1

Project	**IEEE 802.16 Broadband Wireless Access Working Group <http://ieee802.org/16>**
Title	**Email Traffic Model**
Date Submitted	2007-09-06
Source(s)	Yih-Guang Jan, Yang-Han Lee, Ming-Hsueh Chuang, Hsien-Wei Tseng, Jheng-Yao Lin, and Chih-Wei Su Voice: +886-2-2625-2303 E-mail: yihjan@yahoo.com yhlee@ee.tku.edu.tw Institute for Information Industry 7F., No. 218, Sec. 2, Dunhua S. Rd., Taipei City, Taiwan. Department of Electrical Engineering, Tamkang University 151 Ying-chuan Road, Tamsui, Taipei County, Taiwan 25137, R. O. C. [co-authors added here]
Re:	IEEE 802.16m-07/080r3– Call for Comments on Draft 802.16m Evaluation Methodology Document
Abstract	This document contains proposed text for the draft evaluation methodology for IEEE 802.16m technical proposals.
Purpose	For discussion and approval by TGm
Notice	
Release	
Patent Policy	

2

Email Traffic Model

References

[1] Laura A. Dabbish, Robert E. Kraut, Susan Fussell and Sara Kiesler, "Understanding Email Use: Predicting Action on a Message," Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'05), NY: ACM Press, pp.691-700.
[2] V. Bolotin, Y. Levy, and D. Liu," Characterizing Data Connection and Messages by Mixtures of Distributions on Logarithmic Scale, ITC 99, Edinburgh.
[3] G.. Brasche, B. Walke, " Concepts Services, and Protocols of the New GSM Phase 2+ General Packet Radio Service, IEEE Communications Magazine, August 1997.
[4] D. Staehle, K. Leibnitz and P. Tran-Gia," Source Traffic Modeling of Wireless Applications," University of Wurzburg, Institute of Information, Research Report Series, Report No. 261, June 2000.
[5] Bong H. Kim, and Y Hur," Application Traffic Model for WiMAX Simulation," POSDATA, Ltd, April 2007.
[6] M. S. Borella," Source Models of Network Game Traffic", Computer Communications, 23 (4), pp. 403-410.

1. Introduction

Traditionally when it is in the development of Internet traffic model it is usually based on the assumption that the Internet traffic flow arrives according to a memory-less Poisson process, which results in the traffic exhibiting the short-term or short range autocorrelation. However, lately it drew people's attention and one found that the aggregated Internet traffic model depicts the long-term autocorrelation, i.e. the autocorrelation function of the traffic remains significant for all lags and it is identified as Self-Similar Process [6]. These Internet traffics statistical distributions can be identified by either Cauchy, or Pareto or Weibull distribution [6]. In this report we will review the characteristics of Email traffic among various Internet traffics and to develop an appropriate Email traffic model identified from these characteristics. In the literature it has not too many references discussing Email traffic. In [2] and [3] they provided the distribution of the Email size, and it is noted that the 90%-tile Email size varies from 80 Kbytes in the model [4] to 250 Kbytes in [3]. And also in [4] it found that Email size can be approximated by a Cauchy distribution function with $\alpha = 0.8$ and $\beta = 1.0$. In the following some general statistics about Email usages are listed or tabulated from the survey conducted at Carnegie Mellon University containing over 1100 Email addresses [1]. The Email traffic model will then be discussed in sequel.

2. Basic Email Statistics

The general statistics about email usage in the survey is tabulated in Table 1 [1]. On the average every respondent sent 14 messages per day, read 30 of them and kept over 1300 in their inboxes.

Table 1 General Email Usage Characteristics

Message	Mean (Standard Deviation)	Median (Out of N=121)
Message read per day	30 (17)	25
Message sent per day	14 (12)	13
Number of inbox Message	1336 (2785)	105
Number of Email folders	22 912)	25
Times checking Email per day	19 (11)	13

3. Message Level Data

The distribution of messages among the various content types is summarized in Table 2. It is to be noted that it is possible for one message containing one more type. The highest percentage of the message content is to ask for action (34%)

Table 2 Distribution of Message Content Types
(One message may contain more than one type)

Message Content	Percentage
Action request	34
Information request	18
Information Attachment	36
Status update	21
Scheduling	14
Reminder	16
Social	8
Other	12

4. Actions on a message

Two possible actions are considered for people taking on a message, namely the *location action* and *reply action*. In the *location action,* it decides people's action on a message to file, delete, or leave the message after processing it. In the reply action, it considers user's response to a message, the user had already replied to, planned to reply, or did not plan to reply a message. The detailed breakdown of the distribution of messages by reply and location actions can be referred to Figure 1.

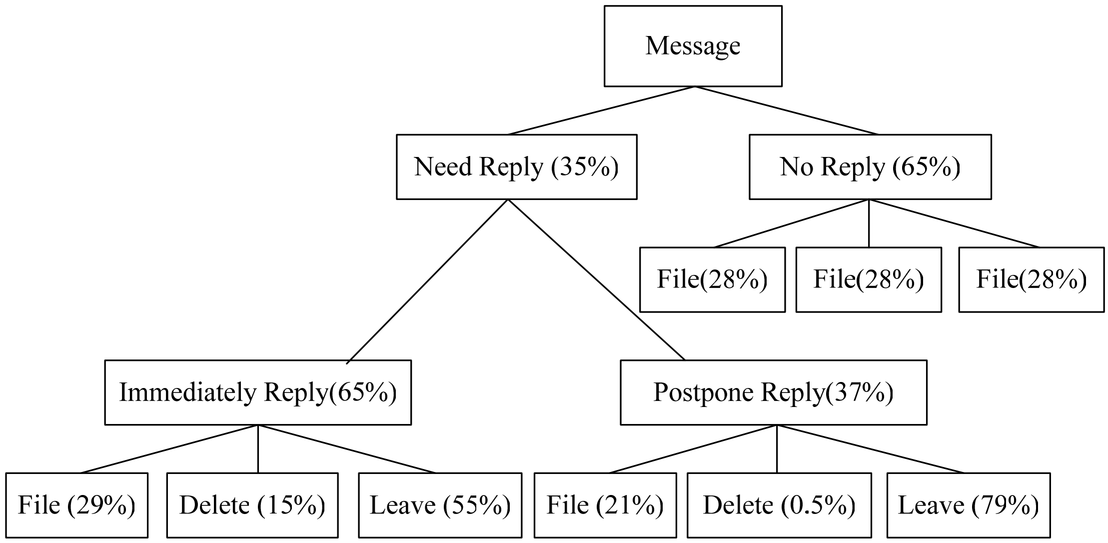

Figure 1 Distribution of Messages by *Reply* and *Location* Actions
(Categories are Mutually Exclusive)

5. Email Protocol

The mostly used Email protocols are POP3 and MAPI (Messaging Application Programming Interface) which is supported by Microsoft Outlook and Exchange Server. The MAPI protocol is in the application layer. In the Outlook, each E-mail involves eleven active TCP connections during Email invoking phase, and each Email transaction consists of multiple MAPI segment transactions in series and each MAPI segment is again segmented into smaller segments. The maximum MAPI segment is 16896 bytes and this information is indicated in the first package of a MAPI segment. Outlook finishes the MAPI segment with ACK acknowledgement transmission, while the Exchange server waits for the MAPI segment completion indication packet before sending the next one. The last packet in the MAPI segment set the "PUSH" bit in the TCP packet to transmit all of the packets in the TCP buffer to the application layer at the receiver side [5].

6. Email Traffic Model

The Email traffic has the burst profile as other Internet traffics, and it is characterized by ON/OFF states. In the ON-state Email traffics are transmitted and in the Off-state it is in the idle period. When the duration of the ON-state is short compared with the ON-OFF session length, then other new Email traffics (sub-sessions) are read and the elapse time between Emails sub-session read is randomly distributed as shown in Figure 2.

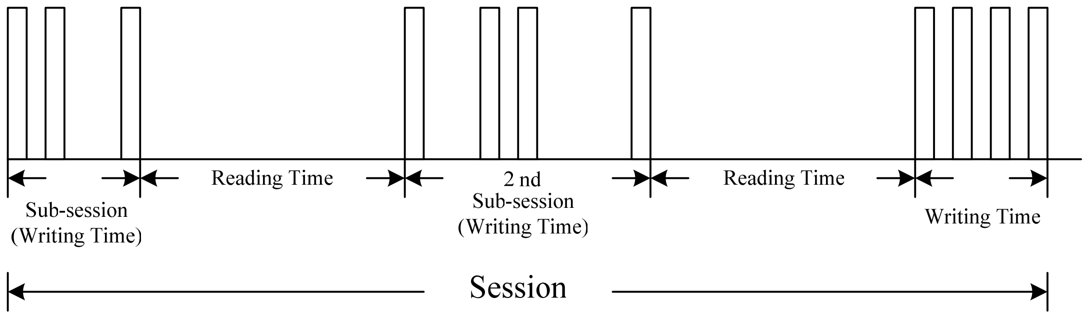

Figure 2 Email Traffic Pattern

Based on the Email traffic patterns, the Email traffic simulation model can be summarized in Table 3 [5] [6].

Table 3 Email Traffic Simulation Model

Component	Distribution	Parameters	PDF
E-Mail Protocol		POP3, MAPI	N/A
E-Mail Average Header Size (Bytes)	Deterministic	1 K	N/A
Number of Read Sub-sessions	Lognormal	Mean μ= 30 Std σ = 17	$f_x = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-(\ln x-\mu)^2}{2\sigma^2}\right] \quad x \geq 0$
Number of Write Sub-sessions	Lognormal	Mean μ= 14 Std σ = 12	$f_x = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-(\ln x-\mu)^2}{2\sigma^2}\right] \quad x \geq 0$
Reading Time Per Sub-session (Seconds)	Pareto	$\alpha = 1.1,, k = 2, m = 65$, mean = 60, maximum = 63	$f_x = \frac{\alpha k^\alpha}{x^{\alpha+1}}, k \leq x \leq m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m$
Writing Time Per Sub-session (Seconds)	Pareto	$\alpha = 1.1,, k = 2, m = 125$, mean = 120, maximum = 123	$f_x = \frac{\alpha k^\alpha}{x^{\alpha+1}}, k \leq x \leq m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m$
E-Mail Size without Attachment (Bytes)			
Read	Cauchy	median $\mu = 22.7$ K, 90%-tile = 80K	$f_x = \frac{A}{\pi((x-\mu)^2+1)}$, A is selected to satisfy 90%-tile value
Write	Cauchy	median $\mu = 22.7$ K, 90%-tile = 80K	$f_x = \frac{A}{\pi((x-\mu)^2+1)}$, A is selected to satisfy 90%-tile value
E-Mail Size with Attachment (Bytes)			

Read	Cauchy	median $\mu = 227$ K , 90%-tile = 800 K	$f_x = \frac{A}{\pi((x-\mu)^2+1)}$, A is selected to satisfy 90%-tile value
Write	Cauchy	median $\mu = 227$ K , 90%-tile = 800 K	$f_x = \frac{A}{\pi((x-\mu)^2+1)}$, A is selected to satisfy 90%-tile value