Incoming Traffic Modeling of Heterogeneous Public Safety Network

Tuyatsetseg Badarch  
Department of Electrical and Computer Engineering  
Northeastern University  
Boston, MA02215, USA  
b_tuyatsetseg@yahoo.com

Otgonbayar Bataa  
School of Information and Communications Technology  
(SICT), Mongolian University of Science and Technology  
(MUST), Mongolia  
otgonbayar@sict.edu.mn

Abstract—Accurate traffic models are necessary for a service provider of emergency communications to properly maintain the capacity planning of the network. To meet the necessity, the good traffic model must be developed that can capture characteristics including occurrences of few and large random incidents and accidents. These occurrences may be described as the unusual spikes and long tails in a probability model term of an actual network load. Therefore, the proposed mixed model for a peak period traffic brings significant results to capture the characteristics than other statistical models. We employ the Lognormal mixture model with high accuracy based on simulations with the real data set of emergency incoming traffic types. The results show that main traffic of a peak period is fitted reasonably by a mixture of Lognormal distributions with two or more components compared to well known statistical candidate distributions such as General Pareto, Weibull, and simple Lognormal.

Keywords— Traffic mixture modeling; MLE of Lognormal mixture; Emergency call traffic; Public Safety Network

I. INTRODUCTION

The ability to access emergency services by dialing fixed numbers is a vital component of public safety and emergency preparedness. In emergency communication networks that serve various public safety personnel, including medical responders, police, and hazard and fire fighters, the incoming voice calls from the affected population contribute to the maximum volume of traffic, which may not be supported by the existing infrastructure because of logistical constraints.

The heterogeneous concept of PSN means that the authorities responsible for public safety can use public communication networks to manage systematic and integrated emergency services more efficiently and reliability. In the feature of integrated services, interconnected VoIP service enables one to make and receive calls to and from PSTN using an Internet connection, possibly a broadband Internet connection, such as Digital Subscriber Line (DSL), cable modem, or wireless broadband.

To the best of the author's knowledge, in literature, there is a lack of research about traffic models of emergency traffic into the heterogeneous Public Safety Network. The traffic over fixed, cellular, and trunked radio networks is traditionally assumed to be approximated by well known standard models such as negative Exponential, Lognormal, Weibull, and General Pareto. However, it is often observed in standard fitting models that the probability of very short occurrences (e.g., long tails) is overestimated, while the area with the highest probability (e.g., spikes) in the empirical histogram is underestimated [1], [2].

Therefore, to overcome the fitting problem in the probability model, recently, the research of this field tends to account for mixture models such as hyper exponential, phase type distribution, erlang mixture, in which a probability distribution is represented as a linear superposition of component distributions [3]-[7]. However, the lognormal mixture model lacks of application with explicit mathematical tools. In this study, we show that the lognormal mixture model provides more robust representation to estimate spikes and long tails of the traffic distribution compared to the well known statistical distributions such as Weibul, Lognormal, and General Pareto.

Many studies have found that whereas there exist measurement based traffic research by showing the cluster based traffic prediction [8], efficient call holding time approach for PSN traffic [9], daily and weekly traffic distribution models [10], multimodal fitting models for the long-tailed network traffic [11], the mixture model for the voice telephone call holding time [12], unlike the existing traffic models, the proposed model presents an accurate model through considering multi parameter case of the mixture of lognormals, as the call holding time is approximated by same model [13], [17].

This paper presents a preliminary investigation of input traffic intensity characteristics through the gateway E1 trunks to the EIN, leading up to the empirical modeling based on the lognormal mixtures of the long-tail distribution.

This article is structured as follows: Section 2 presents the existing emergency network structure, its telephone part components with emergency call delivering features. Section 3 presents data exploration of the existing network for the peak period performance. Section 4 describes the proposed mixed modeling approach with its Maximum Likelihood Estimate (MLE) techniques through the expectation maximization algorithm performance. In section 5, we demonstrate the
proposed model validation and results with the real data statistics, compared to the statistical models.

II. EMERGENCY CALL HANDLING PART OF PUBLIC SAFETY NETWORK

The IP based EIN is the state-of-the-art public emergency service provider in Mongolia [13]. There are four emergency telephone numbers in the country: 103 (ambulance), 102 (police), 101 (fire), and 105 (hazards, disaster). When dialing any one of these numbers from telecommunications networks (PSTN, GSM, CDMA) through PSTN, the emergency call is pushed/forwarded to an emergency call center agency desk of the EIN.

Fig. 1 illustrates the emergency telephone part of the network. The IP based voice call transfer system consists of mainly Automation Location (AL) server, Computer Telephone Integration (CTI) server, Interactive Voice Response (IVR) server and IP based Private Automatic Branch Exchange (IP PABX) which are located in the emergency network.

The incoming voice calls that come from PSTN optical E1 trunks transfer through IP PABX, go to CTI, and then to console desk. IP PABX connects to PSTN through E1 digital lines which provides unit transmission rate in 2.048 Mbps and provides the number of discrete communication channels. Thus, emergency users can get the benefits of a single network for voice, data, and advanced CTI features.

A. Components of Emergency Voice call Integration Part

CTI middleware can be defined as a technology for a capable of a comprehensively controlling, monitoring, and reporting of emergency call statistics which allows interaction between a traditional telephone system and a computer system.

CTI integration is based on a CTI Server which is linked to the IP PABX. The CTI Server can both receive event data from the IP PABX as well command/control the IP PABX. All major switches have a CTI link for connections to a CTI Server (e.g. Avaya, Mitel, NEC, Nortel, Siemens).

CTI provides the trace case call, trace caller’s phone number, trace call history of emergency reception desks. Calls set as emergency calls are responded at once using automatic answering function and call conversation starts using a emergency speaker phone. Routing calls intellectually to the idlest agency within one group supports to receive and handle report within the shortest times.

The IVR allows emergency customer to automate routine inbound queries that are repetitive in nature. Automation of such calls leads to a considerable reduction in the call processing time and enables one to service emergency customers efficiently.

B. Emergency Voice Call Handling Functions of IP PABX

IP PABX of the emergency call center empowers enterprise with automatic answering to emergency call, call recording, voice mail, voice call logging, call forwarding, simultaneous calls as well flexible and scalable web-based management. IP PABX provides these various services according to the main emergency call handling main functions: a call forwarding function, a simultaneous call function, a call pick up function, a call response management function, a call distributing function, and a call trunk tandem function.

III. DATA EXPLORATION AND INCOMING VOICE TRAFFIC FLUCTATION STATISTICS

We explore the real data from bundled digital trunks which connect into the main edge port of IP PBX of the EIN. The experiment period was a peak week including week days, weekends, and the biggest national holiday Naadam, starting from July 8, 2010 to July 14, 2010.

Traffic load is calculated in units of CCS, which represents one hundred seconds of a telephone conversation in a busy hour, then normalized. When a line carries one emergency call continuously for one hour it is said to carry one erlang (6 CCS) of traffic which is expressed by

\[
p = \mu \times t,
\]

where \(p\) - traffic in erlangs, \(\mu\) - mean call arrival rate (call per unit time), \(t\) - mean call holding time in same units as \(\mu\).

There are maximum emergency users who generate as high as 1.3 erlangs of ambulance class as well as light users generating less than 0.002 erlangs of hazard class for the period. The result indicates that the arrivals of emergency events have the cumulative imbalance in terms of their traffic to the system. The incoming traffic of emergency ambulance and police calls reflects the time change during the peak time, when the traffic is increased to the maximum level during an evening time in Fig. 2. The peak period average arrival rates of emergency ambulance, police, fire, and hazard calls are 6.0901, 5.2912, 0.5297, and 0.1981, respectively.
IV. PROPOSED TRAFFIC MODELING WITH LOGNORMAL MIXTURE

The mathematical and simulation part of this paper are supported by the previous study of the Bayesian parametric approach for the mixture model of lognormal distributions of the network which is described in an explicit manner [13].

When traffic \( x = (x_1, ..., x_n) \) are observed, we may consider the component indicators \( y = (y_1, ..., y_n) \) as missing like as in the usual mixture situation, so that \( z = (x, y) \) becomes the complete data [14].

When each mixture component is mapped to a lognormal call holding time distribution using priority probabilities, we notice that the \( n \) independent and identically distributed (i.i.d) call holding time observations \( x = (x_1, ..., x_n) \) come from a finite mixture of \( k \) lognormal components as follows:

\[
\phi_\theta(x) = \sum_{j=1}^{k} \pi_j \phi_{\theta_j}(x) = \sum_{j=1}^{k} \pi_j N(x | \mu_j, \gamma_j)
\]

where \( \phi_{\theta_j}(x) = (\phi_{\theta_j}(x), ..., \phi_{\theta_j}(x)) \) are the component lognormal densities, \( \theta_j = (\mu_{l=1}^j, \mu_{k=1}^j, \gamma_{j=1}^j, ..., \gamma_k) \) are the parameters, and \( \pi_j = (\pi_1, ..., \pi_k) \) are the component weights satisfying \( \sum_{j=1}^{k} \pi_j = 1 \). Therefore, the probability density function of the average traffic may be modeled by a mixture of \( k \) random variables of Gaussian densities on a logarithmic time scale:

\[
\phi_\theta(x) = \frac{1}{x \sqrt{2\pi}} \left[ \pi_1 \text{exp} \left( - (\ln x - \mu_1)^2 / 2\gamma_1^2 \right) + \ldots \right]
\]

\[
\pi_k \text{exp} \left( - (\ln x - \mu_k)^2 / 2\gamma_k^2 \right)
\]

We use Bayesian theorem to compute the expression for the distribution of the unobserved traffic:

\[
r_{ij} = \phi_j(x_i) = p_{ij}(y_{ij} = 1 | x_i; \theta, \pi)
\]

\[
\pi_j^i \phi_{\theta_j}(x_i) = \pi_j^i \phi_{\theta_j}(x_i)
\]

\[
\phi_\theta(x_i) = \sum_{j=1}^{k} \pi_j \phi_{\theta_j}(x_i)
\]

where \( \phi_\theta(x_i) \) is simply lognormal distribution evaluated at and likelihood function given \( x_i \) which is similar to the expression in the right side of (9) in [17]. Given \( \theta' \), a mixture of lognormal distribution is computed for each \( i \) and \( j \).

The expected value of the "complete data \( z = (x, y) \)" log-likelihood \( \phi_\theta(z) \) would be the iterative process. It is formed as a function of the estimate \( \theta \) from (11) in [17] and (3) where \( \theta' \) is the current value at iteration \( t \).

\[
q(\theta, \theta') = \sum_{j=1}^{k} \sum_{i=1}^{n} \ln \pi_j \phi_{\theta_j}(x_i) \phi_j(x_i)
\]

\[
= \sum_{j=1}^{k} \sum_{i=1}^{n} \ln(\pi_j \phi_{\theta_j}(x_i)) + \sum_{j=1}^{k} \sum_{i=1}^{n} [\ln \pi_j \ln x_i - \ln x_i \gamma_j]
\]

\[
- \frac{1}{2} \ln 2\pi - \frac{1}{2\gamma_j} [\ln x_i - \mu_j]^2 \phi_j(x_i)
\]

Taking derivatives with respect to these \( \pi_j, \mu_j, \gamma_j \) and equating them to zero, we define the new estimates \( \pi_j', \mu_j', \gamma_j' \) respectively as follows:

1. The weights of lognormal components:

\[
\pi_j^i = \frac{\sum_{i=1}^{n} \phi_j^i(x_i)}{n}
\]

2. The expected values \( \mu_j \) of the lognormal components :

\[
\mu_j^i = \frac{\sum_{i=1}^{n} \phi_j^i(x_i) \ln x_i}{\sum_{i=1}^{n} \phi_j^i(x_i)}
\]

3. The standard deviations \( \gamma_j \) of the components:

\[
\gamma_j^i = \frac{\sum_{i=1}^{n} \phi_j^i(x_i) (\ln x_i - \mu_j^i) (\ln x_i - \mu_j^i)^T}{\sum_{i=1}^{n} \phi_j^i(x_i)}
\]

B. Algorithm performance of the proposed mixed Lognormal distributions

The algorithm provides computational techniques for distributions that are almost completely unspecified [15], [16].

![Comparison on Emergency incoming traffic patterns](image-url)
The EM algorithm iteratively maximizes $Q(\theta, \theta^{'})$ by the following two steps: 1. E-step: compute $Q(\theta, \theta^{'})$

E-step calculates the expected value of the "complete data" log likelihood from (4) with respect to the unobserved traffic $y$ for all $i = (1,...n)$ and $j = (1,...k)$.

We calculate the expected value of the unobserved traffic data using the observed incomplete traffic data. Computing this expectation requires the posterior probability $\phi_j(x)$ as in (3) and $\pi_j$ for the parameter value $\theta^{'}(5)$.

2. M-step: set $\theta^{t+1} = \arg \max_{\theta} Q(\theta | \theta^{'})$

This step maximizes the expectation we performed in E-step. More specifically, M-step performs the first part containing $\pi_j$ and the second part containing $\theta^*$ iteratively from (4). We note that the parts are not related; we can maximize independently. These two steps are iterated as necessary until the saturation value of the expected complete log-likelihood. At the saturation value, the M-step performs the set of parameters $\theta$, otherwise we repeat the E-step for the next iteration. Although the iteration increases the marginal log-likelihood function, the EM algorithm uses a random restart approach to avoiding a local maximum of the observed data log-likelihood function. In the EM performance, we performed 121-372 iterations to find reasonable fits.

V. MODEL VALIDATION AND FITTING WITH REAL TRACE

In a practical fitting model, it is a challenge to estimate the spikes and tails of frequencies under the mixed model. When simulations of a mixture models are carried out, the model represents significant improvements to hold the spikes and tails of frequencies. As a model validation, visually, the plots including p.d.f, c.d.f, and c.c.d.f were depicted. Whereas p.d.f plots the probability of occurrence of the random variable under study, the c.d.f plots the probability that the random variable will not exceed specific values. The c.c.d.f presents the tail behavior of the model.

C. Results on the Emergency Ambulance Incoming Traffic

Fig. 3 shows the probability results of the distribution fittings, with the Lognormal-5 as the best fitted distribution. The model fitting distance $D_{\text{max}} = 0.00768755$ and error $\varepsilon = 0.46 \times 10^{-4}$, and component parameters $(\pi, \mu, \gamma)$ of the best mixed model are its main maximum likelihood estimation parameters (see Table. I). The weights of component Lognormal distributions are $0.20239335, 0.43246958, 0.15268270, 0.01735559, 0.03889879$ respectively.

We observe clearly that the normalized traffic p.d.f has extreme spikes around 0.3 erlangs and long tails around more than 1.2 erlangs, which shows the mixture of lognormal distributions that can be used to approximate accurately the empirical data with the random spikes and long tails as a best fitted model.

It is clear that the smallest $D_{\text{max}} = 0.00768755$ distance represents that the derived mixed lognormal distribution is better than Generalized Pareto ($D_{\text{max}} = 0.048$), simple Lognormal ($D_{\text{max}} = 0.061$), and Weibull ($D_{\text{max}} = 0.07$) distributions according to the Kolmogorov Smirnov (K-S) test.

The traffic p.d.f, which is derivative of the c.d.f; the highest probability intervals are easily recognized as the peaks in the p.d.f in Fig. 3(a), while c.d.f value is compared to the difference between empirical data and fitting models; the steeper the slope of the c.d.f, the higher the probability of values. Fig. 3(b) shows that the mixed lognormal-5 distribution approximates the empirical data quite well compared to General Pareto distribution, which the second ranked standard fitting model by its $D_{\text{max}}$ and error values of the K-S test. The General Pareto distribution is fitted with the scale parameter $\sigma = 0.21$, shape parameter $\varepsilon = 0.11$, mean 0.265, standard deviation 0.267, and coefficient of variation $cv = 1$.

Although the c.d.f fitting looks perfect, the discrepancy in the set of distributions is clear, while the tail values are not seen clearly well. Therefore, the c.c.d.f on a log-log scale (Fig. 3(c)) presents the tail fitting behavior how the mixed model is the most accurate model.
D. Results on the Emergency Police Incoming Traffic

Fig. 4 presents the emergency policy incoming traffic along with its best fitted mixture model of five lognormal components. The mixed model is significant closer model compared to the General Pareto approximation.

In the traffic with up to 0.65 erlangs for the 98 percent of the full capacity. In the Fig.4, we show the MLE values with the best fitting parameters $D_{max} = 0.00932$ and $\epsilon = 0.39 \times 10^{-5}$ for the policy call traffic, respectively. The weights $\pi_j = (1...5)$ of component lognormal distributions are 0.04214847, 0.13442508, 0.71282226, 0.04436757, and 0.06623662 respectively. In the fire traffic with up to 0.06 erlangs for the 98 percent of the full capacity.

The standard General Pareto model is considered as a good matched model for the empirical data, except the long tail and spikes. The General Pareto distribution is fitted with the scale parameter $\sigma = 0.147$, shape parameter $k = 0.08$, the mean 0.175, standard deviation $\sigma\gamma =1$. Due to the comparably large scale parameter, it does not have a long-tailed behavior. It clearly cannot fit accurately (Figure 4(c)).

E. Results on the Emergency Fire Incoming Traffic

In Fig.5, we show the values of $D_{max} =0.00652$ and $\epsilon = 0.39 \times 10^{-5}$ for the fire call traffic. The weights of components are 0.07472365, 0.17832745, 0.41007972.

In a practical matter of difficulties of tails to distinguish for a model, to illustrate the tail fitting more clearly, we test the c.c.d.f that the large variability of the most fitted model is inherited by the approximating lognormal-5 model, so that the best fit with the mixed one is done. Fig. 5(c) shows the fitting of the tail part of frequencies.

Whereas the medium fire call related incoming load is generated by 0.01 erlangs only, the maximum traffic is generated less than 0.08 erlangs within the peak week.
The traffic of the emergency hazard incoming call is accurately modelled by a mixture of three lognormals. The MLE parameters (all $\pi, \mu, \gamma$) of components, the $D_{\text{max}} = 0.0364377$ value, and $\epsilon = 4.2 \times 10^{-5}$ of the fitting of a mixture of Lognormal-3 through the K-S test. The weights of components are 0.07861348, 0.68476605, and 0.23662047 respectively. We show that the smaller $D_{\text{max}} = 0.0364377$ distance represents that the derived mixed model is better than Generalized Pareto ($D_{\text{max}} = 0.06$), simple Lognormal ($D_{\text{max}} = 0.07$), and Weibull ($D_{\text{max}} = 0.09$) distributions.

The Weibull distribution is fitted with the scale parameter $\sigma = 0.47$, the shape parameter $\alpha = 0.89$, the mean 0.0049, the standard deviation 0.0055, and the $cv = 1.114$, respectively. The tail portion does not fit accurately by these models compared to the best fitted model of Lognormal-3.

G. MLE parameters of Mixture Model

If components of the traffic have lognormal distribution, then the traffic is accurately modeled by the mixed lognormal model.

By comparing the MLE parameter results, we show that moving from simple standard distributions model such as general Pareto to a mixture of Lognormals leads us to more accurate results and less error.

We use comparing the histogram against the Lognormal mixture model, the same plot in log-log scale and quantile-quantile plot [16]. Kolmogorov-Smirnov (K-S) tends to be more sensitive near the center of the distribution than at the tails. Due to this limitation above, we prefer to use the tail behavior pattern through the c.c.d.f plot and, thus, it implies again the accurate validation of the model.

V. Conclusion

For the incoming traffic of the emergency calls, the long-tailed Pareto model provides a better approximation than the common candidates - simple Lognormal and shifted Lognormal distributions; consequently, the Pareto model is compared with the mixture model with simulations were carried out through the parametric approach. Then, it is proved that the incoming traffic of the emergency calls is modelled accurately by a mixture of Lognormal distributions with two or more parameters that can be used to approximate the whole traffic domain, including the spikes and long tails for peak period of the public safety network. The model selection and validation by the comparisons of visual curves, minimum distances, and error values of K-S test presented that the mixed model is the most accurate model. Furthermore, we detailed the tail fitting analysis for the more clear validation. We computed the mixing components of the heavy-tailed traffic model for the network. The mixed model may be the most appropriate model to catch unusual occurrences for peak and long term traffic patterns of a network.

ACKNOWLEDGMENT

Financial support from the International Fulbright Science and Technology PhD program of USA government is gratefully acknowledged. We thank the LG corporation members of EIN project under the support of Mongolia and Korea government for the collaboration during the implementation period of EIN.

REFERENCES