, if a video/voice call attempt is successful, the aggregate data throughput value will drop by about 300-400 Kbps, and additional video/voice calls after that can be made easily without repeated attempts. On the other hand, it is also observed that during the MT5 and MT6 tests in the networks of operators X and Z, some of the existing data calls will be dropped by the network when a video/voice call attempt is successful. This is especially so when video calls are added. In view of this, the order of the MT5 and MT6 tests has been reversed. After the saturation with uplink data calls (MT4), the MT6 test is performed first before the MT5 test. This is because the addition of voice calls rarely caused the data calls to be dropped compared to the video calls, and doing the tests in this order enables us to reduce the number of attempts to reconnect the data calls.

### 3.3 Network Stability

After the data saturation point has been reached in the MT1 test, we observed that the throughput values of the data calls in operator X's network are generally fluctuating, whereas the throughput values of the data calls in the networks of operators Y and Z are relatively more stable. The throughput fluctuations observed in operator X's network are probably due to the dynamic bandwidth resources sharing among the data calls as the radio-link scheduler tries to fairly allocate the resources for each call (to be discussed further in Section 4.1.1). As for the MT4 test in operator Z's network, we have observed multiple uplink data calls being dropped after the data saturation point has been reached. We have also experienced instances where all types of connections are suddenly dropped when the network is saturated. This happened in all of the three operators' networks.

### 3.4 Video and Voice Call Quality in Saturated Networks

In the data-video and data-voice trade-off tests, we observed that as long as a video/voice call is successfully added into the data-saturated network, the quality of the connected video/voice call is satisfactory. On the other hand, in the video-voice trade-off test, the successful addition of voice calls can cause some of the video calls to sometimes experience frozen video frames and one-sided conversation, where one party can view the other party of the conversation but not the other way round. However, for our measurements, we do not consider the video/voice call quality but rather only take into account the number of video/voice calls that can be added into the network, as this parameter is more relevant to our network capacity measurement purposes.

## 4 Measurement Results

We now look into our measurement results from a qualitative and quantitative perspective. Due to space limitations, we have selected and will analyze in the following several graphs that are representative of the measurement results from the 3G operators.

### 4.1 Performance in Saturated 3G Networks

We have generated TCP throughput and round-trip-time (RTT) graphs for all the measurement tests that we have done. The aggregate throughput graphs show the total TCP throughput in the cell that is allocated to all the handsets during the measurement test. From these graphs, we can see the effects on the data calls' total throughput levels due to the addition of video and voice calls in the cell. A similar graph is the individual throughput graph, which shows the TCP throughput of each handset during the measurement test and provides an indication of the radio-link scheduler's fairness in terms of the bandwidth allocation to the handsets in the cell. The RTT graphs show the RTT values of each handset during the measurement test and are typically the response times experienced by a data user in a saturated 3G network.

#### *4.1.1 Downlink Measurements*

Fig. 3 shows the downlink measurement results for the 3G operators. We can see from the TCP downlink aggregate throughput graphs in Figs. 3a, 3b, and 3c that when video or voice calls are added into the data-saturated cell, the TCP downlink aggregate throughput for the data calls are reduced drastically. The TCP throughput reduction is higher when video calls are added compared to voice calls, because a video call requires more bandwidth ($\sim$64 Kbps) compared to a voice call ($\sim$12 Kbps). We also observe that more voice calls can be added into the data-saturated cell compared to video calls. In fact, during the MT3 tests, all the handsets have been fully utilized to make the data calls and the voice calls, and as a result, we are not able to investigate if more voice calls could be made. On the other hand, during the MT2 tests, not all the handsets were fully used up. For example, operator X's results in Fig. 3a show that only 12 video calls could be successfully added into the cell, and from our measurement worksheet, we see that attempts to add more video calls failed. As a comparison, results from the MT7 test at the same measurement site show that 13 video calls and three voice calls were required to fully saturate the cell. Fig. 3a shows that for the MT2 test, the TCP downlink aggregate throughput value after the addition of the 12 video calls is approximately 100 Kbps, which is equivalent to the sum of the bandwidth required for one video call and three voice calls. This indicates that the cell's bandwidth resources have been fully shared between the data and video calls during the MT2 tests. Similar observations are also seen in the results for operators Y and Z.

Observations from the MT2 tests seem to suggest that the radio-link scheduler reserves some minimum bandwidth resources for the data calls. It can be explained that the addition of the initial group of video calls is possible, because the bandwidth resource consumption of the data calls at the peak aggregate throughput is much higher than the minimum data bandwidth threshold $\mathrm{BW}_{\text{data-min}}$. As video calls are added into the cell, bandwidth resources are taken away from the data calls (that is, the radio bearers of the data calls are downgraded to lower bit rates), thereby causing the downlink aggregate throughput value and the downlink individual throughput value of the data calls to drop. Once the downlink aggregate throughput value has been reduced to $\mathrm{BW}_{\text{data-min}}$, attempts to add in more video calls would be unsuccessful.

The TCP downlink individual throughput graphs in Figs. 3d, 3e, and 3f show that the addition of video/voice calls into the data-saturated cell caused the individual data calls' throughput to drop significantly from as high as 370 Kbps to as low as 10 Kbps. We also see that in Fig. 3e,

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

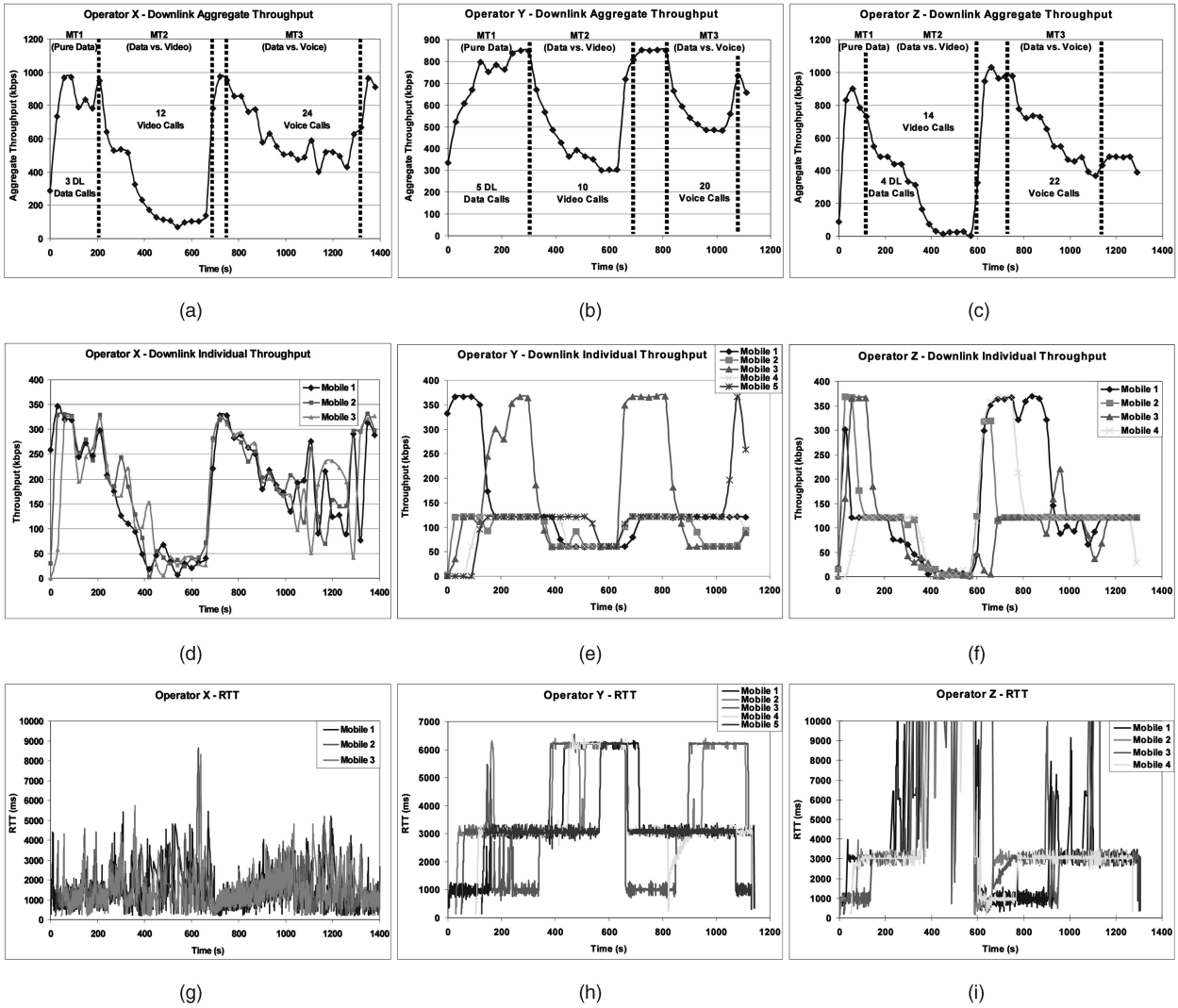



Fig. 3. Downlink measurement results for operators X, Y, and Z.

the TCP throughput values in operator Y's network can be separated into three distinct levels of 370, 125, and 60 Kbps. Fig. 3e also shows that for most of the duration of the downlink measurement test, the different data calls in the cell do not have the same TCP throughput values as each other. While similar results are also obtained in operator Z's network, Fig. 3d on the other hand shows that the data calls in operator X's network generally have the same TCP throughput values as each other. These results suggest that the radio-link schedulers employed in the networks of the three operators implement different bandwidth resources allocation mechanisms, which agrees with our knowledge that the three 3G operators are using different vendors' equipment in their networks (although we do not have access to the algorithms and their parameter values). In terms of fairness in the bandwidth resources allocation, the results suggest that the radio-link schedulers in the networks of operators Y and Z do not treat each data call in the same manner but instead allocate different bandwidth resources to different data calls, thereby causing each data call to achieve different TCP throughput values. On the other hand, the radio-link scheduler in operator X's network is fair in that every data call in the cell is allocated the same bandwidth resources, resulting in all of the data calls generally achieving the same throughput values as each other. The criteria for determining the amount of bandwidth resources to allocate to the data calls could be based on the individual link quality and the amount of data backlog in the network buffer.

The RTT graphs in Figs. 3g, 3h, and 3i show that the RTT values for each data call are inversely proportional to its TCP throughput values. While Fig. 3g shows that the RTT values for the data calls in operator X's network are generally the same as each other, Fig. 3h on the other hand shows that the RTT values for the data calls in operator Y's network can be separated into distinct levels of 1,000, 3,000, and 6,000 ms, inversely corresponding to the different TCP throughput levels observed in Fig. 3e. The RTT results for operator Z's network are similar to those in operator Y's network in that the RTT values can be divided into distinct values of 1,000 and 3,000 ms. The difference is that the RTT values during the MT2 test are exceptionally high, that is, above 100 seconds (out of the range shown in Fig. 3i), corresponding to the low TCP throughput values ($\sim$10 Kbps) shown in Fig. 3f. It is worth noting that even with such high RTT and low TCP throughput values during

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

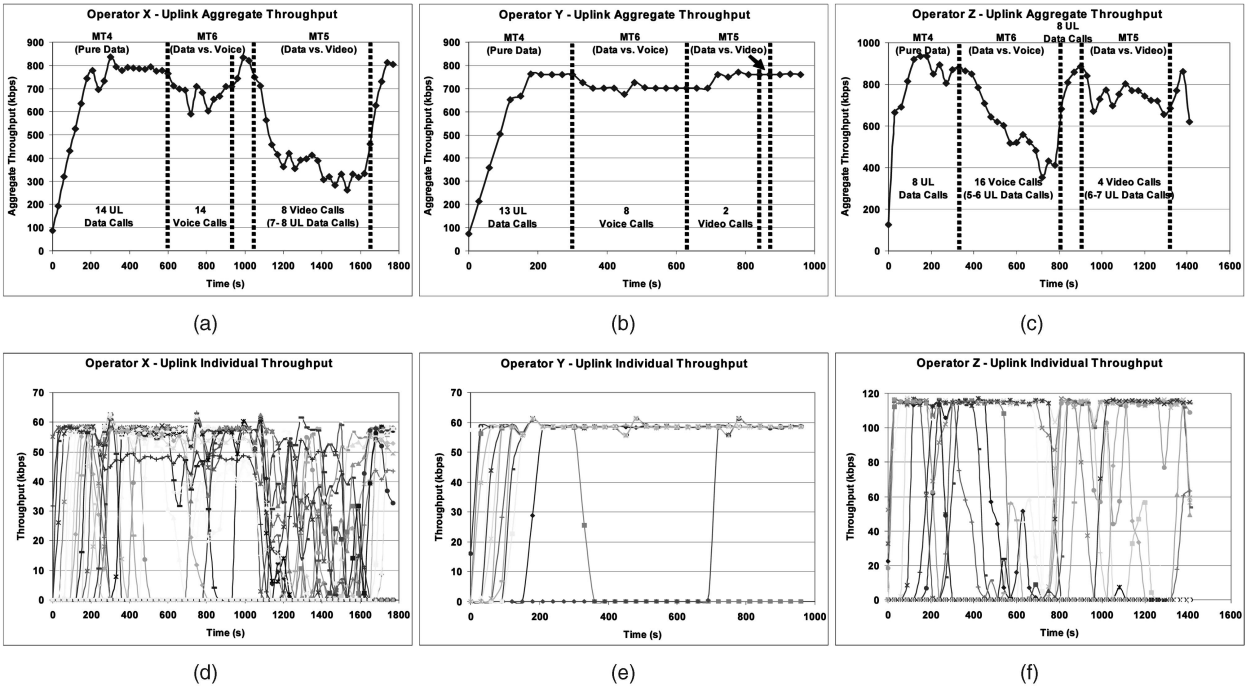



Fig. 4. Uplink measurement results for operators X, Y, and Z.

the MT2 test, the data transfer across the data calls in operator Z's network was still able to continue.

### 4.1.2 Uplink Measurements

The uplink measurement results are shown in Fig. 4. We can see from operator X's results in Fig. 4a that the addition of video calls during the MT5 test caused some of the data calls to be dropped. Even though there were attempts to reconnect the data calls, the number of data calls that could eventually coexist with the video calls in the MT5 test is smaller than that in the MT4 test. In Fig. 4a, we see that after eight video calls were added into the cell, the total number of data calls dropped from 14 to about seven to eight data calls, and attempts to add more video calls in the cell were unsuccessful as well. Similar observations are also seen in operator Z's results in Fig. 4c. Just like in the downlink measurement results, these observations also suggest the existence of $\mathrm{BW}_{\text{data-min}}$. Using a similar explanation as in the downlink case, the addition of the initial group of video calls is possible, because the bandwidth resource consumption of the data calls at the peak aggregate throughput is much higher than $\mathrm{BW}_{\text{data-min}}$. As the video calls are added into the cell, bandwidth resources are taken from the uplink data calls. Unlike downlink data calls, which are allocated higher radio bearer rates, uplink data calls are generally allocated the minimum radio bearer rate of 64 Kbps. Hence, when bandwidth is taken away, some of the uplink data calls would be dropped. The observations also suggest that more than the required number of uplink data calls were dropped, thereby causing the uplink aggregate throughput value to drop lower than $\mathrm{BW}_{\text{data-min}}$. When some of the data calls were reconnected, the uplink aggregate throughput value rose back to $\mathrm{BW}_{\text{data-min}}$. Once this happens, attempts to add in more video calls would fail.

Conversely, our uplink measurement results for operator Y's network (Fig. 4b) generally show that the additions of video/voice calls would not cause the dropping of the existing data calls. In addition, from our measurement worksheet, we see that data calls could no longer be added after uplink data saturation, and attempts to add more voice calls and video calls during the MT6 and MT5 tests were also unsuccessful after the initial successful addition of eight voice calls and two video calls, respectively. These observations suggest that in operator Y's network, once a data call is admitted into the cell, it would not be preempted by new video calls. Coupled with the fact that data calls could no longer be added, it implies that there is an upper limit on the number of data calls that can be admitted into the cell. It also suggests that the call admission control policy in operator Y's network operates in such a way as to guarantee some form of minimum service for all admitted calls. The initial group of voice/video calls could be added into the cell, probably because after the upper limit of the data calls has been reached, there is still some minimal amount of bandwidth resources left to admit the voice/video calls. However, once these remaining bandwidth resources are used up, voice/video calls could no longer be added.

Fig. 4d shows that the maximum TCP uplink individual throughput value for the data calls in operator X's network is about 60 Kbps. Although similar results are obtained for the uplink measurements in operator Y's network, we found that the data calls in operator Z's network can achieve a throughput value of up to 115 Kbps (Fig. 4f). Since the same handsets were used in all our tests, this throughput difference points to a different configuration setting in operator Z's network.

## 4.2 Performance in Lightly Loaded 3G Networks

The MT9 test is designed to investigate the 3G user experience, using a laptop equipped with a 3G data card

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

TABLE 2
Comparison of Link Speed (in Kbps)

| Operator | Overseas Website | | | Local Website | | |
|---|---|---|---|---|---|---|
| | *Max* | *Min* | *Avg.* | *Max* | *Min* | *Avg.* |
| X | 104.3 | 18.0 | 68.7 | 329.4 | 130.6 | 204.6 |
| Y | 200.0 | 22.7 | 89.9 | 368.6 | 33.5 | 228.1 |
| Z | 142.2 | 42.0 | 82.3 | 233.1 | 164.8 | 199.4 |

in a lightly loaded cell scenario, as compared to our other measurement tests, which were conducted with multiple handsets in a fully loaded cell scenario. Table 2 shows the speed test results[2] for the three operators. As expected, the link speed to the local Web site is higher than that to the overseas Web site for all three operators. The average link speeds of the three operators across all the measurement sites are more or less the same, whereas their maximum link speeds to the local Web site are still within the same order of magnitude compared to the maximum downlink throughput values shown in Figs. 3d, 3e, and 3f. This is because the link speeds are bounded by the maximum downlink rate of 384 Kbps offered by the 3G networks.

Due to the blocking of ICMP messages by the 3G networks, we chose to use the TCPing[3] tool for our ping tests. TCPing is a TCP-based ping implementation for the Win32 platform. In measuring the RTT from the 3G networks to various Web sites located in Hong Kong and overseas, TCPing sets up four consecutive TCP connections with the Web sites. The time taken to complete each TCP connection establishment between our laptop client and the Web sites is the RTT. Looking through the source codes of TCPing, it is seen that between successive pings (TCP connection establishments), if the previous RTT is less than 2,000 ms, then TCPing will sleep for a period that is equal to the difference between 2,000 ms and the previous RTT value. Although this sleep feature of TCPing has no effect on the achievable RTT values in 3G networks, it will be shown to have an effect on the achievable RTT values in HSDPA networks, as will be discussed later in Section 4.7.

Fig. 5 shows the ping test results, where the range of RTT values shown is from the 5th percentile to the 95th percentile, with the horizontal line representing the 50th percentile. We see that the RTT values to the overseas Web sites are generally higher than that to the local Web sites due to the long propagation delay to overseas Web sites. In addition, the minimum and average RTT values achieved across all local Web sites for (operator X, operator Y, operator Z) are (187 ms, 109 ms, 140 ms) and (385 ms, 350 ms, 201 ms), respectively. This shows that the minimum RTT value achieved in lightly loaded cells is more than five to nine times lower than the lowest RTT value of 1,000 ms achieved in a fully loaded cell, as shown previously in Figs. 3h and 3i. The RTT values from our ping tests also correspond to those reported in [9], [10], [11], and [12].

We have also performed the ping test by using several broadband network providers (using ADSL, cable modem, and Ethernet) in Hong Kong and found the average RTT value to local Web sites to be 8.69 ms, which is significantly lower than the minimum RTT values achieved with lightly loaded 3G networks. Since the round-trip transmission time of a ping packet over the 3G wireless links is only about 9 ms (50 bytes/200 Kbps + 50 bytes/60 Kbps), this clearly shows that the end-to-end data latency over the 3G wireless networks is dominated by the network processing and queuing delay, which amounts to more than 100 ms at a minimum.

2. Connection speed to the overseas and local (Hong Kong) Internet speed test Web sites http://www.auditmypc.com/broadband-speed-test.asp and http://www.netfront.net/speedtest/testspeed1.htm, respectively.

3. http://www.elifulkerson.com/projects/tcping.php.

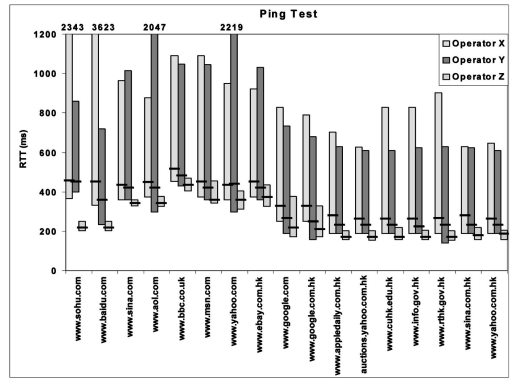



Fig. 5. Ping test results.

### 4.3 TCP Retransmissions in 3G Networks

We have calculated the percentages of TCP retransmissions in all our measurement tests involving data, where it is defined as the percentage ratio of the number of retransmitted bytes over the total number of bytes transmitted during the measurement tests. Fig. 6a shows that the majority of the measurement sites in operator X's network have a TCP retransmission rate of 1 percent, with some sites even reaching up to 6 percent. In contrast, results for operators Y and Z show that all of the measurement sites in the networks of operators Y and Z have TCP retransmission rates below 1 percent, with the majority

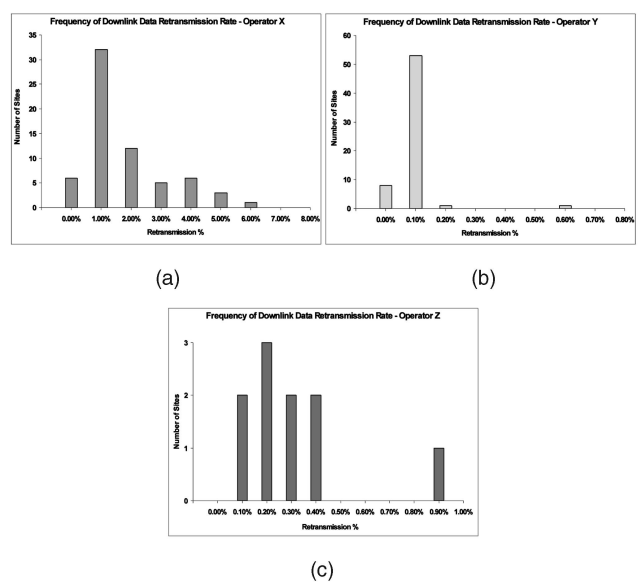
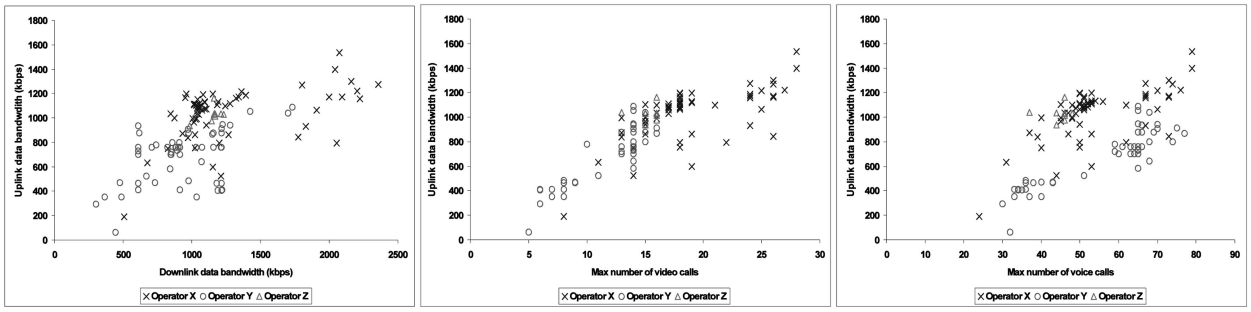



Fig. 6. Percentages of TCP retransmissions for operators X, Y, and Z.

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.




(a) (b) (c)

Fig. 7. Capacity measurement results for operators X, Y, and Z.

of the sites having TCP retransmission rates of 0.1 percent and 0.2 percent, respectively. These results suggest that the networks of operators Y and Z may have been configured to use the RLC Acknowledged Mode, thereby lowering their TCP retransmission rates. The results also suggest a low threshold value for the targeted signal-to-noise ratio in operator X's network to allow for higher network capacity in terms of the data bandwidth capacity. This is in line with our capacity measurement results (to be discussed in the next section), which shows that operator X generally has the higher data capacity results compared to the other two operators.

### 4.4 3G Network Capacity Results

We have collated our measurement results and generated several graphs showing the 3G network capacities in terms of the downlink and uplink data bandwidth capacity and the maximum number of video and voice calls that can be made in the cell. These results are representative of the 3G network capacity as perceived from an end-user's viewpoint.

Fig. 7a presents the measured uplink and downlink data bandwidth capacity results for the three operators. The data points represent the maximum aggregate throughput value out of all throughput values calculated over 30-seconds averaging windows throughout the whole duration ($\sim$1,500 seconds) of the downlink and uplink measurement tests, with each data point corresponding to the results for one measurement site. Generally, we observe a positive correlation between the uplink and downlink data bandwidth capacities. As a whole, the uplink data bandwidth capacity is usually lower than the downlink data bandwidth capacity for a measurement site. Figs. 7b and 7c also show a positive correlation between the measured uplink data bandwidth capacity and the video and voice capacities. This is similar to the observation in Fig. 7a, and we also obtained likewise results when plotting the downlink data bandwidth capacity against the video and voice capacities. In determining the voice capacity, we have utilized the MT7 and MT8 test results to work around the limitation on the number of handsets that we have. From the MT8 test results, we can identify the number of voice calls that are equivalent to one video call. Table 3 shows a summary of the MT8 test results. To calculate the network's voice capacity, we use the formula "$Voice\ capacity = (r \times v) + o$," where $r$ is the ratio of the number of voice calls to one video call, $v$ is the maximum number of video calls in the MT7 test, and $o$ is the number of voice calls in the MT7 test.

From the measurement results, the mean values and standard deviations for the four parameters of downlink and uplink data bandwidth capacities (in Kbps) and the maximum number of video and voice calls that can be made for the three operators are tabulated in Table 4. The values are represented in the format (operator X, operator Y, operator Z). The cumulative distribution graphs for the four parameters are also plotted and shown

TABLE 3
Number of Voice Calls Equivalent to One Video Call

| Operator | Mean | Median | Standard Deviation |
|---|---|---|---|
| X | 2.83 | 3 | 1.20 |
| Y | 4.52 | 4 | 1.71 |
| Z | 2.90 | 3 | 0.88 |

TABLE 4
Mean and Standard Deviations of the Four-Tuple Parameters for Operators X, Y, and Z

| Parameter | Mean | Standard Deviation |
|---|---|---|
| Downlink data BW | (1259.9, 926.4, 1143.4) | (424.04, 296.68, 77.24) |
| Uplink data BW | (1041.3, 686.3, 1029.2) | (208.54, 209.56, 57.32) |
| Video calls | (18.88, 12.13, 15.4) | (4.38, 3.07, 0.97) |
| Voice calls | (54.29, 57.43, 44.9) | (11.52, 13.58, 3.0) |

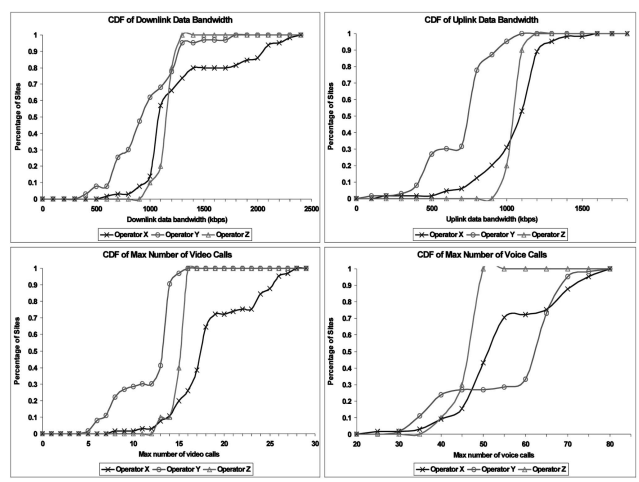



Fig. 8. CDF of the capacity measurement results for operators X, Y, and Z.

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

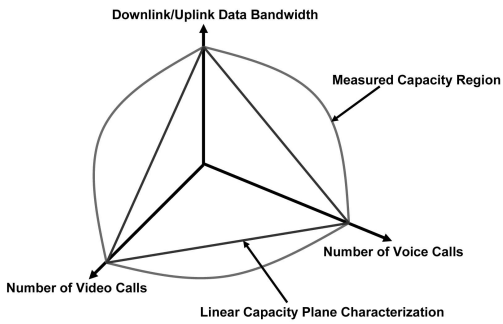



Fig. 9. Capacity region of a cell.

in Fig. 8. Based on our trade-off test results (MT2, MT3, MT5, MT6, and MT8), we have also constructed the capacity region of a cell, as shown in Fig. 9, for the case of heterogeneous services. A capacity region is a graphical representation of the cell capacity in terms of the mixture of circuit-switched and packet-switched services. The analysis in [4], [5], and [6] have suggested that the boundary of the capacity region can be characterized by a linear plane, as shown in Fig. 9. To verify the accuracy of this linear characterization (approximation), we have plotted in Fig. 10 the cross sections of the normalized downlink/uplink bandwidth capacity versus the normalized video call capacity for more than 40 cells of operator X. In Fig. 10, we can observe that the linear capacity plane characterization turned out to be a good conservative estimation for more than 70 percent of the cells. The data points that fall below the linear capacity plane are relatively few and not that far from the 100 percent scaled capacity line in Fig. 10. By further moving the linear capacity plane to the 90 percent scaled capacity line, we are able to obtain the linear capacity plane characterization of the capacity region's boundary for more than 90 percent of the cells. Similar results are also observed for other service trade-offs for the other operators.

### 4.5 Modeling of Measured 3G Network Capacity Results

Fig. 11 shows the measured cell capacities plotted with respect to the received $E_c/N_0$ at their respective measurement sites. The value for the $E_c/N_0$ parameter is the average of the received $E_c/N_0$ measured with six different handsets at the beginning of every measurement test. Figs. 11a and 11b show that the majority of both the downlink and uplink data bandwidth capacity results are spread across the $E_c/N_0$ range of $-2$ dB to $-6$ dB without any identifiable trend. Even for the same value of $E_c/N_0$, the data bandwidth capacities differ substantially across different cells and operators. Similar observations are also shown in Figs. 11c and 11d with respect to the relationship between the video or voice capacities and the received $E_c/N_0$. As such, it is obvious that we cannot use measurements of $E_c/N_0$ alone to predict the network capacity.

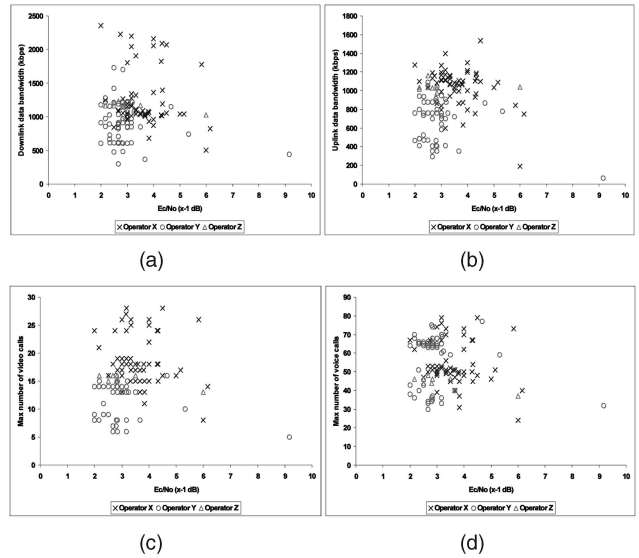



Fig. 11. Cell capacities versus $E_c/N_0$ for operators X, Y, and Z.

We next try categorizing the measurement sites into different classes based on some common site characteristics like relative population densities (rural, suburban, and urban) or the usage nature of the sites (commercial, residential, and industrial) to ascertain if we can identify any distinct clustering among the measured cell capacity results. However, Fig. 12 shows that even when the measurement sites are grouped into rural, suburban, and urban sites, we are still unable to identify any distinct clusters among the measured data bandwidth capacity results of operator Y. Similar results are also obtained when measurement sites are grouped according to the usage nature of the sites and when repeating these efforts on the measured video and voice capacity results for all three operators.

To explore further, we decided to study the capacity distribution patterns of the three operators by mapping the

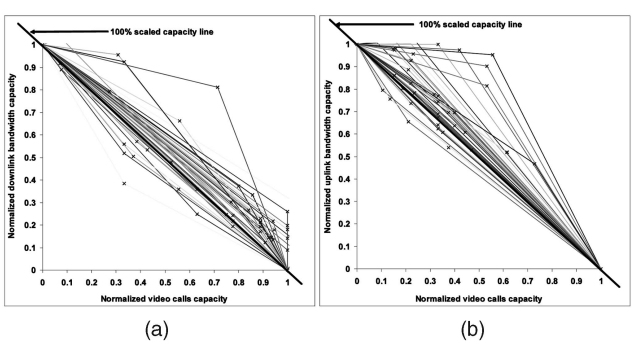



Fig. 10. Normalized capacity trade-offs between downlink/uplink data bandwidth and the number of video calls for operator X.

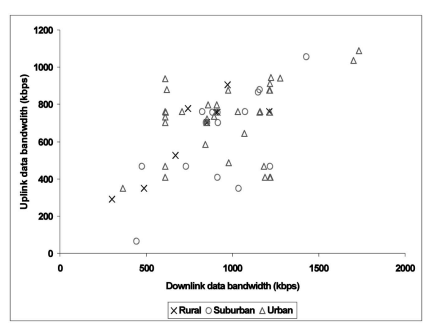



Fig. 12. Data bandwidth capacities, with measurement sites categorized according to population densities for operator Y.

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

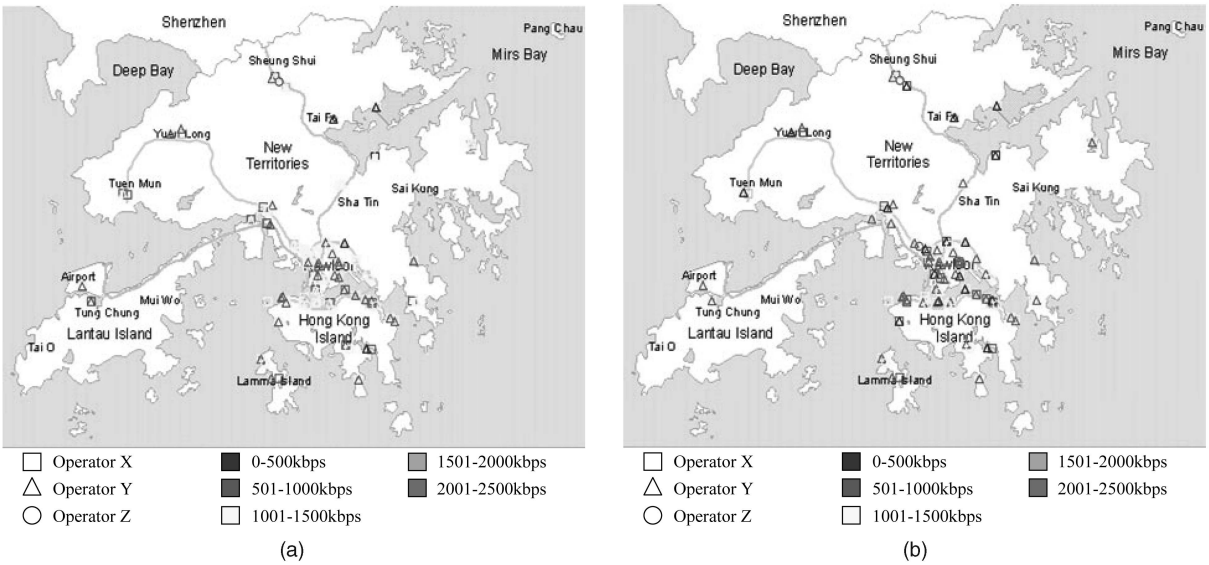



Fig. 13. Distribution of data bandwidth capacities for operators X, Y, and Z. (a) Downlink data bandwidth capacities. (b) Uplink data bandwidth capacities.

data, video, and voice capacity results onto their respective measurement locations in the Hong Kong map. Each measurement site's location is drawn based on its GPS location information. Since the capacity distribution maps for the data, video, and voice capacities are by and large similar to one another, we will only show the capacity distribution map for the data bandwidth capacities.

Fig. 13 shows that the majority of the measurement sites are concentrated in the areas of Kowloon and the north of Hong Kong Island. These densely populated urban areas are the highly busy and important commercial zones in Hong Kong, and hence, the concentration of base stations from all three 3G operators in these areas is very high as well. This is corroborated with the cell site information used previously in the measurement site selection process described in Section 2.2. Covered by relatively small cells, Fig. 13 shows that these areas have medium to high cell capacities. On the other hand, sparsely populated rural areas (northeast part of New Territories and Lamma Island) are often covered by relatively large cells, with low to medium cell capacities. Fig. 13 also shows several areas located in New Territories (for example, Tuen Mun, Tung Chung, Sheung Shui, and the Airport) with medium to high cell capacities. These suburban areas are often covered by relatively large cells, as concluded from the low concentration of base stations there. These observations suggest that the 3G network operators seemed to have customized their network configuration in a cell-by-cell manner according to their own proprietary requirements in order to cope with the projected traffic demand of individual sites and neighborhood.

Based on the observations in Fig. 13, we decided to investigate the relationship between the measured cell capacity and the cell coverage size. This relationship is supported by well-known theoretical models in CDMA systems [1, Chapter 8], [15], as shown in Fig. 14. In our investigation, we used an estimated cell radius value (based on the cell site information that we had) to approximate the target coverage area of a cell. In our scheme, the nearest base station to the measurement site is chosen to be the targeted base station, and the cell radius is estimated to be half the average distance between the targeted base station and the $k$-nearest base stations. We assume the use of omnidirectional antenna at the base stations. However, our approximation method cannot account for other factors that can influence a cell's target coverage area such as the use of a directional antenna, the antenna height and tilt angle, and the landscape in the area. For example, we found instances where the base stations are located somewhat along a linear trajectory on highways. In these instances, $k = 1$ would give a good estimate of the targeted cell's coverage area. On the other hand, values of $k > 1$ would give a better estimate of the cell's coverage area in instances where the $k$-nearest base stations are located around the targeted base station. Using these estimated cell radius values, we try identifying their correlation with the measured cell capacities, with the measurement sites grouped according to their relative population densities.

Fig. 15 shows the measured downlink data bandwidth capacity for operator X, plotted with respect to the estimated cell radius for two values of $k = 1$ and $k = 3$. However, we

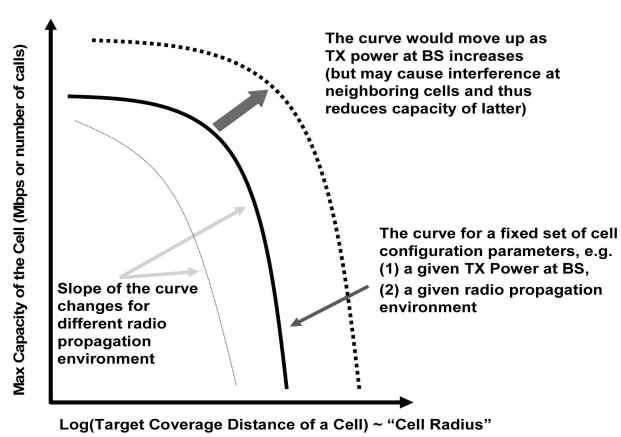



Fig. 14. Theoretical model for cell capacity versus cell radius.

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

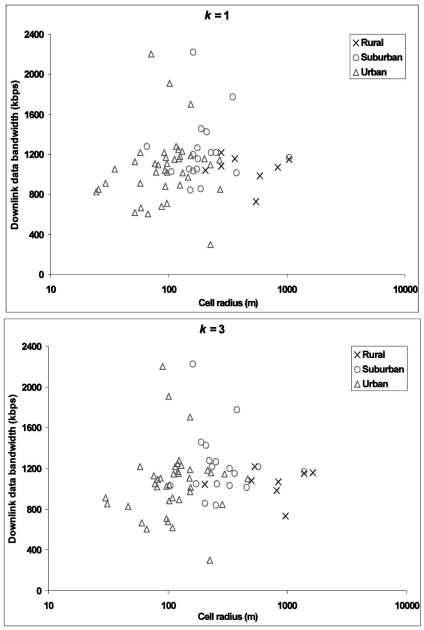



Fig. 15. Downlink data bandwidth capacity versus cell radius for operator X.

are still not able to identify any clear set of clusters among the measured data points. Similar results are also obtained for the other network capacity parameters for all three operators. This observation reinforces the point that the actual settings or parameters employed for each cell seemed to be very diverse across different cells and that the cells cannot be grouped into a small number of sets for categorization purposes. It is obvious that the measured results do not fit the predictions of well-known theoretical capacity models using a reasonably small set of values for the configuration parameters nor can the measured capacity data in one cell be used to predict the capacity of another cell.

### 4.6 Repeated Measurements

In order to investigate the relationship of the 3G network capacity with the time of the day, we have performed repeated measurements at several sites on different days and times of the day. In general, we found that the video call capacity in the first and second measurements do not differ much, with an average and maximum differences of 3.9 percent and 14.3 percent, respectively. However, the data bandwidth capacities can vary substantially in some situations. Table 5 shows the results of the repeated measurements for downlink data bandwidth capacities, where each measurement result has been classified into one of the *network activity factor* categories of "very busy," "moderately busy," and "near idle." This classification is based on our best knowledge of the measurement site's characteristics during the time of measurement. We observe that for seven out of the 10 sites, the repeated measurements were conducted during times when the network is considered to be moderately busy or nearly idle, and as such, there is only a small capacity difference between the first and second measurements (average and maximum differences of 4.0 percent and 12.5 percent, respectively). The only exception is the repeated measurement at site B, which is a suburban measurement site located next to a warehouse and a construction workplace, where we observe a capacity difference of 35.1 percent, even though we presumed the network loading to be moderately busy during the first and second measurements. This considerable capacity difference may have been caused by an unforeseen heavy network loading or changes in the physical and environmental surroundings between the measurement times of the first and second measurements.

On the other hand, the repeated measurements at sites F and G were performed during periods of time when the network loading is considered to be vastly different ("very busy" versus "nearly idle"), and thus, it is not a surprise that there is a big improvement (35.6 percent to 48.5 percent) in the downlink data bandwidth capacity between the first and second measurements. These findings indicate that different measurement times (off-peak hours versus peak hours) can potentially affect the capacity measurement results, particularly in busy commercial zones. However, the majority of our measurements are conducted during off-peak hours, and hence, we are able to minimize any network capacity measurement errors due to factors beyond our control such as the network loading. Furthermore, in the case of the data bandwidth capacity measurements, we

TABLE 5
Downlink Data Bandwidth Capacities (in Kbps) in Repeated Measurements (Measurement Times in Brackets)

| | Operator X | | | | | Operator Y | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Site A | Site B | Site C | Site D | Site E | Site F | Site G | Site H | Site I | Site J |
| | Urban Residential | Suburban Warehouse/ Construction Site | Suburban Residential | Urban Market Place | Urban Ferry Pier | Urban Commercial Offices | Urban Commercial Shopping | Rural Ferry Pier | Suburban Shopping | Urban Residential |
| Very Busy | | | | | | 1097 (15:43) | 892 (18:00) | | | |
| Moderately Busy | 1034 (14:10) | 1395 (14:40) | 2093 (18:46) | 1045 (12:54) | 2193 (15:32) | | | | 1427 (11:10) | 1156 (22:21) |
| | 1094 (01:12) | 2149 (09:25) | | | | | | | 1458 (13:31) | 1157 (15:23) |
| Near Idle | | | 2391 (06:17) | 1052 (22:05) | 2359 (03:10) | 1702 (21:14) | 1733 (02:09) | 669 (20:02) | | |
| | | | | | | | | 670 (15:00) | | |
| Difference | 5.48% | 35.09% | 12.46% | 0.67% | 7.04% | 35.55% | 48.53% | 0.15% | 2.13% | 0.09% |

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

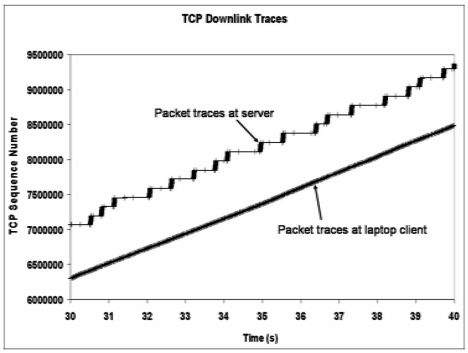



Fig. 16. TCP downlink traces in a lightly loaded HSDPA network.

increase the likelihood of being able to measure the network data capacity by taking the maximum aggregate throughput value out of all throughput values calculated over 30-seconds averaging windows throughout the whole duration (~1,500 seconds) of the measurement tests.

### 4.7 Performance in a Lightly Loaded HSDPA Network

Operator Y's 3G network has recently been upgraded to support HSDPA, with an advertised maximum downlink speed of up to 3.6 Mbps. In order to determine the performance improvements over the 3G networks, we have carried out TCP bulk data transfer measurements (downlink and uplink) at several different locations in Hong Kong by using a single laptop equipped with a USB modem containing an HSDPA data card connecting to our server over a lightly loaded HSDPA network. Packet traces were captured at both the server and the laptop client. Analysis of the TCP downlink traces seemed to suggest that the TCP connection between the server and the laptop client is not end-to-end but rather has been split into two at a TCP Performance-Enhancing Proxy (TCP PEP) [17] in the HSDPA network. TCP PEPs are often employed at the intersection of wired and wireless networks to mitigate performance problems arising from wireless link transmission errors or link disconnections. Nevertheless, the introduction of the TCP PEPs also leads to the violation of the end-to-end semantics of the TCP connections, additional processing delay over the TCP connections, and serious scalability issues in highly loaded networks [17].

Evidence of the existence of the TCP PEP are seen in the snapshots of the packet traces in Fig. 16, in which we see the data packets being sent out in bursts at the server, whereas at the laptop client, the data packets are received in a continuous and regular manner. Closer inspection of the packet traces at the server revealed that the received TCP acknowledgments were periodically advertising a receiver window size of zero (TCP Zero Window), thereby preventing the server from sending more data and causing data packets to be sent in bursts when the TCP receiver window opens again. The examination of the server's packet traces also showed that the maximum advertised TCP receiver window size was 64 Kbytes and that there were very little TCP retransmissions during the data transfer. On the other hand, inspection of the packet traces at the laptop client showed not only that the client's TCP receiver window did not close at all but also that the advertised receiver window value was a steady 128 Kbytes in all the transmitted TCP acknowledgment packets (the TCP Large Window option was enabled in the laptop client). All these observations seemed to suggest the existence of a TCP PEP in the HSDPA network that served to receive, acknowledge, and buffer the data packets from the server and forward the packets over the wireless link to the laptop client.

TABLE 6
TCP Throughput Results in a Lightly Loaded HSDPA Network (in Kbps)

| | Max | Min | Average |
|---|---|---|---|
| TCP Downlink | 2970.64 | 972.69 | 1820.61 |
| TCP Uplink | 365.62 | 59.49 | 265.25 |

As a further validation test, we performed another test, where midway through the TCP downlink data transfer, the USB modem's cable was disconnected from the laptop client. Once this happened, the client program in the laptop immediately terminated. However, the server program continued running in the server, and analysis of the server's packet traces revealed that it was still receiving TCP acknowledgments for its transmitted packets, albeit with a decreasing TCP receiver window size that eventually closed, as indicated by the reception of a TCP Zero Window segment. What was surprising is that long after the disconnection of the USB modem, the server was still sending out TCP Keep-Alive messages and receiving TCP ACK messages as replies from some entity in the HSDPA network. The interval between successive TCP Keep-Alive messages was progressively doubled until it reached a value of 120 seconds, and thereafter, the TCP Keep-Alive messages were sent out every 120 seconds. This went on for about 1,200 seconds until the server finally received a TCP RESET message, which caused the server program to terminate. Since the client program has long since ended, the receptions of the TCP ACK messages and the TCP RESET message at the server pointed to the existence of a TCP PEP in the HSDPA network.

Table 6 shows the TCP throughput results obtained in our measurements. We see that the TCP downlink throughput values range from as low as 972 Kbps to as high as 2,970 Kbps, which is an improvement of 2.6 to 8.0 times compared to the maximum single-user TCP downlink throughput of 370 Kbps achieved in the 3G network. In fact, the maximum TCP downlink throughput of 2,970 Kbps achieved by a single user in operator Y's HSDPA network is about 1.75 times higher than the maximum aggregate TCP downlink throughput of 1.7 Mbps obtained in the 3G network of operator Y, as shown previously in Fig. 7a. Similarly, the TCP uplink throughput results in the HSDPA network of operator Y, with a maximum of 365 Kbps, is more than six times higher than the maximum single-user TCP uplink throughput of 60 Kbps achieved in the 3G network of operator Y.

We have also performed the ping test (using the same TCPing tool that was used in Section 4.2) to several local Web sites and found that the achieved RTT values alternate between 60-125 ms and 2,000-6,000 ms. We also observed the data connection indicator on our USB modem automatically toggling between the "3G" and "HSDPA" modes during the ping test. This observation suggests that if there is not enough data passing through the connection, the

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

network will downgrade the HSDPA connection to a 3G connection, and vice versa. Speculating that this "switching" observation is due to the sleep feature of TCPing when the RTT value is less than 2,000 ms (refer to Section 4.2), we decided to investigate further by removing the sleep function from TCPing and repeating the ping tests. By doing so, we found that the achieved RTT values to local Web sites are now consistently in the range of 60-125 ms, with an average of 81 ms, and we also observed that the data connection indicator on the USB modem is consistently set to the "HSDPA" mode. In addition, these RTT values also correspond with the reported RTT values in [16].

The RTT observations with the HSDPA network prompted us to examine the effect of the sleep function in TCPing on the RTT values achieved over the 3G network. Using the 3G data card of operator Y, we repeated the ping tests to local Web sites by using the versions of TCPing with and without the sleep function. The results show that with TCPing-Sleep, the achieved RTT values are within the range of 125-969 ms, with an average of 175 ms, whereas with TCPing-NoSleep, the achieved RTT values are in the range of 125-953 ms, with an average of 164 ms. This observation indicates that the sleep function of TCPing has no effect on the achievable 3G RTT values to local Web sites. We also see that the minimum RTT values are similar to those reported in Section 4.2 for operator Y. On the other hand, due to the smaller sample size in our new ping tests, the average RTT values are lower compared to those reported in Section 4.2.

For a comparison of the network processing and queuing delays in the HSDPA and 3G networks of operator Y, we note that the round-trip transmission time of a ping packet over the wireless link in an HSDPA network is only about 2.2 ms (50 bytes/2,000 Kbps + 50 bytes/200 Kbps). Therefore, this shows that the network processing and queuing delays in the HSDPA network ($\sim$60 ms at a minimum) is lower than that of the 3G network ($>$ 100 ms at a minimum).

## 5 Conclusion

We have designed and performed a large-scale empirical study on the performance of three commercial 3G networks in Hong Kong in terms of their data, video, and voice capacities. The conclusions of this work are based on the measurement results from real operational networks, albeit with the caveat that the performance can change with different network settings and different vendor equipment and when operators perform software upgrades to their networks. Nevertheless, our results provide researchers with significant real-performance results from three different operational networks. Our key findings include the following:

- Substantial additional processing and queuing latencies for data services ($>$ 100 ms at a minimum) are imposed by the 3G network when compared to wireline IP networks, even under lightly loaded scenarios.
- The average latency for 3G data services increases to beyond 1 second in fully loaded network conditions.
- The minimum RTT values achievable in the HSDPA network ($\sim$60 ms) is about 50 percent of that achievable in the 3G network ($\sim$125 ms) of the same operator. Similarly, the network processing and queuing latency in the HSDPA network is also about 50 percent lower than that of the 3G network.
- There is a maximum eightfold increase in the TCP downlink individual throughput achieved in the HSDPA network (2,970 Kbps) compared to that achieved in the 3G network (370 Kbps) of the same operator.
- There is a maximum sixfold increase in the TCP uplink individual throughput achieved in the HSDPA network (365 Kbps) compared to that achieved in the 3G network (60 Kbps) of the same operator.
- The TCP retransmission probability over the wireless link varies substantially across different operators (from $\ll$ 1 percent to 1$\sim$6 percent), which points to the different operational configurations across different operators in terms of the configured RLC data transfer mode and the threshold value for the targeted signal-to-noise ratio in the 3G networks.
- Our HSDPA measurement results pointed to the existence of a TCP PEP [17] in the HSDPA network.
- We have found that 3G networks operating under near-saturated conditions exhibit highly unpredictable behavior, for example difficulties (requiring repeated attempts) in adding supposedly higher priority video/voice calls into a data saturated network, with a substantial number of observed cases in which not even one video/voice call could be added, and multiple data calls (especially uplink data calls) are being dropped unexpectedly.
- Our 3G measurement results suggest the existence of a guaranteed minimum bandwidth resource for the data services when data calls are present in the network. The results also indicate that this guaranteed data bandwidth threshold value is different across different operators and measurement sites. Such findings further demonstrate the diverse nature of the network resources allocation mechanisms and the call admission control policies employed by different operators.
- The multiservice capacity region of a cell can be conservatively estimated by a plane with axis-intercept values equal to the maximum capacity for the corresponding single-service types, that is, the maximum data bandwidth and the maximum number of video/voice calls.
- The 3G network operators seem to have extensively customized their network configurations in a cell-by-cell manner in order to cope with the projected traffic demand of individual sites and neighborhood, and the target coverage area of the base station. As such, the cell capacity varies widely not only across different operators but also across different measurement sites of the same operator.
- It is practically impossible to predict the actual capacity of a particular cell based on known theoretical models and standard parameters, even when supplemented by key field measurements such as the received signal-to-noise ratio $E_c/N_0$. In other words, it is practically impossible to derive the necessary model parameters, which can provide a good fit between the theoretical prediction and the measured capacity values.

Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.

To the best of our knowledge, this is the first public report on a large-scale empirical study on the capacities and performance of commercial 3G networks carrying live data, video, and voice traffic. Our results are useful to both network planners in planning enough capacity to support new data applications and to application designers in designing flow-control algorithms to handle the bandwidth variability as the video/voice calls arrive and terminate. Our future work will be focused on the investigation of the QoS of voice and video calls (as perceived by the end-user) under varying load in the network.

## ACKNOWLEDGMENTS

The authors would like to thank OnChing Yue for his insightful comments and suggestions for this work. Thanks are also due to Eric Wong, Andy Mak, Daisy Lee, Wilson Ip, Hui Ka Hung, Gina Yuen, and Jack Chui for their assistance during the measurements and processing of the results. This work has been partially supported by OFTA and by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No: CUHK 4126/06E).

## REFERENCES

[1] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third-Generation Mobile Communications,* third ed. John Wiley & Sons, 2004.

[2] *High-Speed Downlink Packet Access (HSDPA) Overall Description Stage 2,* TS 25.308 V 5.6.0, 3GPP, Sept. 2004.

[3] K.S. Gilhousen et al., "On the Capacity of a Cellular CDMA System," *IEEE Trans. Vehicular Technology,* vol. 40, pp. 303-312, May 1991.

[4] V.K. Paulrajan, J.A. Roberts, and D.L. Machamer, "Capacity of a CDMA Cellular System with Variable User Data Rates," *Proc. IEEE Global Telecomm. Conf (GLOBECOM),* 1996.

[5] J.-R. Yang, Y.-Y. Choi, J.-H. Ahn, and K. Kim, "Capacity Plane of CDMA System for Multimedia Traffic," *Electronics Letters,* vol. 33, no. 17, pp. 1432-1433, 1997.

[6] I.-S. Khoo, J-H. Ahn, J-A. Lee, and K. Kim, "Analysis of Erlang Capacity for the Multimedia DS-CDMA Systems," *IEICE Trans. Fundamentals,* vol. 82-A, no. 5, pp. 849-855, May 1999.

[7] A.M. Viterbi and A.J. Viterbi, "Erlang Capacity of a Power Controlled CDMA System," *IEEE J. Selected Areas in Comm.,* vol. 11, pp. 892-900, 1993.

[8] J.S. Evans and D. Everitt, "On the Teletraffic Capacity of CDMA Cellular Networks," *IEEE Trans. Vehicular Technology,* vol. 48, pp. 153-165, Jan. 1999.

[9] K. Pentikousis, M. Palola, M. Jurvansuu, and P. Perälä, "Active Goodput Measurements from a Public 3G/UMTS Network," *IEEE Comm. Letters,* vol. 9, pp. 802-804, Sept. 2005.

[10] P. Reichl, M. Umlauft, J. Fabini, R. Lauster, and G. Pospischil, "Project WISQY: A Measurement-Based End-to-End Application-Level Performance Comparison of 2.5G and 3G Networks," *Proc. Fourth Ann. Wireless Telecomm. Symp. (FTS),* 2005.

[11] M. Kohlwes, J. Riihijärvi, and P. Mähönen, "Measurements of TCP Performance over UMTS Networks in Near-Ideal Conditions," *Proc. 61st IEEE Vehicular Technology Conf. (VTC Spring),* 2005.

[12] J.M. Cano-Garcia, E. Gonzalez-Parada, and E. Casilari, "Experimental Analysis and Characterization of Packet Delay in UMTS Networks," *Proc. Sixth Int'l Conf. Next-Generation Teletraffic and Wired/Wireless Advanced Networking (NEW2AN),* 2006.

[13] A. Mäder, B. Wagner, T. Hossfield, D. Staehle, and H. Barth, "Measurements in a Laboratory UMTS Network with Time-Varying Loads and Different Admission Control Strategies," *Proc. Fourth Int'l Workshop Internet Performance, Simulation, Monitoring and Measurement (IPS-MoMe),* 2006.

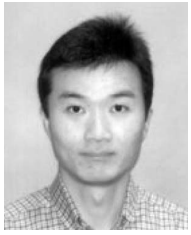
[14] R.M. Joyce, B. Graves, T. Griparis, I.J. Osborne, and T.M. Lee, "Case Study: The Capacity of a WCDMA Network—Orange, UK," *Proc. Fifth IEE Int'l Conf. 3G Mobile Comm. Technologies (3G),* 2004.

[15] V.V. Veeravalli and A. Sendonaris, "The Coverage-Capacity Tradeoff in Cellular CDMA Systems," *IEEE Trans. Vehicular Technology,* vol. 48, pp. 1443-1450, Sept. 1999.

[16] H. Holma and J. Reunanen, "3GPP Release 5 HSDPA Measurements," *Proc. 17th IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm. (PIMRC),* 2006.

[17] J. Border, M. Kojo, J. Griner, G. Montenegro, and Z. Shelby, *Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations,* IETF RFC 3135, June 2001.

[18] W.L. Tan, F. Lam, and W.C. Lau, "An Empirical Study on 3G Network Capacity and Performance," *Proc. IEEE INFOCOM,* 2007.



**Wee Lum Tan** received the BEng degree in electrical engineering from the University of Malaya and the MEngSc degree from the Multimedia University in 1997 and 1999, respectively. He is currently working toward the PhD degree in the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include the experimental measurements, system modeling, and performance analysis of wireless data networks. He is a student member of the IEEE.

**Fung Lam** received the BEng and MPhil degrees from the University of Hong Kong. He is a senior software architect with professional experiences in database technologies, architectural design, software engineering, system development, and IT consultancy. He is actively participating in projects for global financial institutions, logistic companies, and government departments. He was with the Department of Information Engineering, The Chinese University of Hong Kong when this work was done. His research interests include routing algorithms, traffic engineering, performance analysis, system modeling, and wireless networks.

**Wing Cheong Lau** received the BSEng degree from the University of Hong Kong and the MS and PhD degrees in electrical and computer engineering from the University of Texas, Austin. He is currently an associate professor in the Department of Information Engineering, The Chinese University of Hong Kong (CUHK), Shatin, where he is also the director of the Mobile Technologies Center (MobiTeC). From 1997 to 2004, he was with the Performance Analysis Department, Bell Laboratories, Lucent Technologies, Holmdel, New Jersey, as a performance consultant and system architect. Prior to joining CUHK, he was with Qualcomm, San Diego, actively contributing to the design and standardization of IETF and 3G mobility management protocols and architecture. His research interest includes networking protocol design and performance analysis, traffic characterization, system modeling, and network security for high-speed wired and wireless networks. He is a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

View publication stats Authorized licensed use limited to: Univ of Calif Los Angeles. Downloaded on November 9, 2009 at 10:38 from IEEE Xplore. Restrictions apply.