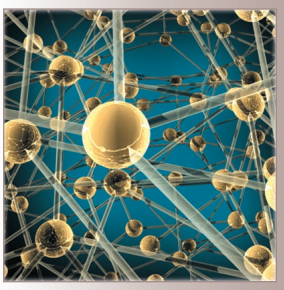

IPsec-Based VoIP Performance in WLAN Environments

The Third-Generation Partnership Project (3GPP) specifies 3G-WLAN interworking that lets mobile stations (MSs) access 3G networks via WLAN-based packet data gateways (PDGs). The specification mandates that packets delivered between an MS and a PDG should be protected by IPsec. This article studies IPsec-based VoIP performance (in terms of throughput, packet loss rate, latency, and jitter) in the 3G-WLAN integration environment. The study is designed to provide guidelines for selecting appropriate system parameter values for VoIP over WLAN.

Ya-Chin Sung and Yi-Bing Lin
National Chiao Tung University, Taiwan

The Third-Generation Partnership Project's technical specification 23.234 defines third-generation mobile telecommunications system and wireless LAN (3G-WLAN) interworking, which extends 3G services to the WLAN environment.[1] A mobile station (MS) in the WLAN accesses the 3G core network through a packet data gateway (PDG), where the security requirements are enforced via an IPsec association between the two, and transmitted packets are protected by IPsec with Encapsulating Security Payload (ESP) in the tunnel mode.[2] In an IPsec tunnel, the tunnel endpoints encrypt and authenticate IP packets, including headers and payloads, and add new IP headers to route the packets between the MS and the PDG. Before sending an IP packet, the MS checks the security policies applied to the packet and performs IPsec encapsulation according to the methods defined by the security association. When receiving an IPsec packet, the PDG validates and decapsulates the packet according to the corresponding security association. Exercising IPsec encapsulation increases the size of the packets transmitted between the MS and the PDG, thus degrading performance.

The 3G-WLAN integration environment supports Voice over IP (VoIP) to provide voice communications over the Internet.[3] VoIP services often use the Session Initiation Protocol (SIP)[4] to control calls and the Real-time Trans-

1089-7801/08/$25.00 © 2008 IEEE

Table 1. Codec attributes.

Codec	Bit rate	Sample period	RTP payload length (sample rate x sample period)	RTP packet rate
G.711	64 Kbps, sampling at an 8 KHz rate with 8 bits per sample	20 ms	160 bytes, one frame per RTP packet	50 packets/sec
G.729	8 Kbps, sampling at a 1 KHz rate with 8 bits per sample	10 ms	2 × 10 bytes, two frames are combined into one RTP packet	50 packets/sec

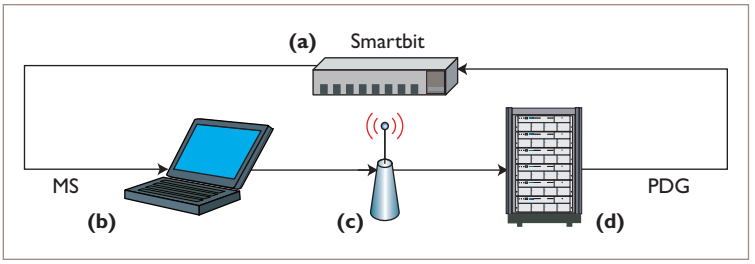

Figure 1. Experimental 3G-WLAN integration system. (a) The Smartbit generates multiple RTP streams. The RTP packets are transmitted through (c) an IPsec tunnel from (b) the MS to (d) the PDG, and then back to the Smartbit. Finally, the Smartbit collects the performance statistics.

port Protocol (RTP)[5] to deliver voice data. Several codecs can be used in RTP calls to meet the bandwidth restrictions. In the 3G-WLAN integration environment, VoIP performance might be degraded due to IPsec encapsulation. In this article, we describe our experimental results investigating IPsec-based VoIP performance between an MS and the PDG.

IPsec-Based VoIP Experimental Environment

Figure 1 illustrates the simplified 3G-WLAN integration system for our experiments. The MS (see Figure 1b) is a laptop equipped with a Pentium M 1.3-GHz CPU and 256 Mbytes of memory. The PDG (see Figure 1d) is a laboratory prototype implemented in a PC equipped with a Pentium IV 3.0-GHz CPU and 1 Gbyte of memory. As Figure 1c shows, the MS communicates with the PDG via a D-Link DL-524 IEEE 802.11b access point (AP), which connects to the PDG through an Ethernet cable with a peak rate of 10 Mbps. The Smartbit (see Figure 1a) measures the RTP performance, which is connected to the MS and the PDG using CAT 5 cables with an RJ-45 interface.[6]

In our experiments, the Smartbit generates multiple RTP streams identified with different source–destination IP address pairs and injects them to the MS. The RTP packets are transmitted over UDP. In generating the RTP streams, we use two kinds of voice codecs (Table 1 summarizes the codec attributes):

- G.711[7] is a high bit-rate codec with a sample period of 20 ms.
- G.729[8] is a low bit-rate codec with a sample period of 10 ms.

When the MS receives the RTP packets from the Smartbit, it encapsulates them with IPsec Encapsulating Security Payload (ESP) in the tunnel mode. The RTP packets are encrypted using the triple Data Encryption Standard (3DES) algorithm and authenticated via the HMAC-SHA-1-96 algorithm.[9,10] The MS sends the encapsulated packets to the PDG via the IPsec tunnel. On receiving the IPsec packets, the PDG executes the IPsec decryption procedure. The Smartbit then collects the decrypted RTP packets and produces the output statistics from the measured packets.

Performance Measurement

Based on the experimental environment described in the previous section, we measure the IPsec overhead in terms of throughput, packet loss rate, latency, and jitter.

Throughput and Packet Loss Rate

Figure 2a illustrates the packet loss rate as measured by the Smartbit. Based on the equation derived by David P. Hole and Fouad A. Tobagi[11] and Wei Wang and his colleagues,[12] we compute the theoretic upper bound of VoIP capacity (in terms of the number of RTP streams) over IEEE 802.11b without packet loss.

As Table 2 shows, our measured capacity without packet loss achieves about 85 percent of the theoretic upper bound capacity. We also found that, after IPsec encryption, the capacities without packet loss degrade by 4 percent for G.729 and 5 percent for G.711.

Rajadurai Rajavelsamy and his colleagues showed that the IEEE 802.11b AP can support

Table 2. Network capacities without packet loss.

Codec	IPsec encrypted	Theoretic upper bound capacity (number of RTP streams)	Measured capacity (number of RTP streams)
G.729	No	24.1	20
	Yes	23.2	19
G.711	No	21.5	18
	Yes	20.7	17

28 IPsec VoIP connections for G.711 with a packet loss rate of less than 1 percent, but the result seems misleading because the reported number of VoIP connections (28) exceeds the theoretic upper bound capacity.[11,12] Our experiments show the consistent results with the theoretic upper bound capacity,[11,12] which indicates that when the packet loss rate is less than 1 percent, an IEEE 802.11b AP can support only 20 and 17 IPsec VoIP connections for G.711, respectively.

To maintain a given packet loss rate, the system can support one IPsec RTP stream less than the original RTP streams (see Figure 2a). For example, with a packet loss rate of 10 percent, the system can support 21.86 original RTP streams or 21.13 IPsec RTP streams for G.729; alternatively, it can support 20.02 original RTP streams or 19.15 IPsec RTP streams for G.711. That is, the IPsec overhead is 3.34 percent for G.729, and the IPsec overhead is 4.35 percent for G.711.

As the system attempts to support more RTP streams and the packet loss rate increases, the IPsec overhead becomes less significant (see Figure 2a). For example, when the packet loss rate increases from 5 to 20 percent, IPsec overhead decreases from 3.80 percent to 3.29 percent for G.729, and from 5.12 percent to 3.55 percent for G.711.

Based on the mathematical analysis in the earlier studies,[11,12] we can calculate the capacity of an IEEE 802.11b AP for IPsec VoIP. The calculation indicates that the packet loss rate increases to 5 and 6.3 percent (for G.729 and G.711, respectively) when the VoIP traffic is one RTP stream larger than the network capacity without packet loss (see Figure 2a).

Figure 2b illustrates the throughput performance. We note that the following relationship holds:

$$\text{Packet Loss Rate} = \frac{\text{ArrivalRate} \times \text{PacketSize} - \text{Throughput}}{\text{ArrivalRate} \times \text{PacketSize}}.$$

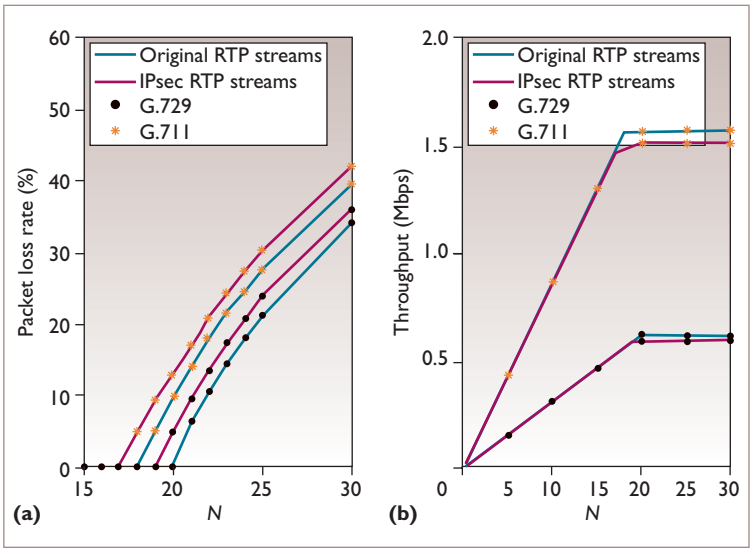

Figure 2. Effects of the number N of RTP streams on (a) packet loss and (b) throughput.

In our experiments, the arrival rate is 50 packets/sec times the number of RTP streams. This figure indicates that the system saturates if we pump more than 20 original RTP streams or 19 IPsec RTP streams for G.729. For G.711, the system saturates if we pump more than 18 original RTP streams or 17 IPsec RTP streams. By exercising IPsec, the maximum throughput for the system degrades by 5 percent for G.729, and 5.56 percent for G.711. When the system isn't saturated, the throughput for supporting both original and IPsec RTP streams is the same.

Latency

Packet processing, delivery, and loss affect the latency performance. Packet processing and delivery both contribute to queuing and thus increase the latency. During packet delivery, the MS might have to retransmit packets because of transmission errors or collisions (that is, radio link congestion). In IEEE 802.11b, a packet is transmitted after a back-off delay. For each retransmission, the average back-off delay doubles. The 802.11b MAC discards the packet after

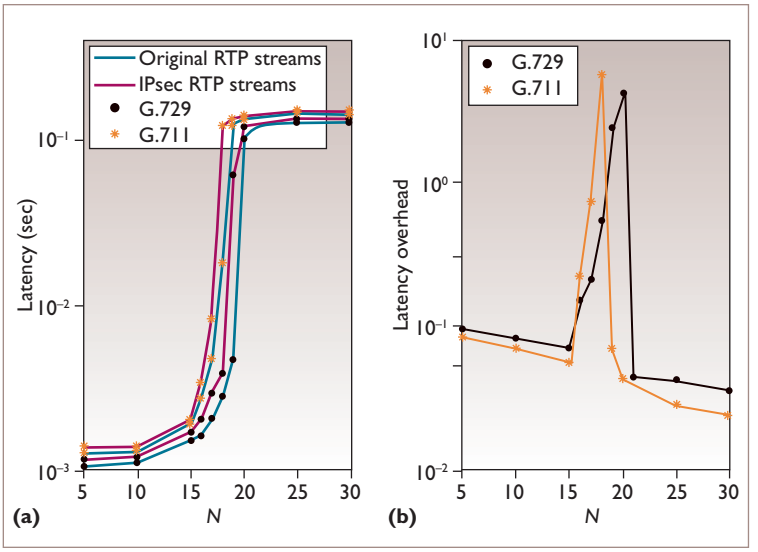

Figure 3. Effects of the number N of RTP streams on (a) the mean latency and (b) latency overhead.

four retransmissions. Packet loss mitigates the queuing effect and therefore stops the latency increase caused by packet retransmission.

Figure 3a shows that the mean latency is an S-shape increasing function of the number N of RTP streams. The S-shape curves are explained in the following three cases.

Case I. When $N < 10$, increasing the number of streams insignificantly affects the latency. In this case, there is no packet loss, and packet retransmission seldom occurs. The latency results from the queuing effect of packet processing. For example, when N increases from 5 to 10, the latency increases from 1.17 ms to 1.30 ms (11.11 percent) for G.729, and from 1.39 ms to 1.41 ms (1.44 percent) for G.711.

Case II. When $10 \le N \le 20$, increasing N significantly increases the latency because of packet processing and packet retransmission. For example, when N grows from 15 to 20, the latency increases from 1.72 ms to 101.22 ms for G.729, and from 2.10 ms to 139.38 ms for G.711.

Case III. When $N > 20$, increasing N only slightly increases the latency due to packet loss. As Figure 2a shows, packet loss significantly increases as N increases for $N \ge 20$. When N increases from 25 to 30, the latency insignificantly increases from 132.99 ms to 134.95 ms (that is, 1.47 percent) for G.729, and from 147.82 ms to 149.23 ms (0.95 percent) for G.711.

The latency performance reported by Rajavelsamy and his colleagues showed the same trend as our results for cases I and II; they didn't investigate case III in their report.[13]

Because the packet size is larger for G.711 than for G.729, the packet processing time is longer as well (see Figure 3a). Similarly, the latencies for IPsec RTP streams are longer than for original streams. Specifically, we can calculate the latency overhead for IPsec as

$$\text{Latency Overhead} = \frac{\text{IPsecRTPLatency} - \text{OriginalRTPLatency}}{\text{OriginalRTPLatency}}.$$

When $N < 15$, the IPsec overhead is less than 9.26 percent due to the insignificant queuing effect that packet processing causes (see Figure 3b). When $15 \le N \le 20$, the latency overhead for IPsec significantly increases — it can be up to 570.97 percent. In this case, IPsec streams experience heavy packet retransmission compared to the original streams. For $N > 20$, the system saturates for both IPsec and the original streams, thus dropping many packets for both as they reach the retransmission limit. The resulting latency overhead drop for IPsec is less than 4.38 percent.

Jitters

Jitter, or the variation of packet inter-arrival time, can create unexpected pauses between utterances, thus affecting VoIP speech intelligibility. A previous study showed that VoIP quality of service (QoS) becomes unacceptable when the average jitter exceeds 35 ms.[6] To reduce jitter, the receiver uses a buffer to store incoming packets before they're played. If the jitter buffer size is too small, network jitter will lead to packet loss and degraded voice intelligibility. If the jitter buffer is too large, packet delays will be lengthy and QoS is degraded. For example, the echo level is easier to perceive with large jitter buffers. NTT Communications specifies that the average jitter should be no more than 0.5 ms for VoIP.[14]

Without the buffer, jitter is an S-shape increasing function of the number N of RTP streams in our experiments (see Figure 4). When $N < 5$, the RTP packets experience fewer link congestions, and the average jitter is less than 0.5 ms. When $5 \le N \le 25$, the average jitter goes up significantly as N increases. When $N > 25$, the system saturates, and the average

jitter increases to about 10 ms. To maintain the same average jitter, the system supports about one less IPsec RTP stream than original RTP streams. For example, to limit the average jitter to 1 ms, the system can't support more than 17.77 original RTP streams or 16.75 IPsec RTP streams for G.729 (that is, the IPsec overhead is 5.74 percent). For G.711, the system can't support more than 15.88 original RTP streams or 15.39 IPsec RTP streams (that is, the IPsec overhead is 9.79 percent).

As Figure 4 indicates, the jitters for the G.711 RTP streams are larger than for G.729 RTP streams. G.711's larger packet size causes more link congestion than G.729.

In our experiments without jitter buffers, the average network jitter (between the MS and the PDG) ranges from 0.44 ms to 14.55 ms. Thus, to eliminate WLAN-caused jitter effects, at least one G.711 or G.729 RTP packet (that is 20 ms) should be buffered to achieve the jitter performance specified by NTT Communications.[14] Note that IPsec overhead doesn't affect the jitter buffer size in this case.

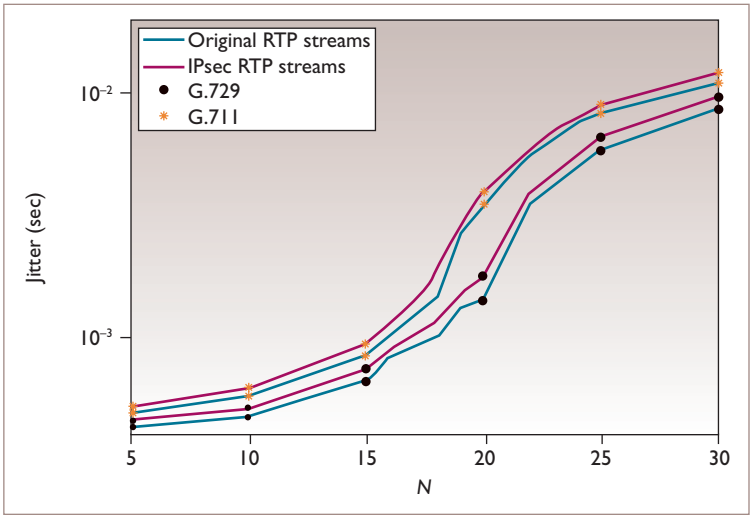

Figure 4. Effects of the number N of RTP streams on jitter without jitter buffer.

Our study provides guidelines for selecting appropriate system parameter setups for VoIP service in the 3G-WLAN integration environment. Specifically, an IEEE 802.11b access point can support 15 IPsec RTP streams with acceptable latency, small jitter, and no packet loss. Our study also indicates that the IPsec overhead is not serious. To maintain the same packet loss rate and jitter, the system will support one less IPsec RTP stream than original RTP streams.

Acknowledgments

Y.-B. Lin's work was supported in part by NSC-96-2219-E-009-019, NSC-97-2221-E-009-143-MY3, Chung-Hwa Telecom, Industrial Technology Research Institute and National Chiao Tun University joint research center, and the Ministry of Education Aim for the Top University and Elite Research Center plan.

References

1. *3GPP TS 23.234, 3GPP System to Wireless Local Area Network (WLAN) Interworking; System Description* (release 7), 3rd Generation Partnership Project (3GPP), 2006; www.3gpp.org/ftp/Specs/archive/23_series/23.234.
2. S. Kent and R. Atkinson, *IP Encapsulating Security Payload (ESP)*, IETF RFC 2406, Nov. 1998; www.ietf.org/rfc/rfc2406.txt.
3. Y.-B. Lin and A.-C. Pang, *Wireless and Mobile All-IP Networks*, Wiley, 2005.
4. J. Rosenberg et al., *SIP: Session Initiation Protocol*, IETF RFC 3261, June 2002; www.ietf.org/rfc/rfc3261.txt.
5. S. Casner et al., *RTP: A Transport Protocol for Real-Time Applications*, IETF RFC 3550, July 2003; www.ietf.org/rfc/rfc3550.txt.
6. Spirent Communications, "SmartVoIPQoS User Guide," 2001, www.spirentcom.com/documents/438.pdf.
7. *ITU-T G.711 Recommendation, Pulse Code Modulation (PCM) of Voice Frequencies*, Int'l Telecommunication Union, Nov. 1988; www.itu.int/rec/T-REC-G/recommendation.asp?lang=en&parent=T-REC-G.711.
8. *ITU-T G.729 Recommendation, Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, Int'l Telecommunication Union, Mar. 1996; www.itu.int/rec/T-REC-G/recommendation.asp?lang=en&parent=T-REC-G.729.
9. *ANSI X9.52-1998, Triple Data Encryption Algorithm Modes of Operation*, Am. Nat'l Standard Inst., 1998.
10. *FIPS 180-1, Secure Hash Standard*, Nat'l Inst. of Standards and Technology, Springer-Verlag, 1995; www.itl.nist.gov/fipspubs/fip180-1.htm.
11. D.P. Hole and F.A. Tobagi, "Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP," *Proc. IEEE Int'l Conf. Comm.*, vol. 1, 2004, pp. 196–201.
12. W. Wang, S.-C. Liew, and V.O.K. Li, "Solutions to Performance Problems in VoIP over an 802.11 Wireless LAN," *IEEE Trans. Vehicular Technology*, vol. 54, no. 1, 2005, pp. 366–384.
13. R. Rajavelsamy et al., "Performance Evaluation of VoIP over 3G-WLAN Interworking System," *IEEE Wireless Comm. and Networking Conf.*, vol. 4, 2005, pp. 2312–2317.
14. NTT Communications, "NTT Com Adds Jitter SLA for

Related Work in VoIP Performance

Numerous researchers have investigated IPsec performance for voice over IP (VoIP) in wireless environments. Arlen Nascimento and his colleagues studied the latency performance for IPsec-based VoIP by evaluating speech quality with the G.711 codec.[1] Specifically, they evaluated speech quality using a subjective method called mean opinion score (MOS), which expresses the speech quality as a score ranging from one to five, where one represents the lowest perceived quality, and five represents the highest perceived quality. Rajadurai Rajavelsamy and his colleagues measured the IPsec-based VoIP performance over 3G-WLAN integration systems, which provide 3G services to users in a WLAN environment.[2] Their study measured VoIP performance using different types of codecs with real experiments instead of mathematical analysis. David Hole and Fouad Tobagi[3] used mathematical analysis and simulation experiments to study VoIP performance in terms of network capacity (that is, the number of supported RTP streams).

In the preceding experiments, the performance measurement tools were run on the mobile station (MS), which might affect the accuracy of the reported results. The QoS measurement tool developed by Nascimento and his colleagues was executed on the MS.[1] Rajavelsamy and his colleagues conducted a study in which four tools were executed on the MS, including RTP tools to send and receive RTP packets, network monitor tools pktstat and Netperf to measure the network traffic and collect the performance data, and Ethereal to record network packet events on the MS.[2] In our experiments, all performance data are collected by Smartbit. The MS only processes the VoIP packets as in the normal operation mode, and its computing power isn't consumed for measurement.

The performance results presented in the previous studies are quite different because they use different experimental setups and define different output measures. The MOS measure considered by Nascimento and his colleagues provides useful insight for voice quality.[1] However, it doesn't reflect the effect of delays. Also, that study didn't conduct IPsec performance in terms of packet loss and jitter. The analytical studies by Hole and Tobagi[3] and Wei Wang and his colleagues[4] showed the IEEE 802.11b access point (AP) capacity for plain VoIP without packet loss. They didn't consider the relationships between throughput and VoIP traffic load. The performance results in Rajavelsamy's study[2] are inconsistent with those by Hole and Tobagi or Wang and his colleagues.[3,4]

No previous studies considered the heavy VoIP traffic issues when the MS is engaged with more than 28 RTP streams with the AP. Our study on heavy traffic provides useful insight to assist VoIP operators in determining what kinds of codec and packet loss concealment techniques to employ. Unlike the others, we also elaborate on IPsec overhead in terms of latency and compare the jitter performance for VoIP with and without IPsec.

References

1. A. Nascimento et al., "Can I Add a Secure VoIP Call?" *Proc. 13th IEEE Int'l Conf. Networks*, vol. 1, 2005, pp. 151–155.
2. R. Rajavelsamy et al., "Performance Evaluation of VoIP over 3G-WLAN Interworking System," *Proc. IEEE Wireless Comm. and Networking Conf.* (WCNC), IEEE CS Press, vol. 4, 2005, pp. 2312–2317.
3. D.P. Hole and F.A. Tobagi, "Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP," *Proc. IEEE Int'l Conf. Comm.*, IEEE CS Press, vol. 1, 2004, pp. 196–201.
4. W. Wang, S.-C. Liew, and V.O.K. Li, "Solutions to Performance Problems in VoIP over a 802.11 Wireless LAN," *IEEE Trans. Vehicular Technology*, vol. 54, no. 1, 2005, pp. 366–384.

Global IP Network Services," Mar. 2005, www.ntt.com/release_e/news05/0003/0329.html.

Ya-Chin Sung is a PhD student in the Department of Computer Science at National Chiao Tung University (NCTU), Hsinchu, Taiwan. Her current research interests include design and analysis of personal communications services networks, network security, and performance modeling. Sung has a BS and an MS in computer science and information engineering from NCTU. Contact her at ycsung@csie.nctu.edu.tw.

Yi-Bing Lin is chair professor and dean of the College of Computer Science at National Chiao Tung University. His current research interests include wireless communications and mobile computing. Lin has an EEBS degree from National Cheng Kung University and a PhD from the Department of Computer Science and Engineering at the University of Washington. He has published more than 200 journal articles, 200 conference papers, and is coauthor of *Wireless and Mobile Network Architecture* (Wiley, 2001), *Wireless and Mobile All-IP Networks* (Wiley, 2005), and *Charging for Mobile All-IP Telecommunications* (Wiley, 2008). He is a fellow of the ACM, the American Association for the Advancement of Science (AAAS), the IET, and the IEEE. Contact him at liny@csie.nctu.edu.tw.