A Stochastic Network Calculus Approach for Traffic Modeling and Performance Analysis in the Internet of Things

Xuan Wu, Xiaorong Zhu and Hongbo Zhu

Nanjing University of Posts and Telecommunications

Nanjing, China

xwun_n@163.com, xrzhu@njupt.edu.cn, zhuhb@njupt.edu.cn

Abstract— **The Internet of Things (IoT) exhibits considerably different traffic patterns than human-type communication. Thus, it is necessary to study new traffic models and analyze the network performances in this context. This paper first abstracts the transmission of Machine-to-Machine (M2M) traffic through the backhaul network as a model of tandem nodes, and uses Coupled Markov Modulated Poisson Process (CMMPP) to model M2M traffic, which can reflect the spatial and temporal correlation of traffic in IoT. Following that, the arrival curve of CMMPP is derived and the service curve of the transmission network is analyzed by using stochastic network calculus (SNC). On this basis, the end-to-end delay bound of M2M traffic is accordingly deduced. Numerical analyses are conducted to verify the accuracy of the results in the end.**

Keywords—Internet of Things; Stochastic Network Calculus; CMMPP; End-to-end delay bound

I. INTRODUCTION

IoT is expected to significantly increase in future wireless networks, in which M2M communication plays a key role to make IoT effective and feasible, including smart home, intelligent transportation, smart grid and video surveillance. Conventional mobile networks that have been optimally designed for H2H services should accommodate this new challenges [1]. In such a case, the study of M2M traffic model and network performance has a guiding significance for optimizing network configuration and resource allocation.

At present, the academic research of M2M is still in primary stage. There are only a few studies on the mathematical modeling of M2M traffic and the performance analysis based on specific network model [2]. The 3rd Generation Partnership Project (3GPP) proposes two kinds of reference models of M2M traffic in [3], which assumes the inter-arrival time of traffic to follow uniform distribution and beta distribution, respectively for describing non-synchronous and synchronous fashion of M2M traffic. Based on the 3GPP traffic models, [4] and [5] respectively put forward the simulation models satisfying the compound Poisson process modulated by beta distribution and CMMPP to model M2M traffic. They become general simulation models because of their accuracies of modeling. It can be seen that M2M communication is developing rapidly, but there is still no standard model of M2M traffic and it is shortage of theoretical analysis at this stage.

Queuing theory is often used as one of the mathematical tools in network performance analysis. However, unique flow specification and service requirements in modern communication networks often make its adoption difficult [6]. SNC is another theory to deal with such problems, having its foundation on the min-plus convolution. It not only can be used to study the system model in a more accurate manner but also can be used to predict and analyze the network performance.

In this paper, we use SNC theory to analyze the typical M2M communication network architecture. The performance evaluation of a network model relies on two main concepts of SNC, stochastic arrival curve and stochastic service curve .The former presents the characteristic of the arrival data, and the latter presents the service capacity that a system guarantees to the flow [7]. Finally, based on these two cores, the end-to-end performance of M2M traffic can be deduced, which provides the basis for the development and deployment of the M2M communication in IoT.

The rest of the paper is organized as follows. In Section II, we describe the system model of M2M communication network. In Section III, we analyze the arrival flow of M2M traffic and give analytical results of the end-to-end performance using SNC. In Section IV, we validate the results with numerical analyses and make a discussion about them. We conclude this paper in Section V.

II. SYSTEM MODEL

A. Network Scenario

Fig.1 depicts the network architecture of M2M communication. Here, the access points are responsible for aggregating all the upload information from various Machine-Type Devices (MTDs), and forwarding the data to the destination M2M server through the carrier network. And then the processed data are reported to the subscribers or MTDs.

In this paper, we choose the backhaul network as the carrier network and call the M2M traffic that flows through the backhaul network as through-traffic. At the same time, the non-M2M traffic competing services with M2M traffic is called cross-traffic. This paper focuses on the end-to-end performance of through-traffic and provides the basis for judging whether the existing network architecture can guarantee the quality of service (QoS) of M2M communication.

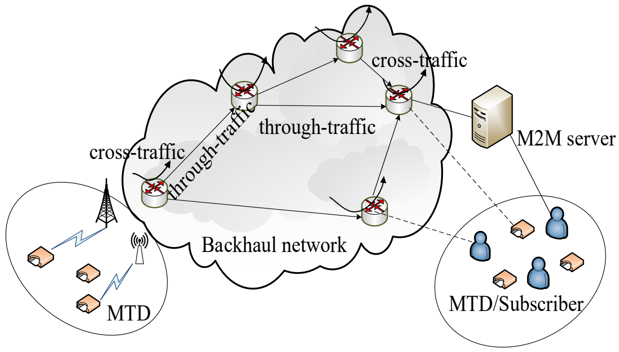

Fig.1. Network architecture of M2M communication

B. CMMPP Model

It should be pointed out that an appropriate traffic model enabling performance evaluation is of great concern. The distinguishing features of M2M communication include uplink-dominant traffic, large scale of MTDs, and various types of applications. In addition, some MTDs generate traffic in a synchronous fashion both in time and space, that is, the event-driven M2M traffic will affect other MTDs around, so that the aggregated traffic generated by MTDs within a certain area shows a synchronous mode. However, two models proposed by 3GPP cannot capture this character.

In order to model this mode, we consider the CMMPP in [5] as M2M traffic model for research. *N* MTDs correspond to *N* MMPP model, and they can be coupled by using unidirectional links influenced by a background process $\varphi(t)$, $\varphi(t) \in [0, 1]$. Further, each MTD is assigned a constant parameter $\delta_n \in [0, 1]$, so the n-th MTD yields $\varphi_n(t) = \delta_n \cdot \varphi(t)$ to associate with the background process. The term δ_n defines the spatial correlation between the machines, and the term $\varphi(t)$ can be seen as the temporal correlation between the machines.

The state transition matrix $\mathbf{P}_n(t)$ of the n-th machine at time *t* is a the convex combination of $\mathbf{P}_1$ and $\mathbf{P}_2$, written as

$$\mathbf{P}_n(t) = \varphi_n(t) \cdot \mathbf{P}_1 + (1 - \varphi_n(t)) \cdot \mathbf{P}_2 \tag{1}$$

At present, we only consider a two-state model, where the states represent regular and alarm operation, respectively. So the two matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ can be defined as

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathbf{P}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \tag{2}$$

The coordinated case $\mathbf{P}_1$ represents that an alarm is triggered in one time slot and then the machines return to regular operation. On the contrarily, the uncoordinated case $\mathbf{P}_2$ represents that an alarm is never triggered.

In addition, to match with the second model in 3GPP models, we add $\varphi(t) = f_T(t; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{t^{a-1}(T-t)^{b-1}}{T^{a+b-1}}, 0 \le t \le T$, that is beta distribution. The shape parameters a and b can be adjusted to responsive application scenarios. The mean of the beta distribution is $\frac{aT}{(a+b)}$, and when n is a positive integer, $\Gamma(n) = (n-1)!$

III. Analysis of Network Model Based on CMMPP

A. Stochastic Arrival Curve of M2M Traffic

The determination of the stochastic arrival curve is the first step in applying SNC theory. We denote $A(s,t)$ the cumulative number of bits arriving at a system in an interval $[s,t)$, so the traffic-amount-centric (t.a.c) model is represented as

$$P\{A(s,t) - \alpha(t-s) > x\} \le f(x) \tag{3}$$

where $\alpha(\tau)$ is the stochastic arrival curve of the traffic model and it describes the stochastic upper bound of the arrivals. Clearly, it is a generalized increasing function, denoted as $\alpha(\tau) \in F$. The above can be simply recorded as $A(t) \sim < f, \alpha >$.

In this paper, we choose the t.a.c model for reference and construct the stochastic arrival curve of the CMMPP model based on the moment generating function(MGF) in [8].

We approach the problem as follows. First, we consider the arrival process is limited by $(\sigma(\theta), \rho(\theta))$, i.e. $\frac{1}{\theta}\log E[e^{\theta A(s,s+t)}] \le \rho(\theta) \cdot t + \sigma(\theta)$, or equivalent to $E[e^{\theta A(s,s+t)}] \le e^{\theta[\rho(\theta) \cdot t + \sigma(\theta)]}$. And then the following inequality can be derived from the Chernoff boundary

$$\begin{aligned} P\{A(s,t) - \alpha(t-s) > x\} &\le P\{e^{\theta[A(s,t) - \alpha(t-s)]} > e^{\theta x}\} \\ &\le e^{-\theta x} E[e^{\theta(A(s,s+t) - \alpha(t))}] \\ &\le e^{-\theta x} e^{\theta[\rho(\theta) \cdot t + \sigma(\theta) - \alpha(t)]} \\ &\le e^{-\theta x} \end{aligned} \tag{4}$$

The above inequality shows that any $\alpha(t)$ satisfying $\alpha(t) \ge \rho(\theta) \cdot t + \sigma(\theta) \ge \frac{1}{\theta}\log E[e^{\theta A(s,t)}]$ is a stochastic arrival curve. In such a case, $f(x)$ equals to $e^{-\theta x}$, and θ is a free parameter greater than 0 .

The n-th arrival traffic generated by a single MTD can be calculated as the total length of the cumulative packets, i.e. $A_n(t) = \sum_{i=1}^{M(t)} L_n^i$. Here, $M(t)$ is the total number of arrival packets between times 0 and t, and $L_n^1, L_n^2, \cdots, L_n^{M(t)}$ are the lengths of each packet which are independent and identically distributed. We assume that the arrival rate of traffic under two states is λ_r and λ_a, so the MGF of $A_n(t)$ is

$$\begin{aligned} \Omega_{A_n}(\theta_n, t) &= E\left[e^{\theta_n A_n(t)}\right] = E\left[e^{\theta_n \sum_{i=1}^{M(t)} L_n^i}\right] \\ &= \sum_m E[e^{\theta_n(L_n^1 + L_n^2 + \ldots + L_n^m)}] \cdot P\{M(t) = m\} \\ &= \sum_m \{E[e^{\theta_n L_n^1}]\}^m \cdot P\{M(t) = m\} \end{aligned} \tag{5}$$

We first calculate the probability density function of $M(t)$

$$P\{M(t) = m\} = \mathbf{\Lambda} e^{\mathbf{Q}_n t} diag\left(\frac{(\lambda_r t)^m}{m!} e^{-\lambda_r t}, \frac{(\lambda_a t)^m}{m!} e^{-\lambda_a t}\right) \mathbf{e} \tag{6}$$

where $\mathbf{e}=[1\ 1]^T$. According to (1)-(2) we arrive at

$$\mathbf{P}_n(t) = \begin{bmatrix} 1-\varphi_n(t) & \varphi_n(t) \\ 1 & 0 \end{bmatrix} \tag{7}$$

$$\mathbf{\Lambda}_n(t) = \left[\mathbf{\Lambda}_{n,1}(t)\ \ \mathbf{\Lambda}_{n,2}(t)\right] = \left[\frac{1}{\varphi_n(t)+1}, \frac{\varphi_n(t)}{\varphi_n(t)+1}\right] \tag{8}$$

Thus, the minimum generation matrix $\mathbf{Q}_n$ of continuous

MMPP can be found as

$$\mathbf{Q}_n = \begin{bmatrix} -\varphi_n(t) & \varphi_n(t) \\ 1 & -1 \end{bmatrix} \tag{9}$$

Then we have

$$\begin{aligned} e^{\mathbf{Q}_n t} &= \mathbf{I} + \sum_{i=1}^{\infty} \frac{\mathbf{Q}_n{}^i}{i!} t^i \\ &= \mathbf{I} - \frac{1}{\varphi_n+1}(e^{-(\varphi_n+1)t} - 1)\mathbf{Q}_n \\ &= \begin{bmatrix} 1 + \frac{e^{-(\varphi_n+1)t}-1}{\varphi_n+1}\varphi_n & -\frac{e^{-(\varphi_n+1)t}-1}{\varphi_n+1}\varphi_n \\ -\frac{e^{-(\varphi_n+1)t}-1}{\varphi_n+1} & 1 + \frac{e^{-(\varphi_n+1)t}-1}{\varphi_n+1} \end{bmatrix} \\ &= \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix} \end{aligned} \tag{10}$$

Use $\mathbf{\Lambda}_n(t)$ at time $t = \frac{aT}{(a+b)}$ as its stationary distribution and denote $\frac{aT}{(a+b)}$ as c. Substituting it and (10) into (6), and assuming that the packet length is fixed to l, we can readily compute the MGF in (5) to be

$$\begin{aligned} \Omega_{A_n}(\theta_n, t) &= \sum_m e^{\theta_n l m} \cdot P\{M(t) = m\} \\ &= \sum_m e^{\theta_n l m} \cdot \Big(\frac{(\lambda_r t)^m}{m!} \cdot e^{-\lambda_r t} \cdot \frac{v_1 + v_3\varphi_n(c)}{\varphi_n(c)+1} \\ &\quad + \frac{(\lambda_a t)^m}{m!} \cdot e^{-\lambda_a t} \cdot \frac{v_2 + v_4\varphi_n(c)}{\varphi_n(c)+1}\Big) \\ &= e^{\lambda_r t(e^{\theta_n l}-1)} \frac{v_1 + v_3\varphi_n(c)}{\varphi_n(c)+1} + e^{\lambda_a t(e^{\theta_n l}-1)} \frac{v_2 + v_4\varphi_n(c)}{\varphi_n(c)+1} \\ &= \frac{1}{\varphi_n(c)+1}\left\{e^{\lambda_r t(e^{\theta_n l}-1)}(1-j) + e^{\lambda_a t(e^{\theta_n l}-1)}(\varphi_n(c)+j)\right\} \end{aligned} \tag{11}$$

The results of the above can be obtained by setting $\frac{1-e^{-(\varphi_n+1)t}}{\varphi_n+1}\big(\varphi_n - \varphi_n(t)\big) = j$ in the last second step. Until now, we can finally get the stochastic arrival curve of the arrival traffic

$$\begin{aligned} \alpha_{n,\theta_n} &= \frac{1}{\theta_n}\log\Omega_{A_n}(\theta, t) \\ &= \frac{1}{\theta_n}\{\log(e^{\lambda_r t(e^{\theta_n l}-1)}(1-j) + e^{\lambda_a t(e^{\theta_n l}-1)}(\varphi_n(c)+j)) \\ &\quad -\log(\varphi_n(c)+1)\} \end{aligned} \tag{12}$$

and $f_n = e^{-\theta_n x}$, θ_n is greater than 0, so we have $A_n(t) \sim < f_n, \alpha_{n,\theta_n} >$.

The method proposed in this paper is also applicable to the CMMPP model with more states. It only needs to transfer $\mathbf{P}_1$ and $\mathbf{P}_2$ to the $n \times n$ order matrix corresponding to n states.

B. Stochastic Service Curve

SNC theory provides one way to describe the stochastic lower bound of the service offered by a system, which is the stochastic service curve introduced in [9].

Definition1 (Stochastic Service Curve) A departure curve $A^*(t)$ is the cumulative amount of traffic that have departed system S by time t, if for all $t, x \geq 0$, there holds

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq \mathrm{g}(x) \tag{13}$$

The system S is said to have a stochastic service curve $\beta \in F$ with bounding function $\mathrm{g} \in \bar{F}$, denoted by $S \sim < g, \beta >$. Here, the operator $\otimes$ represents the min-plus convolution, defined as

$$A \otimes \beta(t) = \inf_{0 \leq s \leq t}\{A(s) + \beta(s,t)\} \tag{14}$$

The service rate of routers is fixed to R due to its relatively constant service capability. Further, we assume routers are work-conserving servers and work with non-preemptive priority scheduling policy, in which each queue forwards data using FCFS. So each router u offers the overall service curve $\beta_u = R_u t$, corresponding to $S_u \sim < 0, \beta_u >$.

We assume that there are N_d flows of through-traffic, accompanied by N_c flows competing with them for service at each router, and the priority of cross-traffic is higher than through-traffic. Here, treat H2H traffic following the Poisson distribution as cross-traffic. The stochastic arrival curve of the i-th flow of cross-traffic satisfies $A_i^\mathrm{c}(t) \sim < f_i^\mathrm{c}, \alpha_{i,\theta_i}^\mathrm{c} t >$, where $f_i^\mathrm{c} = e^{-\theta_i x}$, $\alpha_{i,\theta_i}^\mathrm{c} = \frac{\lambda}{\theta_i}(e^{\theta_i} - 1)t$ [10]. Then, the stochastic arrival curve of aggregated cross-traffic can be

$$A_N^\mathrm{c}(t) \sim < f_N^\mathrm{c}, \alpha_{N,\theta_N}^\mathrm{c} > \tag{15}$$

where $f_N^\mathrm{c} = f_1^\mathrm{c} \otimes f_2^\mathrm{c} \cdots \otimes f_{N_\mathrm{c}}^\mathrm{c}, \alpha_{N,\theta_N}^c = \sum_{i=1}^{N_\mathrm{c}} \alpha_{i,\theta_i}^\mathrm{c}$. Similarly, N_d flows of aggregated through-traffic satisfy $A_N^\mathrm{d}(t) \sim < f_N^\mathrm{d}, \alpha_{N,\theta_N}^\mathrm{d} >$, where $f_N^\mathrm{d} = f_1^\mathrm{d} \otimes f_2^\mathrm{d} \cdots \otimes f_{N_\mathrm{c}}^\mathrm{d}, \alpha_{N,\theta_N}^\mathrm{d} = \sum_{i=1}^{N_\mathrm{d}} \alpha_{i,\theta_i}^\mathrm{d}$.

Theorem1 (Leftover Service Curve) Consider a system S with input $A_1(t) \sim < f_1, \alpha_1 >$, $A_2(t) \sim < f_2, \alpha_2 >$ and suppose $S \sim < g, \beta >$, then the stochastic service curves offered to $A_1(t)$ and $A_2(t)$ satisfy

$$\begin{aligned} S_1 &\sim < f_2 \otimes g, \beta - \alpha_2 > \\ S_2 &\sim < f_1 \otimes g, \beta - \alpha_1 > \end{aligned} \tag{16}$$

Considering of theorem1 and the service model $S_u \sim < 0$, $\beta_u >$ provided by a router, we can get that the service offered to N_d flows by router u satisfies $S_u^\mathrm{d} \sim < f_N^\mathrm{c}, R_u t - \alpha_{N,\theta_N}^\mathrm{c} >$. Let $f_N^\mathrm{c} = g_u^\mathrm{d}, R_u t - \alpha_{N,\theta_N}^\mathrm{c} = \beta_u^\mathrm{d}$, so that $S_u^\mathrm{d} \sim < g_u^\mathrm{d}, \beta_u^\mathrm{d} >$. The concatenation property using min-plus convolution simplifies the service curve of U routers provided to through-traffic

$$S^\mathrm{NET} \sim < g^\mathrm{NET}, \beta^\mathrm{NET} > \tag{17}$$

where $\beta^\mathrm{NET}(t) = \beta_1^\mathrm{d} \otimes \beta_{2,-\theta'}^\mathrm{d} \otimes \cdots \otimes \beta_{U,-(U-1)\theta'}^\mathrm{d}(t)$, $g^\mathrm{NET}(x) = g_{1,\theta_1'}^\mathrm{d} \otimes g_{2,\theta_2'}^\mathrm{d} \cdots \otimes g_{U,\theta_U'}^\mathrm{d}(x)$, $\beta_{u,-(u-1)\theta'}^\mathrm{d}(t) = \beta_u^\mathrm{d}(t) - (n-1)\theta'$, $g_{u,\theta_u'}^\mathrm{d}(x) = g_u^\mathrm{d}(x) + \frac{1}{\theta_u'}\int_x^\infty g_u^\mathrm{d}(y)dy$ and all free parameters θ', θ_1', $\cdots$, θ_U' are greater than 0.

C. Analysis of End-to-end Delay

Analysis of the end-to-end performance measures of networks helps to determine whether the specific network itself can satisfy the QoS of different services and to provide the basis for performance evaluation and optimal design of the backhaul network. This paper mainly focuses on the end-to-end delay. [11] gives the characterization of delay as follows.

Definition2 (Delay) A system's delay at time t is defined as:

$$D(t) = \inf\{i: A(t) \le A^*(t+i)\} \tag{18}$$

Theorem2 (Delay Bound) Consider a system S with input$A(t)$. Suppose the input has $A(t) \sim < f, \alpha >$ and server S provides a service model $S \sim < g, \beta >$. Then, no matter whether the arrival process is independent of th service process or not, the delay $D(t)$ is bounded by

$$P\{D(t) > w(\alpha + x, \beta)\} \le f \otimes g(x) \tag{19}$$

where

$$w(\alpha + x, \beta) = \sup_{s \ge 0}\{\inf\{\tau \ge 0: \alpha(s) + x \le \beta(s+\tau)\}\} \tag{20}$$

it's the maximum horizontal distance between $\alpha + x$ and β.

Based on the stochastic arrival curve of CMMPP and the stochastic service curve of routers in the backhaul network, the end-to-end delay bound of M2M traffic in IoT can be deduced as bellow.

First, according to (20), (19) can be converted to

$$P\{D(t) > x\} \le f \otimes g(\inf_{s \ge 0}[\beta(s) - \alpha(s-x)]) \tag{21}$$

And then applying min-plus convolution and substituting $A_N^{\mathrm{d}}(t) \sim < f_N^{\mathrm{d}}, \alpha_{N,\theta_N}^{\mathrm{d}} >$ and $S^{\mathrm{NET}} \sim < g^{\mathrm{NET}}, \beta^{\mathrm{NET}} >$ into (21), we can get

$$P\{D(t) > x\} \le \inf_{0 \le k \le x}\{f_N^{\mathrm{d}}(\inf_{s \ge 0}[\beta^{\mathrm{NET}}(s) - \alpha_{N,\theta_N}^{\mathrm{d}}(s-x)] - k) + g^{\mathrm{NET}}(k)\} \tag{22}$$

Here,

$$\begin{aligned} f_N^{\mathrm{d}}(x) &= f_1^{\mathrm{d}} \otimes f_2^{\mathrm{d}} \otimes \cdots \otimes f_{N_{\mathrm{c}}}^{\mathrm{d}}(x) = e^{-\theta_1 x_1} \otimes \cdots \otimes e^{-\theta_{N_{\mathrm{d}}} x_{N_{\mathrm{d}}}} \\ &= \inf_{x_1 + \cdots + x_{N_{\mathrm{d}}} = x} \sum_{i=1}^{N_{\mathrm{d}}} e^{-\theta_i x_i} = N_{\mathrm{d}} e^{\frac{-\theta x}{N_{\mathrm{d}}}} \end{aligned} \tag{23}$$

$$\begin{aligned} \beta^{\mathrm{NET}}(t) &= \beta_1^{\mathrm{d}} \otimes \beta_{2,-\theta'}^{\mathrm{d}} \otimes \cdots \otimes \beta_{U,-(U-1)\theta'}^{\mathrm{d}}(t) \\ &= (R_1 t - \alpha_{N,\theta_N}^{\mathrm{c}}) \otimes (R_2 t - \alpha_{N,\theta_N}^{\mathrm{c}} - \theta' t) \otimes \cdots \otimes (R_U t - \alpha_{N,\theta_N}^{\mathrm{c}} - (U-1)\theta' t) \\ &= \{\min_{1 \le u \le U}[R_u - \sum_{i=0}^{N_{\mathrm{c}}} \frac{\lambda}{\theta_i}(e^{\theta_i} - 1) - (u-1)\theta']\}t \\ &= rt \end{aligned} \tag{24}$$

Similarly,

$$\begin{aligned} g^{\mathrm{NET}}(x) &= g_{1,\theta_1'}^{\mathrm{d}} \otimes g_{2,\theta_2'}^{\mathrm{d}} \otimes \cdots \otimes g_{U,\theta_U'}^{\mathrm{d}}(x) \\ &= N_{\mathrm{c}}(1 + \tfrac{N_{\mathrm{c}}}{\theta\theta_1'})e^{-\frac{\theta}{N_{\mathrm{c}}}x} \otimes \cdots \otimes N_{\mathrm{c}}(1 + \tfrac{N_{\mathrm{c}}}{\theta\theta_U'})e^{-\frac{\theta}{N_{\mathrm{c}}}x} \\ &= N_{\mathrm{c}} \cdot U \cdot (1 + \tfrac{N_{\mathrm{c}}}{\theta\theta'})e^{-\frac{\theta}{UN_{\mathrm{c}}}x} \end{aligned} \tag{25}$$

So the end-to-end delay bound of the M2M traffic is

$$P\{D(t) > x\} \le \inf_{0 \le k \le x}\{N_{\mathrm{d}} e^{\frac{-\theta}{N_{\mathrm{d}}}(\inf_{s \ge 0}[rs - \alpha_{N,\theta_N}^{\mathrm{d}}(s-x)] - k)} + N_{\mathrm{c}} \cdot U \cdot (1 + \tfrac{N_{\mathrm{c}}}{\theta\theta'})e^{-\frac{\theta}{UN_{\mathrm{c}}}k}\} \tag{26}$$

IV. Analysis of Numerical Results

A. Validation and Settings

In order to validate the derived results of the performance of M2M traffic, and to study the influencing factors of the end-to-end delay bound in the backhaul network, the relevant numerical analyses will be conducted.

Some important parameters are shown in table I. We suppose the numbers of flows of through-traffic and cross-traffic are the same, $N_{\mathrm{d}} = N_{\mathrm{c}} = 50$. TR 37.868 advices that the value of $a = 3$ and $b = 4$ are assumed in the study. In addition, the appropriate values of free parameters should be chosen for greater multiplexing gain. Here, we assume that $\theta_n = \theta = 1$, $\theta_n' = \theta' = 3149900$.

TABLE I. Network Settings

Parameters	Values
Transmission rate of router R_u	1Gbps~1.5Gbps
Arrival rate of M2M traffic at regular λ_{r}	0.0033pkt/s
Arrival rate of M2M traffic at alarm λ_{a}	10pkt/s
Length of M2M packet l	120Bytes
Parameters of beta distribution a, b, T	$a = 3, b = 4,$ $T = 10s$
Arrival rate of Poisson flow λ	0.8Mb/s

B. Results and Analysis

Fig.2 clearly reveals that the amount of cross-traffic is greatly more than that of through-traffic (M2M traffic). Although the stochastic arrival curve is the stochastic upper bound of traffic, not the actual arrival traffic, but it also confirms that M2M traffic is small data from the side.

Fig.3 analyzes the impact of three influencing factors on the end-to-end delay bound respectively. All three figures in Fig.3 illustrate that the violation probability decreases with the increasing of end-to-end delay bound.

Specifically, Fig.3 (a) shows that in the case of a certain violation probability, the end-to-end delay increases gradually

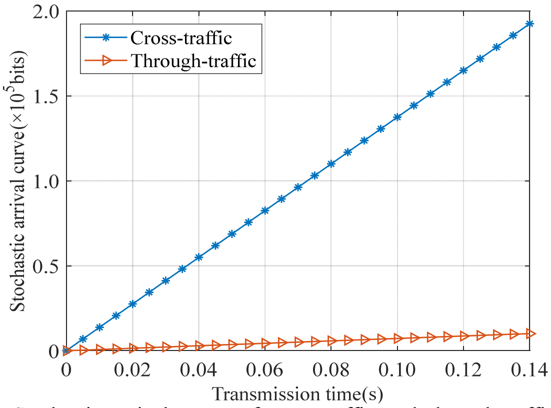

Fig.2. Stochastic arrival curve of cross-traffic and through-traffic during transmission time

with more numbers of routers. Therefore, in order to minimize the number of nodes in transmission path for lower delay and better real-time transmission, routing decisions should be considered in network planning.

It is seen in the Fig.3 (b) that the end-to-end delay bound increases when we fix the number N_{d} and change N_{c} from 30 to 90. This is because the cross-traffic has higher priority access to service due to the non-preemptive priority scheduling policy, resulting in increased delay of through-traffic. At the same time, as the total flows becomes more, the increasing magnitude of the end-to-end delay is gradually smaller. This fact shows that SNC can reflect statistical multiplexing of traffic in networks and make analysis more accurate.

Fig.3 (c) illustrates the relationship between the violation probability and the end-to-end delay bound of M2M traffic under different scenarios. We set $\delta_n$with two different values as: $\delta^{(\mathrm{H})}$=0.8 and $\delta^{(\mathrm{L})}$=0.2.Then, three scenarios have been studied: (1)all machines have high correlation, i.e., $\delta_n = \delta^{(\mathrm{H})}, \forall n$; (2) $\delta_n = \delta^{(\mathrm{L})}, \forall n$; (3) $\delta_n = \delta^{(\mathrm{H})}$ for half of the machines and $\delta_n = \delta^{(\mathrm{L})}$for the second half. We can see that scenario1 with high correlation has biggest delay. Because at this time, the probability that all machines are in alarm state is high, which can brings more traffic resulting in longer delay.

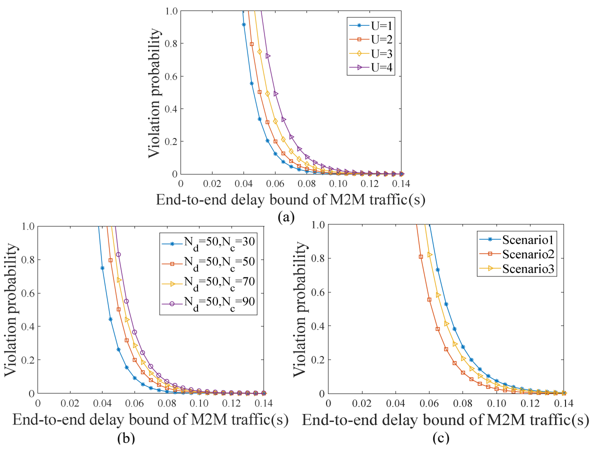

Fig.3. Relationship between violation probability and end-to-end delay bound of M2M traffic based on three influencing factors

And the delays of scenario3 and scenario2 reduce in turn reasonably.

V. Conclusion

In this paper, we take full advantage of SNC theory to figure out the stochastic arrival curve of CMMPP model considering the characteristics of M2M traffic. At the same time, the stochastic service curve is analyzed based on the abstracted model of tandem nodes, and then the end-to-end delay bound of M2M traffic is finally derived in the context of joint M2M and H2H traffic. The results show that the method is reasonable and can provide references to evaluate whether the network can guarantee the specific service for M2M traffic or not.

Acknowledgment

This work was supported by Natural Science Foundation of China (61372125), National Science and Technology Major Project of Ministry of Science and Technology of China (2017ZX03001008), and Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (16KJA510005).

References

[1] D. Niyato, W. Ping, and I. K. Dong. “Performance modeling and analysis of heterogeneous machine type communications”. *IEEE Transactions on Wireless Communications*, vol. 13, pp. 2836-2849, May 2014.

[2] S. A. Alqahtani. “Analysis and modelling of power consumption-aware priority-based scheduling for M2M data aggregation over long-term-evolution networks”. *IET Communications*, vol. 11, pp. 177-184, Jan. 2017.

[3] 3GPP. “Study on RAN improvements for machine-type communications”. Technical Report 37.868, Sep. 2012.

[4] C. D. Paiva, R. D. Vieira, and Y. M. Sail. “Random access capacity evaluation with synchronized MTC users over wireless networks”. *IEEE Vehicular Technology Conference*, Ottawa, Ontario, Canada, May 2011.

[5] M. Laner, P. Svoboda, N. Nikaein, et al. “Traffic models for machine type communications”. *Tenth International Symposium on Wireless Communication Systems*, Ilmenau, Germany, Jan. 2013.

[6] Jiang, Yuming. "Stochastic network calculus for performance analysis of Internet networks —An overview and outlook." *International Conference on Computing, NETWORKING and Communications IEEE*, 2014:638-644.

[7] J. Huang, Y. Sun, Z. Xiong, et al. “Modeling and Analysis on Access Control for Device-to-Device Communications in Cellular Network: A Network-Calculus-Based Approach”. *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 1615-1626, Mar. 2016.

[8] M. Fidler and A. Rizk. “A Guide to the Stochastic Network Calculus”. *IEEE Communications Surveys & Tutorials*, vol. 17, pp. 92-105, Jan. 2017.

[9] Y. Liu and Y. Jiang. “Stochastic Network Calculus”. London, Springer, 2009.

[10] S. Mao, S. S. Panwar, S. Mao, et al. “A survey of envelope processes and their applications in quality of service provisioning”. *IEEE Communications Surveys & Tutorials*, vol. 8, pp. 2-20, Feb. 2006.

[11] Z. Li, Y. Gao, B. A. Salihu, et al. “Network calculus delay bounds in multi-server queueing networks with stochastic arrivals and stochastic services”. *IEEE Global Communications Conference*, San Diego, Dec. 2015.