

측정 자료를 기반으로 한 SMS 트래픽의 모형화

정회원 최 천 원*, 이 정 수**, 김 민 석**, 하 장 용**, 준회원 하 준*

Measurement based Modeling of SMS Traffic

Cheon Won Choi*, Jeoung Soo Lee**, Min Seok Kim**, Jang Yong Ha** *Regular Members*

Jun Ha* *Associate Member*

요 약

서비스의 다양화가 진행되는 이동 통신망에서는 데이터 서비스 사용량의 증가에 따라 음성 서비스의 사용량에 조율되어 온 교환 용량의 산출 방식에 대한 수정이 요구되고 있으며, 이를 위해 데이터 서비스 트래픽에 대한 모형화의 필요성이 대두되고 있다. 본 논문에서는 이동 통신망에서 활발한 사용이 있는 데이터 서비스의 하나인 SMS를 대상으로 트래픽을 측정 및 분석하고 이러한 측정 자료를 기반으로 SMS 트래픽 모형을 구축한다. 본 논문에서 개발되는 SMS 트래픽 모형은 측정 자료를 토대로 구성된 SMS 호 도착 프로세스와 데이터 트래픽의 고유 성질이 반영된 호 지속 시간으로 특징지어지는 가입자 큐잉 시스템을 근간으로 하며 단순성과 측정 자료 반영의 용이성을 갖는다. 이러한 SMS 트래픽 모형의 정확성을 검증하기 위해 측정 자료와 모의 실험을 통해 구한 SMS 호 발생량을 비교하여 평균 SMS 호 발생량이 통계적으로 동일함을 확인한다.

ABSTRACT

In the progression of service diversification at wireless cellular networks, the prevailing use of data service invokes a revised rule of predicting required switching capacity and modeling data service traffic becomes of utmost necessity as a basis of such revision. In this paper, we measure the traffic flow involved in provisioning SMS and construct an SMS traffic model based on the measurements. The proposed SMS traffic model is founded on the subscriber queueing system characterized by a call arrival process based on measurements and call duration time reflecting distinctive properties of data service traffic. The simple structure enables an easy improvement of the model with support of supplementary measurements. To evaluate the accuracy of the model, we compare the SMS call generation processes yielded by measurement and simulation, and confirm that the two average amounts of SMS calls generated in reference hour are statistically identical.

I. 서 론

이동 통신망에서는 단순한 음성 서비스(plain old telephone service: POTS) 제공의 제약을 넘어서 WWW 서비스 등 데이터 서비스를 포괄하는 서비스의 다양화가 진행되고 있다. 또한 데이터 서비스의 사용자 및 사용량도 증가하는 추세이며, 이러한 서비스의 다양화와 데이터 서비스 사용량의 증가

경향은 제 3 세대 이동 통신망의 본격적인 보급과 함께 심화될 것으로 예상된다^{[1][2][3][4]}. 전통적으로 이동 통신망에서는 음성 서비스가 총 서비스 사용량의 대부분을 차지하여 왔으며, 따라서 서비스의 원활한 제공을 위한 교환국(mobile switching center: MSC)의 교환 용량도 음성 서비스의 사용량에 조율되어 왔다. 그러나 예상되는 서비스 다양화와 데이터 서비스 사용량의 증가에 비추어 이러한

* 단국대학교 전기전자컴퓨터공학부(cchoi@dku.edu),

** (주) SK Telecom

논문번호 : 020412-0927, 접수일자 : 2002년 9월 27일

※ 본 연구는 (주) SK Telecom의 지원으로 수행되었습니다.

교환 용량의 산출 방식은 수정되어야 하고, 이를 위하여 데이터 서비스의 제공으로 발생하는 트래픽의 성질에 대한 분석 및 모형화가 선행되어야 한다.

데이터 트래픽의 관찰은 초기에는 근거리 통신망(local area network: LAN)인 Ethernet에서 이루어졌으며^{[5][6]}, 최근에는 사용량이 급증하고 있는 유선 인터넷에서 트래픽의 관찰이 수행되고 있다. 이러한 관찰로부터 데이터 트래픽은 burstiness와 self-similarity 등 음성 트래픽과 비교하여 상이한 성질을 갖고 있다는 사실이 문헌을 통해 발표되었다^{[5][6][7]}. 그러나 이동 통신망에서 음성 서비스로 인해 발생하는 트래픽에 대해서는 실측 자료가 있는 반면 데이터 서비스의 트래픽에 대해서는 실측 자료가 미흡하고 검증된 트래픽 모형도 부재한 상태이다.

현재 이동 통신망에서 활발한 사용이 있는 데이터 서비스 중 하나는 단문 서비스(short message service: SMS)이다^{[2][3][4]}. 본 논문에서는 이동 통신망에서 제공되는 데이터 서비스의 트래픽을 모형화하기 위한 기초 연구로서 SMS의 트래픽을 측정하고 이를 토대로 SMS 트래픽의 모형을 제시한다. 이러한 모형화는 다음과 같이 수행된다. 이동 통신망에서 제공되는 SMS를 대상으로 첫째, MSC에서 시간별 발생하는 SMS의 호 수를 측정하고 이러한 측정 자료에 대해 통계적 분석을 수행한다. 둘째, 각 서비스 사용자에게 대해 단일 서버 및 단일 버퍼(single-server and single-buffer)의 가입자 큐잉 시스템(subscriber queueing system)을 구성하고 이러한 큐잉 시스템의 SMS 호 도착 프로세스를 측정 자료에 기초하여 nonhomogeneous Poisson process로 설정하며 데이터 트래픽의 고유 성질을 반영하여 호 지속 시간에 Pareto 분포를 부여한다. 마지막으로 SMS 호 발생 프로세스를 각 가입자 큐잉 시스템에서 수락되는 호가 도착하는 시간의 중첩된 프로세스로 모형화한다.

본 논문은 다음과 같이 구성된다. 2 절에는 SMS 트래픽의 측정 및 통계적 분석의 결과가 서술되고, 3 절에서는 SMS 호 발생 프로세스의 모형화가 수행된다. 4 절에서는 SMS 트래픽 모형에 대한 모의 실험이 수행되고, 모의 실험의 결과와 측정 자료와의 비교를 통해 모형의 정확성이 검증된다.

II. SMS 트래픽의 측정 및 통계적 분석

본 절에서는 이동 통신망의 MSC에서 측정된 SMS 트래픽 자료를 소개하고 이에 대한 통계적 분

석을 수행한다.

표 1은 하루 중 음성 서비스 호가 가장 많이 발생하는 시간대로 정의되는 기준 시간대에 천 명의 가입자당 발생한 SMS 호 수를 보여 준다. SMS 호 수의 측정은 8 개월 동안 수행되었다. 이 동안 매달 보편성을 갖는 10 일을 선정하여 각 날 기준 시간대에 발생하는 SMS 호 수를 측정하고 해당 일의 가입자 수를 이용하여 표 1을 구성하였다.

표 1. 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수.

	M1	M2	M3	M4	M5	M6	M7	M8
D 1	56	56	67	92	107	127	129	109
D 2	59	58	66	93	108	132	128	108
D 3	61	57	72	89	101	137	127	114
D 4	61	53	73	96	100	133	118	113
D 5	57	59	78	101	103	130	124	109
D 6	60	69	82	102	107	132	120	117
D 7	55	64	85	103	105	133	111	108
D 8	60	68	83	104	99	138	113	108
D 9	61	70	85	101	96	134	120	121
D10	58	66	85	105	111	125	119	122

표 1에서 Mm 은 m 번째 달 그리고 Dd 는 해당하는 달의 d 번째 측정일을 의미한다. 표 1에 주어진 m 번째 달의 측정일 Dd 에 가입자 천 명당 발생한 SMS 호 수를 x_{dm} 이라고 하자. 또한 m 번째 달의 기준 시간대에 발생하는 SMS 호 수의 평균(mean)과 분산(variance)의 추정값을 각각 $\hat{\mu}_m$ 과 $\hat{\sigma}_m^2$ 라고 하자. 이러한 추정값을 moment 추정 방식으로 다음과 같이 정한다^{[8][9]}.

$$\hat{\mu}_m = \frac{1}{10} \sum_{d=1}^{10} x_{dm} \tag{1}$$

$$\hat{\sigma}_m^2 = \frac{1}{9} \sum_{d=1}^{10} (x_{dm} - \hat{\mu}_m)^2.$$

월별 기준 시간대에 발생하는 SMS 호 수가 정규 분포(normal distribution)를 갖는지 여부를 파악하기 위해 Kolmogorov와 Smirnov의 적합성 검정(goodness-of-fit test)을 수행한다^{[8][10]}. 표 2는 이러한 검정 결과를 보여 준다. (표 2의 평균과 표준 편차는 식 (1)에 의해 계산된 추정값이다.) 표에 나타난 p -value로부터 SMS 호 수가 정규 분포를 갖는다는 가설을 기각할 근거가 미흡함을 알 수 있다. 이러한 정규성(normality)에 근거하여 월별 기준 시

표 2. 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수에 대한 Kolmogorov- Smirnov 검정 결과.

month	mean	standard deviation	Kolmogorov-Smirnov statistic	p-value
M1	58.8	2.1	0.646	0.799
M2	62.1	6.1	0.529	0.942
M3	77.6	7.5	0.719	0.679
M4	98.7	5.8	0.832	0.493
M5	103.7	4.6	0.420	0.995
M6	132.1	3.8	0.463	0.983
M7	120.8	6.4	0.536	0.936
M8	112.9	5.5	0.770	0.593

간대에 발생하는 SMS 호 수의 평균의 동질성 (homogeneity)을 검증하는 분산 분석(analysis of variance)을 수행한다^{[8][9][10]}. 표 3은 평균의 동질성을 확인하기 위한 F 검정의 결과를 보여 준다. 이 표의 p-value로부터 월별 기준 시간대에 발생하는 SMS 호 수의 평균이 동일하다는 가설을 기각할 근거가 충분함을 알 수 있다.

표 3. 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수에 대한 F 검정 결과.

source	degrees of freedom	mean square	F statistic	p-value
month	7	7361.5	249.2	0.000
day	72	29.5		
total	79			

한편 월별 기준 시간대에 발생하는 SMS 호 수의 분산이 동일한지의 여부를 확인하기 위해 Levene의 검정을 수행한다^[10]. 표 4는 이러한 검정 결과를 보여 주며 p-value로부터 월별 분산이 동일하다는 가설을 기각할 근거가 충분함을 알 수 있다. 표 2, 3, 4의 검정 결과를 종합하여 월별 기준 시간대에 가입자 천 명당 발생하는 SMS 호 수는 모두 정규 분포를 따르나 그 평균과 분산은 서로 상이함을 알 수 있다.

표 4. 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수에 대한 Levene 검정 결과.

Levene statistic	3.267
degrees of freedom 1	7
degrees of freedom 2	72
p-value	0.005

평균과 분산 외에 월별 기준 시간대에 발생하는 SMS 호 수를 대표하는 주요 파라미터는 99th percentile이다^{[8][9][11]}. 이러한 m번째 달의 99th percentile의 추정값을 $\hat{\gamma}_m$ 이라고 하자. 일반적으로 99th percentile은 다음과 같이 정의되는 sample 99th percentile로 추정된다^[11].

$$\hat{\gamma}_m = x_{(10.99 \times 101)_m} \tag{2}$$

식 (2)에서 $x_{(h)_m}$ 은 m 번째 달의 SMS 호 수 샘플 중 h 번째로 작은 샘플을 의미하고 [a]는 a 보다 크거나 같은 최소 정수이다. 한편 표 2에 주어진 Kolmogorov와 Smirnov의 적합성 검정을 통해 SMS 호 수가 정규 분포를 따름이 확인되었으므로 정규 분포의 성질을 이용하여 99th percentile을 다음과 같이 추정할 수 있다^{[8][9]}.

$$\hat{\gamma}_m = \hat{\mu}_m + 2.33 \hat{\sigma}_m \tag{3}$$

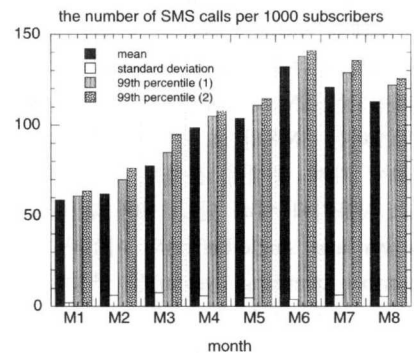


그림 1. 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수의 평균, 표준 편차, 99th percentile의 추정값.

그림 1은 표 1에 주어진 자료로부터 구한 월별 기준 시간대에 가입자 천 명당 발생한 SMS 호 수의 평균, 표준 편차, 99th percentile의 추정값을 보여 준다. 이 그림에서 평균과 표준 편차의 추정값은 식 (1)에 의해 구해지고 99th percentile의 추정값인 99th percentile(1)과 99th percentile(2)는 각각 식 (2)와 (3)에 의해 구해진다.

표 5는 하루 중 매 시간대에 천 명의 가입자당 발생한 SMS 호 수를 보여 준다. 이러한 시간대별 SMS 호 수의 측정은 보편성을 갖는 11 일 동안 수행되었고 해당 일의 가입자 수를 이용하여 표 5가 구성되었다. 표 5에서 Dd는 d 번째 측정일을 의미하고 Hh는 하루 중 (h-1) 시부터 h 시까지의 시간대 (h-1,h)를 나타낸다. 측정일 Dd의 시간대 Hh

에 발생한 SMS 호 수를 y_{hd} 라고 하자. 모든 $d \in \{1, \dots, 11\}$ 에 대해 샘플 $\{y_{hd}: h=1, \dots, 24\}$ 가 분포 F_d 로부터 얻어졌다고 하자. 이러한 분포 $\{F_d: d=1, \dots, 11\}$ 의 독립성과 동일성을 검정하기 위해 chi-square 검정을 수행한다^{[8][9]}. 이 검정에 사용되는 Pearson의 chi-square 통계량은 다음과 같이 정의된다.

$$\chi^2 = \sum_{h=1}^{24} \sum_{d=1}^{10} \frac{\left[y_{hd} - \frac{y_{h \cdot} \cdot y_{\cdot d}}{y_{\cdot \cdot}} \right]^2}{\frac{y_{h \cdot} \cdot y_{\cdot d}}{y_{\cdot \cdot}}} \quad (4)$$

식 (4)에서 $y_{h \cdot} = \sum_{d=1}^{11} y_{hd}$, $y_{\cdot d} = \sum_{h=1}^{24} y_{hd}$, $y_{\cdot \cdot} = \sum_{h=1}^{24} \sum_{d=1}^{11} y_{hd}$. 표 6은 표 5의 자료에 대한 chi-square 검정 결과를 보여 준다.

표 5. 하루 중 시간대별 가입자 천 명당 발생한 SMS 호 수.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
H1	80	89	72	77	79	72	87	87	87	80	86
H2	49	51	53	51	50	51	49	45	42	46	44
H3	26	28	25	27	28	29	21	20	23	28	21
H4	15	14	13	14	15	17	12	10	12	16	11
H5	8	9	8	9	9	10	8	6	7	9	6
H6	7	7	7	7	7	7	6	5	6	7	5
H7	9	9	8	9	8	7	8	10	11	7	12
H8	16	17	14	16	15	10	16	26	26	11	29
H9	27	28	24	30	24	16	29	42	38	19	45
H10	48	48	46	51	37	25	48	62	53	29	68
H11	57	58	56	54	52	35	59	68	60	42	72
H12	66	66	65	63	59	46	68	77	69	50	70
H13	75	71	68	74	63	54	77	79	67	62	73
H14	74	72	76	76	62	58	84	84	68	61	75
H15	72	72	71	74	67	56	77	75	65	65	76
H16	75	75	70	76	63	54	76	80	69	62	80
H17	68	67	64	71	62	59	74	76	65	64	77
H18	65	68	66	74	60	58	65	79	66	67	75
H19	69	72	72	76	57	53	69	79	67	70	79
H20	71	64	70	72	57	57	71	81	72	77	82
H21	85	77	81	85	77	83	86	88	85	88	86
H22	90	82	90	93	80	94	90	99	93	102	98
H23	97	87	88	90	90	97	101	109	98	107	111
H24	96	84	91	100	91	97	97	107	102	111	106

표 6. 하루 중 시간대별 가입자 천 명당 발생한 SMS 호 수에 대한 chi-square 검정 결과.

Pearson's chi-square statistic	144.816
degrees of freedom	230
p-value	1.000

표 6에 주어진 p-value로부터 분포 $\{F_d: d=1, \dots, 11\}$ 가 독립이고 아울러 동일하다는 가설을 기각할 근거가 미흡함을 알 수 있다. 따라서 이러한 검정 결과를 바탕으로 시간대별 SMS 호의 발생 빈도를 다음과 같이 추정한다. $h=1, \dots, 24$ 에 대해 $\xi(h)$ 를 임의의 SMS 호가 하루 중 시간대 $(h-1, h]$ 에 발생할 확률이라고 하면 $\xi(h)$ 의 추정값 $\hat{\xi}(h)$ 를 다음과 같이 정한다.

$$\hat{\xi}(h) = \frac{1}{11} \sum_{d=1}^{11} \frac{y_{hd}}{\sum_{g=1}^{11} y_{gd}} \quad (5)$$

그림 2에는 측정 자료를 이용하여 구해진 시간대별 SMS 호 발생 빈도가 도시되어 있다.

한편 각 시간대에 발생하는 SMS 호 수가 Poisson 분포를 따르는지의 여부를 확인하기 위해 Kolmogorov와 Smirnov의 적합성 검정을 수행한다.

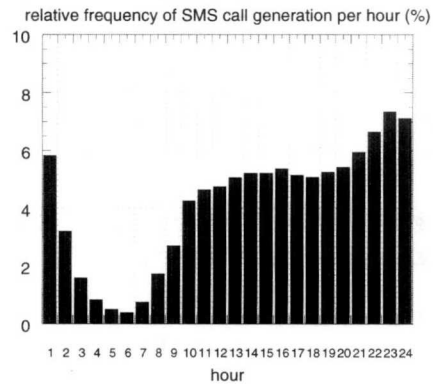


그림 2. 시간대별 SMS 호 발생 확률의 추정값.

표 7은 이러한 검정 결과를 보여주며 Kolmogorov-Smirnov 통계량에 따른 p-value로부터 SMS 호 수가 Poisson 분포를 갖는다는 가설을 기각할 근거가 일반적으로 미흡함을 알 수 있다.

III. SMS 호 발생 프로세스의 모형화

본 절에서는 SMS 호가 발생하는 시간의 프로세스를 모형화한다. 이를 위해 우선 SMS 호와 관련된 가입자별 큐잉 시스템을 구축하고, 이러한 큐잉 시스템의 기본 요소인 SMS 호의 서비스 시간과 SMS 호가 가입자에 도착하는 시간의 시퀀스를 규정한다. 이어서 가입자 큐잉 시스템에 도착하는

표 7. 하루 중 시간대별 가입자 천 명당 발생한 SMS 호 수에 대한 Kolmogorov-Smirnov 검정 결과.

hour	mean	Kolmogorov-Smirnov statistic	p-value
H1	81.5	0.614	0.845
H2	48.3	0.741	0.642
H3	25.1	0.621	0.836
H4	13.5	0.471	0.980
H5	8.1	0.676	0.751
H6	6.5	1.063	0.208
H7	8.9	0.714	0.688
H8	17.8	0.813	0.523
H9	29.3	0.677	0.749
H10	46.8	0.630	0.823
H11	55.7	0.491	0.970
H12	63.5	0.637	0.811
H13	69.4	0.376	0.999
H14	71.8	0.458	0.985
H15	70.0	0.610	0.851
H16	70.9	0.718	0.681
H17	67.9	0.430	0.993
H18	67.5	0.598	0.867
H19	69.4	0.643	0.803
H20	70.4	0.644	0.802
H21	83.7	0.983	0.289
H22	91.9	0.747	0.632
H23	97.7	0.427	0.993
H24	98.4	0.414	0.995

SMS 호 중 수락된 호의 도착 시간으로 가입자별 SMS 호의 발생 시간을 모형화한다.

1. 가입자 큐잉 시스템

SMS 호의 시도와 관련하여 각 가입자를 호가 도착하면 이를 버퍼에 저장하고 서비스 규칙(service discipline)에 따라 버퍼에 저장된 호 중 하나를 선택하여 호 지속 시간동안 서비스를 제공하는 큐잉 시스템으로 모형화할 수 있다. 일반적으로 한 가입자는 동시에 둘 이상의 SMS 호에 참여할 수 없으므로 호가 지속되는 동안 새로운 호의 발생 시도가 있다면 이 시도는 실패한다. 이러한 특성은 단일 서버 및 단일 버퍼(single-server single-buffer)의 가입자 큐잉 시스템을 구축하여 반영될 수 있다^{[12][16]}. 즉, 호의 서비스가 진행되는 동안에 또 다른 호가 도착한다면 이 호는 버퍼에 저장되지 않고 큐잉 시스템에서 거부된다.

2. SMS 호 지속 시간

음성 트래픽과 구분되는 데이터 트래픽의 주요 성질은 burstiness와 self-similarity로서, burstiness는

호 (혹은 세션) 지속 시간 동안 전송되는 데이터 양이 시 구간별로 균일하지 못한 성질이며 self-similarity 성질은 단위 시간당 전송되는 데이터 양의 time-average와 ensemble-average가 flow 수 혹은 시간의 증가에 따라 매우 낮은 속도로 수렴함으로써 나타난다^{[5][6][7]}. 본 절에서는 이 들 성질이 반영된 SMS 호 지속 시간의 모형을 제시한다.

데이터 서비스에서 burstiness 성질로 인해 호 지속 시간은 데이터 전송이 이루어지는 활성 시간과 그렇지 않은 비활성 시간의 반복으로 구성된다. 한편 분산이 매우 큰 분포를 활성 시간에 부여하여 self-similarity 성질을 활성 및 비활성 시간의 반복 모형에 반영할 수 있다. 이러한 분포의 하나는 Pareto 분포로서, 만약 확률 변수 S가 Pareto 분포를 갖는다면

$$P(S \leq t) = 1 - \left[\frac{t}{\alpha} \right]^{1 + \frac{\alpha}{\beta}} \tag{6}$$

식 (6)에서 $\alpha = \min\{S\}$ 이고 $\beta = E(S)$ 이다^{[5][13]}. 식 (6)으로부터 Pareto 분포는 평균은 유한하나 분산은 무한대임을 알 수 있다.

일반적으로 SMS 호 지속 시간은 단일 활성 시간으로 구성된다. 따라서 본 논문에서는 SMS 호 지속 시간을 식 (6)의 Pareto 분포를 갖는 확률 변수로 모형화한다.

3. SMS 호 도착 프로세스

본 절에서는 2 절에 서술된 하루 중 시간대별 발생하는 SMS 호 수에 대한 관찰에 기초하여 3.1 절의 가입자 큐잉 시스템에 SMS 호가 도착하는 프로세스를 모형화한다.

그림 2에 나타난 바와 같이 SMS 호 발생 빈도는 시간대별로 차이를 나타낸다. 한편 표 7의 Kolmogorov와 Smirnov의 검정 결과는 SMS 호 발생이 Poisson 분포를 따른다는 가설을 높은 유의 수준으로도 기각할 수 없음을 보여 준다. 이러한 추정 및 검정 결과로부터 유추하여 가입자 k의 가입자 큐잉 시스템에서 SMS 호 도착 프로세스를 intensity 함수 $\lambda_k^*(t)$ 의 nonhomogeneous Poisson process로 설정한다^{[14][15]}. 임의의 달 Mm 동안 모든 $h \in \{1, \dots, 24\}$ 에 대해 시간대 (h-1, h]에 가입자 k의 큐잉 시스템에 도착하는 평균 SMS 호 수를 $\eta_k^*(m, h)$ 라고 하자. 그러면 이 달의 intensity 함수 $\lambda_k^*(m, t)$ 는 다음과 같이 구성된다.

$$\lambda_k^*(m, t) = \sum_{h=1}^{24} \eta_k^*(m, h) \cdot I_{(h-1 < t \text{ mod } 24 \leq h)} \quad (7)$$

식 (7)에서 intensity 함수값은 다음과 같이 추정된다. 시간대 $(h_o-1, h_o]$ 를 기준 시간대라고 하고 이러한 기준 시간대에 발생하는 가입자 k 의 평균 SMS 호 수 $\eta_k(m, h_o)$ 의 추정값을 $\hat{\eta}_k(m, h_o)$ 라고 하자. 만약 모든 가입자가 SMS 호 발생과 관련하여 동질성을 갖는다고 가정하면, 식 (1)에 주어진 가입자 천 명당 기준 시간대에 발생한 평균 SMS 호 수의 추정값 $\hat{\mu}_m$ 을 이용하여 $\hat{\eta}_k(m, h_o)$ 를 다음과 같이 정한다.

$$\hat{\eta}_k(m, h_o) = \frac{\hat{\mu}_m}{1000} \quad (8)$$

2 절에 서술된 바와 같이 $\zeta(h)$ 는 $h \in \{1, \dots, 24\}$ 에 대해 (임의의 가입자의) 임의의 SMS 호가 하루 중 시간대 $(h-1, h]$ 에 발생할 확률을 나타내고 이 확률의 추정값 $\hat{\zeta}(h)$ 는 식 (5)에 주어져 있다. 모든 가입자가 SMS 호 발생과 관련하여 동질성을 갖는다고 가정하면, 임의의 달 Mm 동안 시간대 $(h-1, h]$ 에 발생하는 가입자 k 의 SMS 호 수 $\eta_k(m, h)$ 는 다음과 같이 추정된다.

$$\hat{\eta}_k(m, h) = \hat{\eta}_k(m, h_o) \frac{\hat{\zeta}(h)}{\hat{\zeta}(h_o)} \quad (9)$$

한편 주어진 시간대 $(h-1, h]$ 안에서는 intensity 가 균일하므로 가입자 k 의 큐잉 시스템을 M/G/1/1 시스템으로 근사할 수 있다. 따라서 M/G/1/1 큐잉 시스템의 성질로부터 시간대 $(h-1, h]$ 에 가입자 k 큐잉 시스템에 도착하는 평균 SMS 호 수 $\eta_k^*(m, h)$ 와 발생하는 평균 SMS 호 수 $\eta_k(m, h)$ 에 대해 다음의 관계식을 얻는다^{[12][16]}.

$$\eta_k(m, h) = \eta_k^*(m, h) [1 - \eta_k^*(m, h) E(S)] \quad (10)$$

식 (10)에서 S 는 SMS 호 지속 시간이다. 식 (9)에 주어진 $\eta_k(m, h)$ 의 추정값을 이용하여 $\eta_k^*(m, h)$ 는 다음과 같이 추정된다.

$$\hat{\eta}_k(m, h) = \hat{\eta}_k^*(m, h) [1 - \hat{\eta}_k^*(m, h) E(S)] \quad (11)$$

마지막으로 식 (11)의 추정값 $\hat{\eta}_k^*(m, h)$ 를 이용하여 임의의 달 Mm 동안 intensity 함수값의 추정을 다음과 같이 완성한다.

$$\hat{\lambda}_k^*(m, t) = \sum_{h=1}^{24} \hat{\eta}_k^*(m, h) \cdot I_{(h-1 < t \text{ mod } 24 \leq h)} \quad (12)$$

4. SMS 호 발생 프로세스

임의의 달 Mm 동안 임의의 가입자 k 의 가입자 큐잉 시스템에 SMS 호가 도착하는 시퀀스를 $\{A_k^*(m, n), n=0,1,\dots\}$ 이라고 하자. 이러한 SMS 호 중 단일 버퍼의 가입자 큐잉 시스템에서 거부되지 않은, 즉 수락된 SMS 호의 도착 시간을 $\{A_k(m, n), n=0,1,\dots\}$ 라고 하자. 이 때 $\{A_k(m, n), n=0,1,\dots\}$ 은 $\{A_k^*(m, n), n=0,1,\dots\}$ 의 subsequence이다. 달 Mm 의 가입자 수를 $U(m)$ 이라고 하자. 그러면 중첩된(superposed) 시퀀스 $\bigcup_{k=1}^{U(m)} \{A_k(m, n), n=0,1,\dots\}$ 을 임의의 달 Mm 동안의 SMS 호 발생 프로세스로 설정한다.

가입자 큐잉 시스템은 단일 서버 및 단일 버퍼 큐잉 시스템이므로 각 가입자는 임의의 시간에 최대 한 개의 호에 참여할 수 있다. 따라서 가입자 수가 $U(m)$ 이라면 최대 $U(m)$ 개의 SMS 호가 동시에 진행될 수 있으나 실제 이동 통신망에서는 무선 링크의 용량 제약으로 인해 이러한 최대 $U(m)$ 개의 호에 대한 동시 지원이 이루어지지 못한다. 그러므로 본 논문의 SMS 트래픽 모형은 호 발생률에 대한 낙관적인(optimistic) 추정값을 제공한다.

IV. SMS 호 발생 프로세스 모형의 검증

본 절에서는 3 절에 서술된 SMS 트래픽 모형을 바탕으로 모의 실험을 통해 SMS 트래픽을 발생시키고, 기준 시간대에 발생하는 SMS 호 수의 샘플을 추출하여, 평균, 분산, 99th percentile 등 주요 파라미터의 추정값을 구한다. 이어서 측정 자료로부터 구한 주요 파라미터의 추정값과 모의 실험을 통해 얻은 추정값을 비교하여 SMS 트래픽 모형의 정확성을 검토한다.

모의 실험에서 SMS 호 지속 시간으로는 Pareto 분포 모형이 적용되고 이때 평균 호 지속 시간은 1 msec, 최소 호 지속 시간은 평균 호 지속 시간의

반으로 설정된다. 또한 모의 실험에서 8 개월간 매 달 10 일 동안 SMS 트래픽을 발생시키며 가입자 수는 측정 자료로부터 구해진 매 달 평균 가입자 수가 사용된다. 이 때 모든 가입자의 동질성이 가정 된다.

표 8. 모의 실험과 측정을 통해 구한 월별 기준 시간대에 발생한 SMS 호 수.

	M1		M2		M3		M4	
	simu.	meas.	simu.	meas.	simu.	meas.	simu.	meas.
D1	0.577	0.592	0.599	0.652	0.740	0.872	1.062	1.140
D2	0.602	0.603	0.613	0.660	0.733	0.861	1.072	1.144
D3	0.623	0.602	0.608	0.658	0.792	0.863	1.026	1.145
D4	0.628	0.602	0.564	0.659	0.812	0.862	1.109	1.141
D5	0.584	0.602	0.633	0.660	0.857	0.863	1.168	1.147
D6	0.615	0.602	0.732	0.658	0.908	0.862	1.182	1.144
D7	0.564	0.603	0.679	0.662	0.940	0.860	1.192	1.144
D8	0.614	0.604	0.727	0.660	0.918	0.862	1.203	1.145
D9	0.621	0.603	0.749	0.659	0.941	0.860	1.172	1.145
D10	0.596	0.603	0.705	0.658	0.937	0.862	1.213	1.147
mean	0.602	0.602	0.659	0.661	0.863	0.859	1.144	1.140
S.D.	0.003	0.021	0.003	0.065	0.003	0.083	0.002	0.067

	M5		M6		M7		M8	
	simu.	meas.	simu.	meas.	simu.	meas.	simu.	meas.
D1	1.260	1.222	1.485	1.547	1.481	1.386	1.227	1.271
D2	1.277	1.224	1.533	1.545	1.468	1.382	1.211	1.271
D3	1.199	1.227	1.595	1.544	1.454	1.387	1.277	1.273
D4	1.182	1.228	1.550	1.543	1.346	1.383	1.265	1.270
D5	1.213	1.228	1.519	1.535	1.422	1.381	1.227	1.269
D6	1.261	1.228	1.541	1.538	1.370	1.381	1.316	1.273
D7	1.238	1.220	1.545	1.543	1.268	1.386	1.214	1.267
D8	1.171	1.225	1.610	1.537	1.292	1.384	1.210	1.274
D9	1.138	1.224	1.557	1.538	1.368	1.382	1.355	1.273
D10	1.317	1.226	1.462	1.544	1.359	1.384	1.374	1.273
mean	1.225	1.226	1.541	1.540	1.384	1.383	1.271	1.268
S.D.	0.003	0.055	0.004	0.045	0.002	0.073	0.002	0.062

표 8은 8 개월 동안 매 달 10 일의 기준 시간대에 발생한 SMS 호 수를 보여 준다. 이 표에서 월별 SMS 호 수는 측정과 모의 실험의 두 가지 방법으로 구해 졌다. 한편 월별 평균과 표준 편차는 식 (1)의 moment 추정 방식으로 추정된 값이다.

표 1에 대한 Kolmogorov와 Smirnov의 검정 결과는 정규성을 입증하므로 표 8에 주어진 측정 자료와 모의 실험 결과를 비교하는 분산 분석을 수행한다. 표 9는 평균의 동일성을 검증하기 위한 F 검정과 분산

의 동일성을 검증하기 위한 Levene 검정의 결과를 보여 준다. 표 9의 p -value로부터 측정과 모의 실험의 두 가지 방법으로 구한 기준 시간대에 발생한 SMS 호 수의 평균은 동일하다는 가설을 기각할 근거가 미흡함을 알 수 있다. 그러나 분산이 동일하다는 가설을 기각할 근거는 충분함을 보여 준다.

표 9. 모의 실험과 측정을 통해 구한 월별 기준 시간대에 발생한 SMS 호 수에 대한 F 검정과 Levene 검정 결과.

month	F statistic	p -value	Levene statistic	p -value
M1	0.013	0.909	20.544	0.000
M2	0.012	0.914	51.391	0.000
M3	0.030	0.864	36.530	0.000
M4	0.041	0.842	44.729	0.000
M5	0.001	0.978	23.981	0.000
M6	0.013	0.911	9.381	0.007
M7	0.001	0.979	22.960	0.000
M8	0.038	0.848	23.959	0.000

그림 3과 4에는 측정과 모의 실험의 두 가지 방법으로 구해진 기준 시간대에 발생한 SMS 호 수의 평균과 표준 편차의 추정값이 도시되어 있다. 한편 그림 5는 식 (3)에 의해 구해진 99th percentile의 추정값을 보여 준다. 그림 3과 4에서 평균의 동일성과 분산의 상이성을 보여 주는 표 9의 검정 결과를 확인할 수 있다. 또한 그림 4에서 모의 실험 방법으로 구한 표준 편차의 추정값은 측정 자료에서 얻은 표준 편차의 추정값에 비해 작음을 알 수 있다. 한편 평균에 비해 표준 편차는 상대적으로 매우 작음을 보여 준다. 이로 인해 그림 5에 나타난 바와 같이 분산의 상이성에도 불구하고 측정과 모의 실험 방법으로 구한 99th percentile의 추정값은 상당한 유사성을 갖는다.

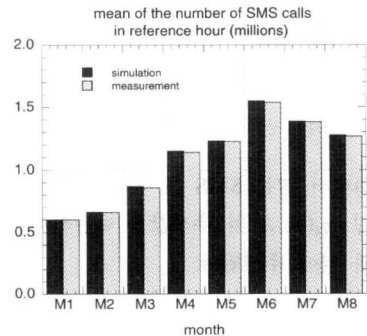


그림 3. 월별 기준 시간대에 발생하는 SMS 호 수의 평균의 추정값.

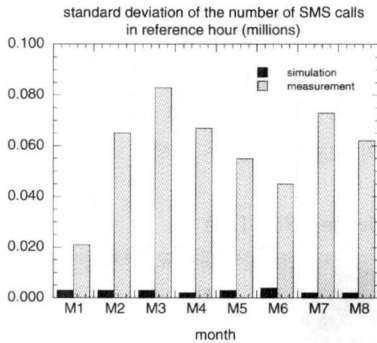


그림 4. 월별 기준 시간대에 발생하는 SMS 호 수의 표준 편차의 추정값.

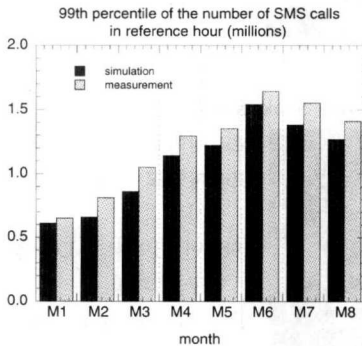


그림 5. 월별 기준 시간대에 발생하는 SMS 호 수의 99th percentile의 추정값.

이러한 분산의 상이성은 SMS 트래픽 모형에 내재한 다음 두 가지 요인에 기인한다고 판단된다. 한 가지는 모의 실험에서는 가입자 수로 한 달간의 평균 가입자 수가 사용되는 반면 측정 자료는 매일 변동하는 가입자 수를 반영하고 있는 점이다. 다른 요인은 식 (8)과 (9)에 전제된 SMS 호의 발생에 대해 모든 가입자가 동질성을 갖는다는 가정이다. 일별 가입자 수의 변동과 가입자의 이질성을 SMS 트래픽 모형에 반영할 경우, 기준 시간대에 발생하는 SMS 호 수의 분산은 증가하게 되고 따라서 모의 실험을 통해 얻은 분산 및 표준 편차의 추정값 측정 자료로부터 구한 추정값에 보다 근접하게 된다.

V. 결론

본 논문에서는 이동통신망에서 제공되는 SMS로 인해 발생하는 호 수를 측정하고 이러한 측정 자료를 바탕으로 하루 중 시간대별 SMS 호의 발생 빈도와 월별 기준 시간대에 발생하는 SMS 호 수의 평균, 분산, 99th percentile 등의 추정값을 구하였

다. 이어서 시간대별 SMS 호의 발생 빈도에 대한 관찰을 바탕으로 SMS 호 도착 프로세스를 nonhomogeneous Poisson process로 모형화하고 SMS 호 지속 시간에 Pareto 분포 모형을 적용하여 단일 서버 및 단일 버퍼의 가입자 큐잉 시스템을 구성하였다. 마지막으로 이러한 가입자 큐잉 시스템에서의 수락된 호 도착 프로세스로서 SMS 호의 발생 프로세스를 모형화하였다.

본 논문에서 개발된 SMS 트래픽 모형에서 모의 실험 방법으로 추출한 기준 시간대에 발생하는 평균 SMS 호 수의 추정값은 측정 자료로부터 구한 평균의 추정값과 상당히 유사함을 확인하였다. 그러나 기준 시간대에 발생하는 SMS 호 수의 분산에서는 이러한 추정값이 차이를 나타내었다. 이러한 현상은 SMS 트래픽 모형에서 가정된 가입자의 동질성에 기인한다고 판단된다. 따라서 연령, 지역 등의 조건으로 가입자를 분류하여 가입자의 군을 구성한 후, 가입자 군별 측정 자료를 확보하여 가입자의 이질성을 모형에 반영함으로써 본 SMS 트래픽 모형의 정확성이 개선될 것으로 예상된다.

참고 문헌

- [1] P. Chaudhury, W. Mohr, and S. Onoe, "The 3GPP Proposal for IMT-2000," *IEEE Communications Magazine*, Vol. 37, No. 12, pp. 72-81, December 1998.
- [2] V. Garg, *IS-95 CDMA and cdma2000*. Prentice Hall, 1999.
- [3] H. Holma and A. Toskala, *WCDMA for UMTS*. John Wiley & Sons, 2000.
- [4] D. Knisely, S. Kumar, S. Laha, and S. Nanda, "Evolution of Wireless Data Services: IS-95 to cdma2000," *IEEE Communications Magazine*, Vol. 36, No. 10, pp. 140-149, October 1998.
- [5] R. Addie, M. Zukerman, and T. Neame, "Broadband traffic Modeling : Simple Solutions to Hard Problems," *IEEE Communications Magazine*, Vol. 36, No. 8, pp. 88-95, August 1998.
- [6] W. Leland, "On the Self-similar Nature of Ethernet Traffic," *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, pp. 1-15, February 1994.
- [7] A. Andersen and B. Nielsen, "A Markovian

