WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2014; **14**:450–459
Published online 6 February 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/wcm.2194

RESEARCH ARTICLE

Deriving the distributions for the numbers of short message arrivals

Hui-Nien Hung[1], Yi-Bing Lin[2*] and Chao-Liang Luo[2,3]

[1] Institute of Statistics National Chiao Tung University, Hsinchu, Taiwan
[2] Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
[3] Telecommunication Laboratories Chunghwa TelcomCo., Ltd., Taiwan

ABSTRACT

In the broadband era, narrowband short message service (SMS) is still the most popular wireless data service. Many studies have been conducted to investigate the performance of SMS based on the arrival rates of short messages. From Chunghwa Telecom's commercial SMS call data records, we observed that even if the SMS arrival rates are the same, the distributions for the number of SMS arrivals per half hour are quite different for various observed days. We further identify that for the SMS traffic in a specific day, there are non-burst and burst periods. This paper investigates the SMS behaviors on weekdays, weekends, and holidays (specifically, new years' days and eves). With the assistance of kernel-based fitting method, we derive the SMS arrival number distributions of various traffic types and observed days. Our approach fits each SMS arrival number distribution by three cubic polynomial functions that can accurately capture the SMS behaviors. On the basis of the SMS arrival number distributions derived from our model, the mobile operators have better understanding about the volumes of short messages in different times and days, which can be used to design more flexible short message charging rates. Copyright © 2012 John Wiley & Sons, Ltd.

KEYWORDS

arrival distribution; kernel-based fitting; mobile telecommunications network; short message service (SMS)

***Correspondence**

Yi-Bing Lin, Department of Computer Science and Information Engineering, National Chiao Tung University.
E-mail: liny@cs.nctu.edu.tw

1. INTRODUCTION

Short message service (SMS) contributes about 60% of the mobile data service revenue today [1]. This statistic indicates that even in the broadband area, narrowband SMS is still the most popular wireless data service. Many business applications, such as stock service, personal identification verification service, weather casting service or daily news service, can be delivered to the customers through SMS [2–5]. Figure 1 shows the SMS architecture for universal mobile telecommunications system (UMTS) [6–9]. In this architecture, a short message sent to a user equipment (UE, Figure 1(c)) can be originated from another UE (Figure 1(a)), where the short message is first sent to the short message-service center (SM-SC, Figure 1(g)) through the originating UMTS terrestrial radio access network (Figure 1(d)), the mobile-originating mobile switching center (Figure 1(e)) and the inter-working mobile switching center (Figure 1(f)). Upon receipt of a short message, the SM-SC sends the message to the gateway MSC (GMSC, Figure 1(h)). The GMSC interrogates the home subscriber server (Figure 1(i)) to identify the mobile terminating MSC (Figure 1 (j)) of the recipient and forwards the message to the mobile terminating MSC. Finally, the short message is delivered to the terminating UE (Figure 1(c)) via the terminating UMTS terrestrial radio access network (Figure 1(k)). In Chunghwa Telecom, (CHT, the largest telecom operator in Taiwan), the SM-SC and SMS-GMSC are collocated.

The short message can also be sent from an external application (typically on the Internet; see Figure 1(l)) to the SM-SC through the external short message entity (Figure 1(b)) by using short message peer-to-peer protocol [6]. In Chunghwa Telecom, 30.48% of the short messages are originated from the external short message entity.

When the SM-SC receives a short message, a call data record (CDR) is created for billing and other administration purposes. The CDR records the SMS arrival time information, which is important for SMS traffic engineering. For example, in the work of Petros *et al.*, the SMS

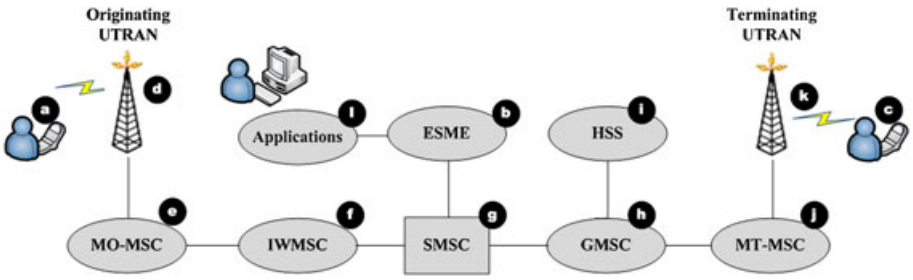


Figure 1. UMTS-based SMS architecture.

arrival rate was used for modeling SMS reliability to investigate a significant risk for emergency alerts in New Year during 2005 [10]. Sou *et al*. investigated the SMS resending traffic through the SMS arrival rate [11]. Wu *et al*. proposed a new mechanism for real-time content monitoring and filtering through SMS arrival [12]. Although these studies provide very useful insights into the SMS behaviors, conclusions of these studies could be strengthened if the SMS arrivals were described with more accurate distributions for the numbers of SMS arrivals instead of their means (i.e., arrival rates). For the description purpose, we use the short term 'arrival distribution' to represent 'distribution for the number of SMS arrivals'.

In this paper, the collection data from Chunghwa Telecom's SMSC are more than 10 million short message generated from 100,000 users between 2007 and 2010. We show that the SMS arrival behaviors are significantly different for various traffic types and observed days. Then, with the assistance of the kernel-based fitting method [13], we derive eight types of SMS arrival distributions. The paper is organized as follows. Section 2 describes histograms for the number of SMS arrivals from a macro view. Section 3 presents the fitting model to derive the SMS arrival distributions. Finally, Section 4 concludes this study and outlines the future work.

2. SHORT MESSAGE SERVICE ARRIVAL BEHAVIOR: A PRELIMINARY VIEW

This section presents the SMS arrival histograms based on more than 10 million SMS CDRs created in different time periods. Let $N_{F,d}(T)$ be the average number of SMS arrivals during the 30-min period $[T, T + 30\,\text{min})$ on the dth day of February 2010. Let S_1 be the set of the weekdays in February 2010, and S_1^* be the set of the weekends in February 2010. Denote $N_F(T)$ and $N_{F*}(T)$ as

$$N_F(T) = \frac{\sum_{d\in S_1} N_{F,d}(T)}{|S_1|} \text{ and}$$
$$N_{F*}(T) = \frac{\sum_{d\in S_1^*} N_{F,d}(T)}{|S_1^*|} \quad (1)$$

That is, $N_F(T)$ is the average number of SMS arrivals during $[T, T + 30\,\text{min})$ for a weekday in February 2010, and $N_{F*}(T)$ is that for a weekend.

Figure 2 plots the histograms of $\text{Log}(N_F(T))$ and $\text{Log}(N_{F*}(T))$. Several trends are observed in the figure.

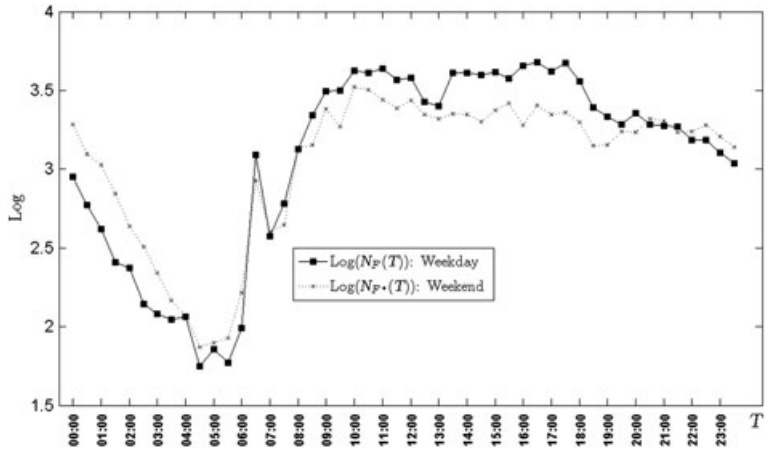


Figure 2. Average numbers of SMS arrivals per half hour in a weekday and a weekend.

In general, the number of SMS arrivals in a weekday are larger than that in a weekend during the working hours (6:30 to 19:30), and the result reverses for the non-working hours (20:00 to 6:00). At meal times (i.e., around 7:00, noon and 18:00), the number of SMS drops for both weekdays and weekends. During night hours (22:00 to 4:00), the number of SMS is a decreasing function of time.

Denote $N_{N,y}(T)$ as the number of SMS arrivals during the 30-min period $[T, T + 30\,\text{min})$ on the New Year's Day for year y, and $N_{N^*,y}(T)$ as that on the New Year's Eve for year y. Similarly, denote $N_{L,y}(T)$ and $N_{L^*,y}(T)$ as those in Lunar New Year's Day and Lunar New Year's Eve for year y, respectively. Let $S_2 = \{2007, 2008, 2009, 2010\}$, and let

$$\begin{aligned} N_N(T) &= \frac{\sum_{y\in S_2} N_{N,y}(T)}{|S_2|}, \\ N_{N^*}(T) &= \frac{\sum_{y\in S_2} N_{N^*,y}(T)}{|S_2|} \\ N_L(T) &= \frac{\sum_{y\in S_2} N_{L,y}(T)}{|S_2|}, \text{ and} \\ N_{L^*}(T) &= \frac{\sum_{y\in S_2} N_{L^*,y}(T)}{|S_2|} \end{aligned} \tag{2}$$

That is, $N_N(T)$ is the average number of SMS arrivals during the 30-min period $[T, T + 30\,\text{min})$ in the New Year's Days for 2007, 2008, 2009, and 2010. Similarly, $N_{N^*}(T)$ is the average number of New Year's Eves, $N_L(T)$ is that of Lunar New Years' Days, and $N_{L^*}(T)$ is that of Lunar New Years' Eves, respectively.

Figure 3(a) plots the $\text{Log}(N_{N^*}(T))$ and the $\text{Log}(N_{L^*}(T))$ curves for (Lunar) New Year's Eve. Figure 3(b) plots the $\text{Log}(N_N(T))$ and the $\text{Log}(N_L(T))$ curves for (Lunar) New Year's Day. These curves show that the trends of SMS traffic for the New Year's Day (Eve) and the Lunar New Year's Day (Eve) are similar. Both $N_{N^*}(T)$ and $N_{L^*}(T)$ tend to decrease from 0:00 to 6:00, and then increase from 6:00 to midnight. The peak traffic occurs at midnight of the (Lunar) New Year's Eve. In general, the $N_{L^*}(T)$ and the $N_{N^*}(T)$ values are larger than the $N_L(T)$ and the $N_N(T)$ values. Also, the trends of the (Lunar) New Year curves are similar to that of the (Lunar) New Year's Eve curves except that $N_N(T)$ ($N_L(T)$) decreases from 10:00 to midnight.

To closely investigate the histograms of $\text{Log}(N_F(T))$ and $\text{Log}(N_{F^*}(T))$ curves in Figure 2, we replace the 30-min period T by 1-min period t between 8:00 and 9:59 in Figure 4.

In Figure 4, $N_F(t)$ and $N_{F^*}(t)$ represent the average number of SMS arrivals during $[t, t + 1\,\text{min})$ in a weekday and a weekend, respectively. We found that a large volume of short messages are issued periodically in Figure 4. In every 30-min period, a large volume of short messages occur in a 3-min period called the *burst period* (see areas (a), (b), (c), and (d) in Figure 4). The traffic in the remaining 27 min is much smaller than that in the 3-min period, and is called the non-burst period. For example, in (8:30–9:00), the burst period occurs in (8:30–8:32) (area (b) in Figure 4). These SMS bursts are generated by some commercial users who periodically broadcast large numbers of short messages to their customers. Let $N_{F,B,d}(T)$ be the average number of the burst period in $(T, T + 30\,\text{min})$ in the dth day of February 2010, and $N_{F,N,d}(T)$ be the average SMS number of non-burst period in $(T, T + 30\,\text{min})$. Let $N_{L,B,y}(T)$ be the average number of burst traffic during $(T, T + 30\,\text{min})$ in the Lunar New Year's Day for year $y \in S_2$, and $N_{L,N,y}(T)$ be the average number of non-burst traffic. That is

$$N_{L,y}(T) = N_{L,B,y}(T) + N_{L,N,y}(T).$$

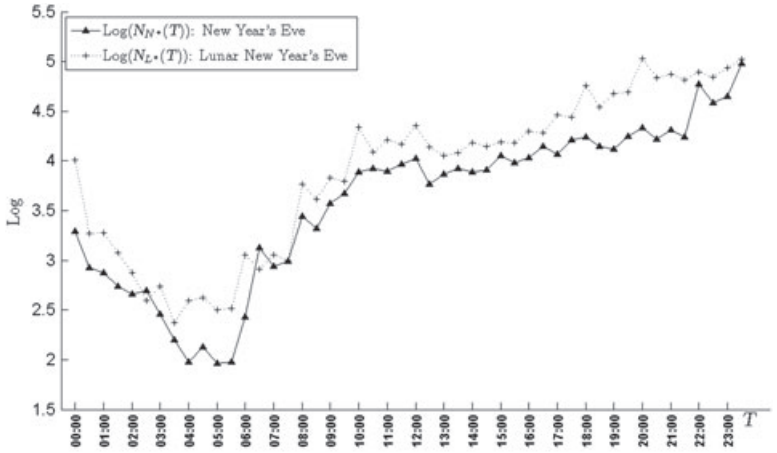
Similarly, we define $N_{L^*,B,y}(T)$ and $N_{L^*,N,y}(T)$ for Lunar New Years' Eves. Denote

$$\begin{aligned} N_{F,B}(T) &= \frac{\sum_{d\in S_1} N_{F,B,d}(T)}{|S_1|}, \\ N_{F,N}(T) &= \frac{\sum_{d\in S_1} N_{F,N,d}(T)}{|S_1|}, \\ N_{F^*,B}(T) &= \frac{\sum_{d\in S_1^*} N_{F,B,d}(T)}{|S_1^*|}, \\ N_{F^*,N}(T) &= \frac{\sum_{d\in S_1^*} N_{F,N,d}(T)}{|S_1^*|}, \\ N_{L,B}(T) &= \frac{\sum_{y\in S_2} N_{L,B,y}(T)}{|S_2|}, \\ N_{L,N}(T) &= \frac{\sum_{y\in S_2} N_{L,N,y}(T)}{|S_2|}, \\ N_{L^*,B}(T) &= \frac{\sum_{y\in S_2} N_{L^*,B,y}(T)}{|S_2|}, \text{ and} \\ N_{L^*,N}(T) &= \frac{\sum_{y\in S_2} N_{L^*,N,y}(T)}{|S_2|}. \end{aligned}$$

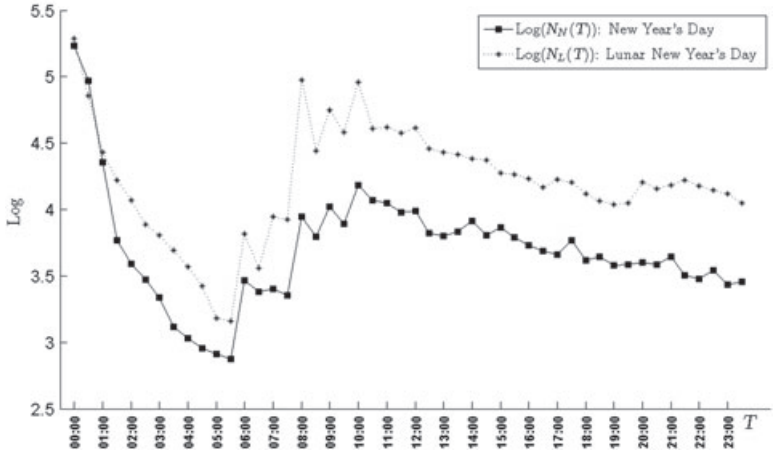
Figure 5 shows the curves of $N_{F,B}(T)$ (weekday burst) and $N_{F,N}(T)$(weekday non-burst). The figure indicates that most burst periods occur in (6:00, 18:00). Compared with the period (6:00, 18:00), there are fewer business activities in (22:00, 6:00), and therefore much less burst periods are observed in this time interval. On the average, the SMS volumes in the burst periods are 39.29% (weekday) and 45.58% (weekend) larger than that in the non-burst periods. To address such variance of SMS traffic in different time intervals, it is desirable to derive various arrival distribution functions based on different observed days and burst types. We will focus on eight traffic types (see Table I) in the remainder of this paper.

3. FITTING OF ARRIVAL DISTRIBUTIONS

The volume of SMS arrivals to a mobile telecom network is typically very large, and these arrivals can be viewed as



(a) New Year's Eve and Lunar New Year's Eve



(b) New Year's Day and Lunar New Year's Day

Figure 3. Average numbers of SMS arrivals per half hour for a (Lunar) New Year's Day and a (Lunar) New Year's Eve.

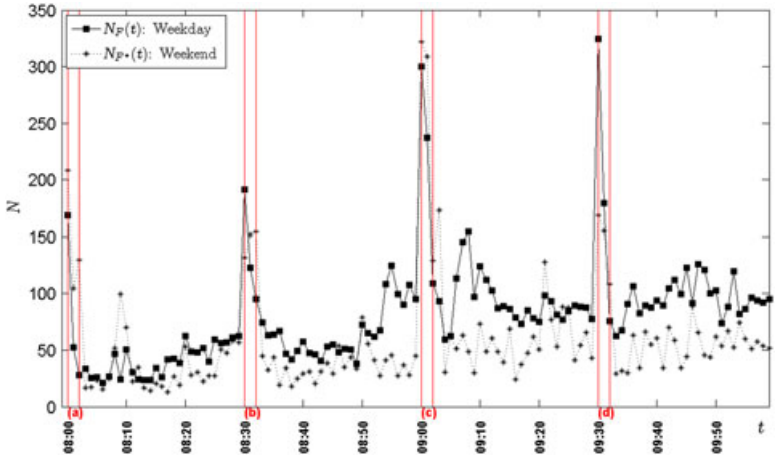


Figure 4. Average numbers of SMS arrivals per minute between 8:00 and 9:59 in a weekday and a weekend.

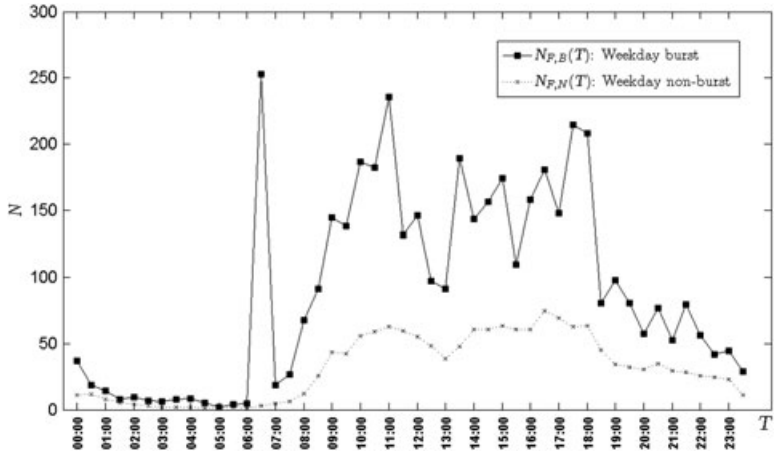


Figure 5. The numbers of SMS arrivals in the burst and the non-burst periods in a weekday.

Table I. Eight types of SMS arrival distributions.

Type	Description
F, *B*	Burst traffic in a weekday
F, *N*	Non-burst traffic in a weekday
*F**, *B*	Burst traffic in a weekend
*F**, *N*	Non-burst traffic in a weekend
L, *B*	Burst traffic in a Lunar New Year's Day
L, *N*	Non-burst traffic in a Lunar New Year's Day
*L**, *B*	Burst traffic in a Lunar New Year's Eve
*L**, *N*	Non-burst traffic in a Lunar New Year's Eve

statistically independent. Therefore, for each type-x traffic (where $x \in \{(F,B), (F,N), (F^*,B), (F^*,N), (L,B), (L,N), (L^*,B), (L^*,N)\}$), the traffic can be modeled by a non-homogeneous Poisson process with the arrival rate function $r_x(T)$. Let S be the set of the days considered in (1) and (2), then $N_x(T)$ is the average number of type-x SMS arrivals in $(T, T+30\,\text{min})$ over the days in S. Statistically, $N_x(T)$ can be seen as a random sampled data, and $r_x(T) = \lim_{|s|\to\infty} N_x(T)$, which is the expected number of SMS arrivals that occur per unit time. Because the measured data from CHT's commercial operation are drawn from a limited size of $S = S_1, S_1^*$, or S_2 in this study, $N_x(T)$ is an approximation of $r_x(T)$.

The $N_x(T)$ function can be fit by a non-parametric estimation $\hat{r}_x(T)$ through a kernel-based approximation method [13,14]. For every T, this method considers the points $T^* \in \mathbb{T}(\lambda, T) = [T-(\lambda/2), T+(\lambda/2)]$ and then scales T^* with the Nadaraya–Waston kernel weighted factor $K_\lambda(T^*, T)$ [10,11] to derive the estimation $\hat{r}_x(T)$, where

$$K_\lambda(T^*, T) = \begin{cases} \left(\frac{3}{4}\right)\left[1-\left(\frac{|T^*-T|}{\lambda}\right)^2\right], & \text{if } \left|\left(\frac{|T^*-T|}{\lambda}\right)\right| < 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

On the basis of (3), we express $\hat{r}_x(T)$ as

$$\hat{r}_x(T) = \frac{\sum_{T^*\in\mathbb{T}(\lambda,T)} K_\lambda(T^*, T)\, N_x(T^*)}{\sum_{T^*\in\mathbb{T}(\lambda,T)} K_\lambda(T^*, T)} \quad (4)$$

In (4), $K_\lambda(T^*, T)$ is a weight assigned to T^* based on its distance from T, and parameter λ specifies the width of the neighborhood used to estimate $r_x(T)$. The value of parameter λ can be determined by using cross validation method [13] or observed directly from the measured data. When λ is larger, the $\hat{r}_x(T)$ curve becomes smoother. As a non-bust traffic example, Figure 6(a) shows the curves for $N_{F,N}(T)$ and $\hat{r}_{F,N}(T)$ with $\lambda = 2$, 4, and 6. When λ is larger, the $\hat{r}_{F,N}(T)$ curve becomes farther away from $N_{F,N}(T)$. The errors between $\hat{r}_{F,N}(T)$ and $N_{F,N}(T)$ are 3.03%, 10.61% and 22.92% for $\lambda = 2$, 4, and 6, respectively. Clearly, when $\lambda = 2$, the $\hat{r}_{F,N}(T)$ curve is close to $N_{F,N}(T)$ and still smooth enough. Therefore, we choose $\lambda = 2$ for nonparametric estimation.

As a burst traffic example, Figure 6(b) shows the curves for $N_{L^*,B}(T)$ and $\hat{r}_{L^*,B}(T)$ with $\lambda = 2$, 4, and 6. The figure indicates that $\lambda = 2$ is not smooth, and $\lambda = 4$ is a better choice for $\hat{r}^*_{L^*,B}(T)$.

The nonparametric $\hat{r}_x(T)$ effectively smoothens the arrival rate function and provides useful insight to describe the SMS arrival distribution. However, for computational

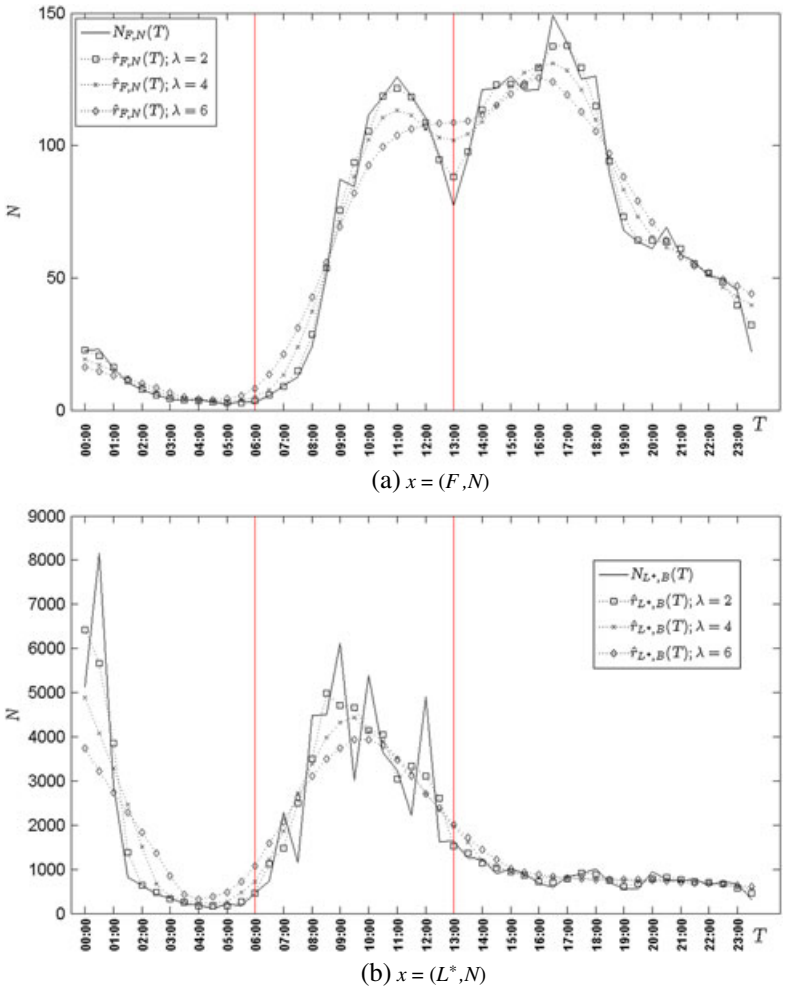


Figure 6. $N_x(T)$, $\hat{r}_x(T)$ and segments for $\hat{r}^*_x(T)$.

purposes, it is desirable to approximate $N_X(T)$ by a parametric function $\hat{r}^*_x(T)$ that satisfies two criterions:

Criterion 1. *$\hat{r}^*_x(T)$ is a continuous and piecewise polynomial function with some known knots and degree less than or equal to three.*

Criterion 2. *$\hat{r}^*_x(T)$ is close to $N_X(T)$ in the sense that the objective function $\Omega = \sum_T \left[\hat{r}^*_x(T) - N_X(T)\right]^2$ is minimized.*

Criterion 1 provides the guideline for generating the parametric form of $\hat{r}^*_x(T)$ such that the degree of the polynomial function is limited to 3. The objective function Ω in Criterion 2 measures the distance between $\hat{r}^*_x(T)$ and $N_X(T)$. To find the appropriate knots for $\hat{r}^*_x(T)$ that satisfy Criterion 1, the non-parametric $\hat{r}_x(T)$ plays an assistance role. For example, the $\hat{r}_{F,N}(T)$ curve ($\lambda = 2$) in Figure 6(a) suggests that $N_{F,N}(T)$ curve can be appropriately partitioned into 3 segments with two knots at $T = 6:00$ and $T = 13:00$, where each segment can be fit by a cubic polynomial function that satisfies Criterion 1. For the purpose of deriving $\hat{r}^*_x(T)$, we first replace the time period T by an index j. Define set $S_3 = \{(T/30\,\text{min}) + 1\} = \{1, 2, 3, \ldots, 48\}$. Then, both $N_X(T)$ and $\hat{r}^*_x(T)$can be transformed to $N_X(j)$ and $\hat{r}^*_x(j)$ where $j \in S_3$ For the $N_{N,F}$ curve in Figure 6, we select two knots at $j = 13$ (i.e., $T = 6:00$) and $j = 27$ (i.e $T = 13:00$), which are around the breakfast and the lunch times. For $0 \le j \le 13$, $\hat{r}^*_{N,F}(j)$ is a decreasing curve. For $13 \le j \le 27$, $\hat{r}^*_{N,F}(j)$ increases and then decreases. For $27 \le j \le 48$, $\hat{r}^*_{N,F}(j)$ also increases and then decreases. Other traffic types show similar trends, and can be segmented by the same knots. Therefore, $\hat{r}^*_x(j)$ is partitioned into three segments fit by

$$\hat{r}^*_x(j) = \begin{cases} \hat{r}^*_{x,1}(j), & 1 \le j \le 13 \\ \hat{r}^*_{x,2}(j), & 13 \le j \le 27 \\ \hat{r}^*_{x,3}(j), & 27 \le j \le 48 \end{cases} \quad (5)$$

Where

$$\hat{r}^*_{x,1}(13) = \hat{r}^*_{x,2}(13) \text{ and } \hat{r}^*_{x,2}(27) = \hat{r}^*_{x,3}(27) \quad (6)$$

The cubic polynomial arrival rate functions can be expressed as

$$\hat{r}^*_{x,1}(j) = a_{1,1}j^3 + a_{1,2}j^2 + a_{1,3}j + a_{1,4} \text{ for } 1 \le j \le 13, \quad (7)$$

$$\hat{r}^*_{x,2}(j) = a_{2,1}j^3 + a_{2,2}j^2 + a_{2,3}j + a_{2,4} \text{ for } 13 \le j \le 27, \quad (8)$$

and

$$\hat{r}^*_{x,3}(j) = a_{3,1}j^3 + a_{3,2}j^2 + a_{3,3}j + a_{3,4} \text{ for } 27 \le j \le 48 \quad (9)$$

From (6), we have

$$\begin{aligned} a_{2,4} &= 2197a_{1,1} + 169a_{1,2} + 13a_{1,3} + a_{1,4} \\ &\quad - 2197a_{2,1} - 169a_{2,2} - 13a_{2,3} \quad (10) \\ a_{3,4} &= 2197a_{1,1} + 169a_{1,2} + 13a_{1,3} + a_{1,4} \\ &\quad - 17486a_{2,1} + 560a_{2,2} + 14a_{2,3} - 19683a_{3,1} \\ &\quad - 729a_{3,2} - 27a_{3,3} \quad (11) \end{aligned}$$

Substitute (10) and (11) into (7)–(9) to yield

$$\hat{r}^*_{x,1}(j) = a_{1,1}j^3 + a_{1,2}j^2 + a_{1,3}j + a_{1,4} \quad (12)$$

$$\begin{aligned} \hat{r}^*_{x,2}(j) &= 2197a_{1,1} + 169a_{1,2} + 13a_{1,3} + a_{1,4} \\ &\quad + \left(j^3 - 2197\right)a_{2,1} + \left(j^2 - 169\right)a_{2,2} \\ &\quad + (j - 13)\,a_{2,3}, \end{aligned} \quad (13)$$

and

$$\begin{aligned} \hat{r}^*_{x,3}(j) &= 2197a_{1,1} + 169a_{1,2} + 13a_{1,3} + a_{1,4} \\ &\quad + 17486a_{2,1} + +560a_{2,2} + 14a_{2,3} \\ &\quad + \left(j^3 - 19683\right)a_{3,1} + \left(j^2 - 729\right)a_{3,2} \\ &\quad + (j - 27)\,a_{3,3}. \end{aligned} \quad (14)$$

Equations (12)–(14) can be represented in a matrix format

$$\hat{\boldsymbol{r}}^*_{\boldsymbol{x}} = \boldsymbol{J}\boldsymbol{a} \quad (15)$$

where $\hat{\boldsymbol{r}}^*_{\boldsymbol{x}} = \left[\hat{r}^*_{x,1}(1), \hat{r}^*_{x,1}(2), \ldots, \hat{r}^*_{x,1}(12), \hat{r}^*_{x,2}(13), \ldots, \hat{r}^*_{x,2}(26), \hat{r}^*_{x,3}(27), \ldots, \hat{r}^*_{x,3}(48)\right]^{\mathrm{T}}$ is a 48-component column vector, $\boldsymbol{a} = \left[a_{1,1}, a_{1,2}, a_{1,3}, a_{1,4}, a_{2,1}, a_{2,2}, a_{2,3}, a_{3,1}, a_{3,2}, a_{3,3}\right]^{\mathrm{T}}$ is a 10-component parameter vector, and $\boldsymbol{J}$ is a 48×10 matrix that consists of three sub-matrixes $\boldsymbol{J_1}\boldsymbol{J_2}$, and $\boldsymbol{J_3}$. That is

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J_1} \\ \boldsymbol{J_2} \\ \boldsymbol{J_3} \end{bmatrix} \quad (16)$$

In (16), $\boldsymbol{J_1}$ is a 12×10 matrix, $\boldsymbol{J_2}$ is a 14×10 matrix, and $\boldsymbol{J_3}$ is a 22×10 matrix. Let $\boldsymbol{J_{n,i}}$ be the i-th row vector of sub-matrix n (where $n = 1, 2, 3$). Then for $n = 1$ and $i = 1, \ldots, 12$, we have

$$\boldsymbol{J_{1,i}} = \left[i^3 i^2\ i\ 0\ 0\ 0\ 0\ 0\ 0\right]_{1\mathrm{x}10}$$

For $n = 2$ and $i = 1, \ldots, 14$, we have

$$\begin{aligned} \boldsymbol{J_{2,i}} = \Big[&2197\ 169\ 13\ 1\ \left((i+12)^3 - 2197\right) \\ &\times \left((i+12)^2 - 169\right)(i-1)\,0\ 0\ 0\Big]_{1\mathrm{x}10} \end{aligned}$$

For $n = 3$ and $i = 1, \ldots, 22$, we have

$$\begin{aligned} \boldsymbol{J_{3,i}} = \Big[&2197\ 169\ 13\ 1\ 17486\ 560\ 14\left((i+26)^3 - 19683\right) \\ &\times \left((i+26)^2 - 729\right)(i-1)\Big]_{1\mathrm{x}10} \end{aligned}$$

On the basis of (15), we derive the parameter vector $\boldsymbol{a}$ by using Criterion 2. Specifically, we plug (15) into the objective function

$$\begin{aligned} \Omega &= \sum_T \left[\hat{r}^*_x(T) - N_x(T)\right]^2 \\ &= \sum\nolimits_{j \in S_3} \left[\hat{r}^*_x(j) - N_x(j)\right]^2, \end{aligned}$$

where

$$\begin{aligned} S_3 &= \left\{\frac{T}{30\,\mathrm{min}} + 1\right\} \\ &= \left(\hat{\boldsymbol{r}}^*_{\boldsymbol{x}} - \boldsymbol{N_x}\right)^T \left(\hat{\boldsymbol{r}}^*_{\boldsymbol{x}} - \boldsymbol{N_x}\right) \end{aligned}$$

where $\boldsymbol{N_x}$ is a 48-component vector representing $N_x(j)$

$$\boldsymbol{N_x} = \left[N_x(1), N_x(2), \ldots, N_x(48)\right]^T \quad (17)$$

From (15) and (17), we have

$$\Omega = (\boldsymbol{Ja} - \boldsymbol{N_x})^T (\boldsymbol{Ja} - \boldsymbol{N_x}) \quad (18)$$

where $(\boldsymbol{Ja} - \boldsymbol{N_x})$ is a 48-component vector. To minimize this objective function, we take partial derivatives of Ω with respect to parameters $\boldsymbol{a}$ and set them equal to 0. That is, from (18), we solve

$$\begin{aligned}\Omega' &= \frac{\partial\left[(\boldsymbol{Ja}-\boldsymbol{N_x})^T(\boldsymbol{Ja}-\boldsymbol{N_x})\right]}{\partial \boldsymbol{a}} \\ &= \frac{\partial\left[\left(\boldsymbol{a}^T\boldsymbol{J}^T-\boldsymbol{N}_x^T\right)(\boldsymbol{Ja}-\boldsymbol{N_x})\right]}{\partial \boldsymbol{a}} \\ &= \left\{\left[\frac{\partial\left(\boldsymbol{a}^T\boldsymbol{J}^T-\boldsymbol{N}_x^T\right)}{\partial \boldsymbol{a}}\right](\boldsymbol{Ja}-\boldsymbol{N_x})\right\}^T \\ &\quad + (\boldsymbol{Ja}-\boldsymbol{N_x})^T\left[\frac{\partial(\boldsymbol{Ja}-\boldsymbol{N_x})}{\partial \boldsymbol{a}}\right] \\ &= \left[\boldsymbol{J}^T(\boldsymbol{Ja}-\boldsymbol{N_x})\right]^T + (\boldsymbol{Ja}-\boldsymbol{N_x})^T\boldsymbol{J} \\ &= 2(\boldsymbol{Ja}-\boldsymbol{N_x})^T\boldsymbol{J}\end{aligned} \tag{19}$$

If we set $\Omega' = 0$, then from (19) we have $[\boldsymbol{Ja}-\boldsymbol{N_x}]^T\boldsymbol{J} = \boldsymbol{a}^T\boldsymbol{J}^T\boldsymbol{J}-\boldsymbol{N}_x^T\boldsymbol{J} =$, which leads to

$$\boldsymbol{a}^T\boldsymbol{J}^T\boldsymbol{J} = \boldsymbol{N}_x^T\boldsymbol{J} \tag{20}$$

By multiplying $\left(\boldsymbol{J}^T\boldsymbol{J}\right)^{-1}$ from right in both sides of (20), we have

$$\boldsymbol{a}^T = \boldsymbol{N}_x^T\boldsymbol{J}\left(\boldsymbol{J}^T\boldsymbol{J}\right)^{-1} \tag{21}$$

By transposing the matrices of both sides of (21), we have

$$\begin{aligned}\boldsymbol{a} &= \left[\left(\boldsymbol{J}^T\boldsymbol{J}\right)^{-1}\right]^{\mathrm{T}}\boldsymbol{J}^T\boldsymbol{N_x} \\ &= \left(\boldsymbol{J}^T\boldsymbol{J}\right)^{-1}\boldsymbol{J}^T\boldsymbol{N_x}\end{aligned} \tag{22}$$

Equation (22) guarantees that Criterion 2 is satisfied. On the basis of this equation, the computed $\boldsymbol{a}$, $a_{2,4}$, and $a_{3,4}$ values for eight types of traffic are showing in Table II. Define the error between $\hat{r}_x^*(j)$ and $N_x(j)$ as

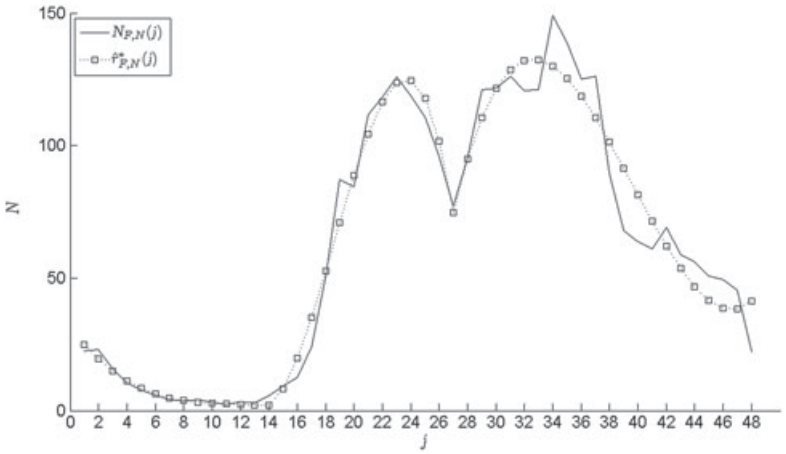
$$\boldsymbol{error} = \frac{\hat{r}_x^*(j) - N_x(j)}{N_x(j)} \tag{23}$$

Figures 7(a) and (b) show the $N_x(j)$ and the $\hat{r}_x^*(j)$ curves, where $x = (F, N)$ and (L^*, B). In Figure 7(a), the error (i.e., Equation (23)) between the two curves is 1.09%, and in Figure 7(b), the error is 8.11%.

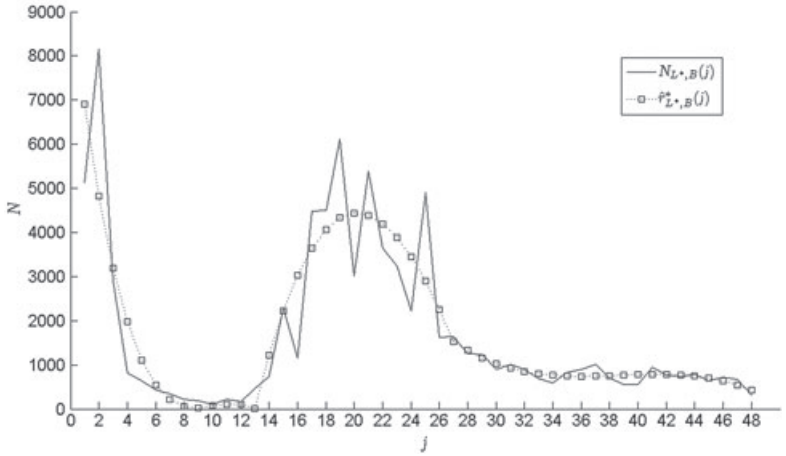
Table III shows the errors between $\hat{r}_x^*(T)$ and $N_x(T)$ for different x. The table indicates that $\hat{r}_x^*(j)$ accurately fits $N_x(j)$ for non-burst traffic types with errors between 0.25% and 6.79%. On the other hand, the errors for burst traffic types are between 8.11% and 15.43%. The larger error incurred by the burst $\hat{r}_x^*(j)$ than the non-burst one on the same day is due to the fact that the variance of burst traffic is larger than that of non-burst traffic. Furthermore, the burst traffic happens once every half an hour and the

Table II. The $a_{n,l}$ values for different x($n = 1, 2, 3$, and $l = 1, 2, 3, 4$).

	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$
F, *B*	0.42	−6.30	17.37	38.22
F, *N*	−0.02	0.60	−7.21	31.53
*F**, *B*	0.03	1.76	−36.82	167.58
*F**, *N*	0.01	0.48	−11.45	64.59
L, *B*	−6.00	228.00	−2596.00	9182.00
L, *N*	−9.00	260.10	−2370.10	7130.10
*L**, *B*	−4.38	115.99	−923.53	2173.90
*L**, *N*	−0.61	16.40	−135.56	353.15
	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,4}$
F, *B*	−0.61	33.77	−588.86	3415.30
F, *N*	−0.24	13.35	−229.73	1257.90
*F**, *B*	−0.14	4.79	−10.41	−309.42
*F**, *N*	−0.15	8.66	−151.35	845.96
L, *B*	−0.10	−40.10	2180.10	−1983.70
L, *N*	0.01	−44.00	1500.10	−12,875.10
*L**, *B*	−5.77	320.50	−5559.90	30,945.00
*L**, *N*	−0.75	44.66	−812.82	4.68
	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$
F, *B*	0.19	−22.07	846.03	−10,214.00
F, *N*	0.071	−8.21	315.18	−3807.80
*F**, *B*	0.01	−1.18	49.17	−455.01
*F**, *N*	0.01	−1.02	39.75	−433.22
L, *B*	−1.00	60.00	−2305.10	3019.10
L, *N*	−0.01	7.00	−343.10	5838.10
*L**, *B*	−0.91	94.97	−3004.60	30,450.00
*L**, *N*	−0.88	100.28	−3614.10	42,299.00



(a) $x = (F,N)$



(b) $x = (L^*,B)$

Figure 7. $N_x(j)$ and $\hat{r}^*_x(j)$ $(j = \{(T/30\text{min}) + 1\})$.

Table III. The errors between $\hat{r}^*_x(T)$ and $N_x(T)$.

x(non-burst)	F, N	F^*, N	L, N	L^*, N
error	1.09%	0.25%	5.95%	6.79%
x(burst)	F, B	F^*, B	L, B	L^*, B
error	13.24%	15.34%	8.11%	14.43%

non-burst traffic happens every minute, which implies that more samples are collected for the non-burst traffic, and therefore can be approximated more accurately.

4. CONCLUSION

This paper derived the arrival distribution functions for SMS. We considered SMS arrival distributions for different traffic types and observation days. Specifically, we modeled the short message arrivals as non-homogeneous Poisson processes. Then we compute the arrival rate functions based on the measured data from CHT's commercial operation. On the basis of the SMS arrival distributions derived from our model, the mobile operators have better understanding about the volumes of short messages in different times and days, which can be used to design more flexible short message charging rates. For example, peer-to-peer SMS has a higher charging rate at busy hour (17:00) and business users have a lower charging rate if their applications do not send SMS at [0,30] min each hour.

We observed that the SMS arrival curves have two major turnover points around the breakfast and the lunch times. Therefore, we partitioned a day into three time zones, and approximated each zone by a cubic polynomial arrival rate function. For non-burst traffic, the errors between the derived arrival rate functions and the measured data are between 0.25% and 6.79%. For burst traffic, the errors are between 8.11% and 15.43%. Our study

indicated that the errors of the arrival rate functions for burst traffic are acceptable for network planning purposes of commercial SMS operation. Although the user behavior of other telecom operators may be different from that of Chunghwa Telecom, they can apply our model with their measured data to predict the potential SMS volume in their commercial operation, and then modify their network configurations to achieve SMS traffic load balancing.

REFERENCES

1. Halepovic E, Williamson C, Ghaderi M, Wireless data traffic: A decade of change. *IEEE Network* 2009: 20–26.
2. Lin P, Wu S-H, Chen C-M, Liang C-F. Implementation and performance evaluation for a ubiquitous and unified multimedia messaging platform. *Springer Wireless Networks* 2009; **15**(2): 163–176.
3. Huang-Fu C-C, Lin Y-B, Chung-HwaRao H. iP2P: A peer-to-peer system for mobile devices. *IEEE Wireless Communications* 2009: 30–36.
4. Markett C, Arnedillo I, Sánchez SW, Tangney B. Using short message service to encourage interactivity in the classroom. *Computers & Education* 2006; **46**(3): 280–293.
5. Sou S-I, Lin Y-B, Wu Q, Jeng J-Y. Modeling prepaid application server of VoIP and messaging services for UMTS. *IEEE Transactions on Vehicular Technology* 2007; **46**(3): 1434–1441.
6. forum SMS. Short Message Peer-to-Peer Protocol (SMPP) specification version 5.0, 2003.
7. 3GPP. *Mobile Application Part (MAP) specification TS 29.002,* 3rd Generation Partnership Project (3GPP), Sep. 2010.
8. 3GPP. *Technical realization of the Short Message Service (SMS). TS 23.040,* 3rd Generation Partnership Project (3GPP), Sep. 2008.
9. 3GPP. *Network Architecture. TS 23.002*, 3rd Generation Partnership Project (3GPP), Sep. 2010.
10. Petros Z, Xiaoquao M, Starsky HYW. A study of the short message service of a nationwide cellular network. *ACM Internet Measurement Conference* 2006: 263–268.
11. Sou S-I, Lin Y-B, Lou C-L. Cost analysis of short message retransmissions. *IEEE Transactions on Mobile Computing* 2009; **9**: 215–225.
12. Wu N, Wu M, Chen S. Real-time monitoring and filtering system for mobile SMS. *IEEE Conference on Industrial Electronics and Applications* 2008: 1319–1324.
13. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, (2nd edn). Springer: New York, 2009.
14. Simonoff JS. *Smoothing Methods in Statistics*. Springer: New York, 1996.

AUTHORS' BIOGRAPHIES

Hui-Nien Hung received the Ph.D. degree in statistics from the University of Chicago, Chicago, Illinois in 1996. He is currently a professor at the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. His research interests include applied probability, biostatistics, statistical inference, statistical computing, and industrial statistics.

Yi-Bing Lin is Vice President and Life Chair professor of the College of Computer Science, National Chiao Tung University (NCTU), and a Visiting professor of King Saud University. He is also with the Institute of Information Science and the Research Center for Information Technology Innovation, Academia Sinica, Nankang, Taipei, Taiwan, R.O.C. Lin has authored books on Wireless and Mobile Network Architecture (Wiley, 2001), Wireless and Mobile All-IP Networks (John Wiley, 2005), and Charging for Mobile All-IP Telecommunications (Wiley, 2008). Lin has received numerous research awards including the 2005 NSC Distinguished Researcher and 2006 Academic Award of Ministry of Education. Lin is an ACM Fellow, an AAAS Fellow, an IEEE Fellow and an IET Fellow.

Chao-Liang Luo is currently working toward the Ph.D. degree at the Department of Computer Science and Engineering, National Chiao Tung University. In 2001, he joined the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., and was involved in the implementation of value-added services in mobile networks. In 2005, he was with the short message service team. Since then, he has been involved in the design of the Next Generation Network (NGN), mobile packet switched data and multimedia services, and the study of mobile network evolution. His research interests include the design and analysis of personal communications services network, 3G networks, wireless Internet, mobile computing, and performance modeling.