2017 29th International Teletraffic Congress

State Modulated Traffic Models for Machine Type Communications

Olav N. Østerbø*, Daniel Zucchetto†, Kashif Mahmood*, Andrea Zanella†, Ole Grøndalen*
* Telenor Research, Norway
† Dept. of Information Engineering, University of Padova, Italy

***Abstract*—Machine-to-machine (M2M) traffic is variegate and finding a traffic model which can cover a wide range of M2M sources is challenging. In this paper we address this challenge by proposing an extension of legacy renewal processes for modeling of M2M traffic sources. To this end, we first describe the model and derive some performance parameters, as the overall packet arrival distribution and its moments. We then discuss the packet generation process and consider the counting variable in a time interval, and give the mean and the Laplace transform of the z-transform for this variable. Successively, we present the asymptotic expansion for the variance and the Index of Dispersion of Counts (IDC). We derive the expression of the two first coefficients of this expansion in the general case, while more explicit expressions are provided for some special cases. More specifically, for the special case of a source model with two states, geometric distribution of the numbers of arrivals in each state, and exponential inter-arrival times, we solve for the model parameters in terms of mean, variance and two IDCs. The model is then applied to real M2M traces obtained from an operational network. Albeit the match is not perfect, yet the proposed model captures the main features of the traces, in particular the large burstiness in the packet arrival process.**

***Index Terms*—M2M traffic models, Modulated Renewal Process, Index of dispersion of counts, M2M traffic traces.**

I. Introduction

Machine type communication (MTC) is one of the biggest factor dictating the design of 5G networks. The challenge, however, is that the traffic generated in MTC is very disparate both in volume and shape [1]. For example, there could be water level measuring sensors generating few bytes of data every hour while, on the other side, there could be surveillance cameras hogging the network with massive amount of data. Likewise there could be sensors *uploading* data while on the other hand there could be applications which require *downloading* data such as weather maps for farmers. Secondly, groups of machine type sources, unlike most human generated traffic, may initiate transactions that are correlated both in space and time [2]. So, modeling this diverse canvas of machine type devices is essential to understand the performance of both current and future networks and for accurate network dimensioning.

In this work we present a model which makes it possible to represent a wide range of machine-to-machine (M2M) sources with very different traffic characteristics, ranging from highly regular to bursty types of traffic sources. To this end we propose to apply what we call modulated Renewal Processes for packet level traffic modeling of single M2M type of sources. A modulated Renewal Process is an extension of ordinary Renewal Process (RP) [3] where the distribution between arrivals may change according to the state of a modulating Markov Chain. To make the description manageable we assume the number of arrivals in a state, with a particular inter-arrival time distribution, can be modeled by an integer random variable with a particular distribution that is specific for that particular state. Hence, the total time spent in a particular state (for the modulated RP) is the sum of the inter-arrival times between the arrivals that occur in that particular state. After the total time in a state has expired, the source will move to another state according to the modulating Markov Chain, and start a new RP arrival sequence with its specific distributions (both for the inter-arrival time and the number of arrivals). The modulated RP is therefore a generalization of legacy RPs where the statistical distribution of the arrival process changes depending on a state variable. In addition, we may also specify the packet size distribution, which again may depend on the state variable. This type of parameterized model makes it possible to represent a wide array of M2M sources with varying traffic characteristics.

Modeling based on modulated RP has pros and cons. Sometimes, when the source type and its traffic generation pattern are well known, e.g., alternating between some known deterministic pattern, the modulated RP will not necessary give accurate description. However, when the patterns are more random, the traffic generation will fit well with a general stochastic modeling approach. Some of the appealing properties of the modulated RP traffic model are that they are easy to understand and simulate, and a very broad range of M2M source types can be described with such models, both for regular and bursty traffic. However, there are some drawbacks attached to this modeling approach such as the fact that the modulated RP models involve a large number of parameters, and it is therefore not easy to choose the *best* model and estimate the parameters based on recorded traces. Secondly, it is difficult to model aggregated traffic streams and analytical model based on aggregates are difficult to analyze.

Traffic modeling for MTC has gained a lot of attention in the recent past. Most, if not all, of these approaches are based on state-based modeling [1], [2] as quite often traffic sources have their inherit natural states, depending on different processes in the communication, e.g., waiting states, thinking times and states where sources are active. A well known two-state source model is the ON/OFF model [1], where the source changes between ON state (where packets are sent) and OFF state (where no packets are sent). Traffic modeling for M2M

978-0-9883-0453-6/17 $31.00 © 2017 ITC
DOI 10.1109/ITC.2017.12

last mile wireless access is proposed in [4] where the analysis is limited to event driven and fixed scheduling traffic sources. A coupled Markovian arrival process is proposed in [5] for MTC in an automotive industry and the weakness of traditional simple models with exponential inter-arrival time distribution is highlighted. There is one thing common in most of the related work on traffic modeling for MTC and that is the need for a generic traffic model which can capture the diverse set of MTC use-cases. This paper aims to fill this gap.

II. Packet level model

We shall briefly describe the general packet level model. The idea behind this type of arrival process is to generalize the legacy renewal model, where we allow the distribution between arrivals to change according to the state of a modulating Markov Chain. Further, the sojourn time in a state is determined by the number of arrivals in that particular state, described by an integer random variable, and by the inter-arrival times.[1]

A. Description of the Modulated RP

Let us start by giving a formal description of the modulated RP. The modulated RP is described by the following stochastic variables (see Fig. 1):

- The (modulation) state variable I_k at k'th jump is the state of a Markov Chain, with state space $\Omega = \{1, 2, \ldots, N\}$ and transition probability matrix $Q = (q_{ij})$, with $i, j \in \Omega$, and $q_{ii} = 0$ for all $i \in \Omega$.
- If the modulating Markov Chain has performed k transitions up to time t, then the state of the system at a generic time t is $J_t = I_k$;
- When the modulation state variable I_k is in state $i \in \Omega$, i.e. $I_k = i$, then
  - packet inter-arrival times $\{T_k^i\}$ are independent and identically distributed (iid), with cumulative distribution function (CDF) equal to $G_i(t) = P(T^i \leq t)$, where T^i is the canonical inter-arrival time random variable;
  - the corresponding packet lengths $\{V_k^i\}$ are also iid, with CDF $U_i(x) = P(V^i \leq x)$, where V^i is the canonical packet-size random variable;
  - the total number M^i of packet arrivals in state $i \in \Omega$ is discrete random variable with probability mass distribution $p_i(m) = P(M^i = m)$, for $m = 1, 2, \ldots$.

For the analysis, we also define the Probability Density Functions (PDFs) for the inter-arrival times, $g_i(t) = G_i'(t)$, and for the packet size in each state, $u_i(x) = U_i'(x)$. The associated Laplace Stieltjes Transforms (LSTs) are given by $f_i(s) = \int_{t=0}^{\infty} e^{-st} dG_i(t)$ and $v_i(y) = \int_{x=0}^{\infty} e^{-yx} dU_i(x)$, respectively, while the moment Generating Function (GF) of M^i is given by $P_i(z) = \sum_{m=1}^{\infty} z^m p_i(m)$. Furthermore, we denote $t_i^{(k)} = E[(T^i)^k]$, $v_i^{(k)} = E[(V^i)^k]$ and $m_i^{(k)} = E[(M^i)^k]$ as the k'th moment of inter-arrival times, packet lengths and total

[1]Another possibility is to model the time spent in a state as a separate random variable. This approach, however, would make the model more complicate and it is not considered in this study.

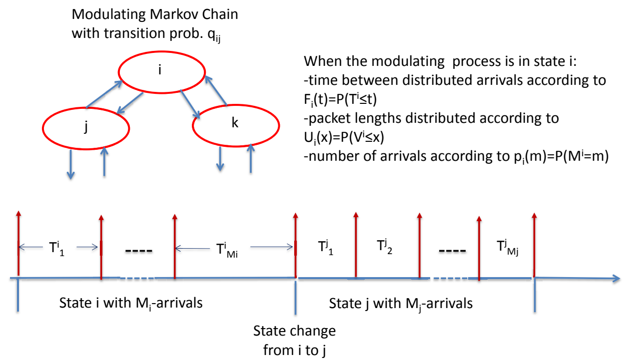


Fig. 1. Traffic generation by the N state Modulated Renewal Process.

number of packet arrivals in state $i \in \Omega$, respectively. For ease of writing, the mean values are represented as $t_i = t_i^{(1)}$, $v_i = v_i^{(1)}$ and $m_i = m_i^{(1)}$, respectively.

Now, the the sojourn time in state i where $i \in \Omega$ is readily found as the (stochastic) sum

$$S^i = T_1^i + T_2^i + \ldots + T_{M^i}^i \tag{1}$$

where all the T_k^i are iid and distributed as T^i.

Given the mutual independence of the involved variables, the transform of the joint distribution of the sojourn time and the number of arrivals in state i is given by $w_i(s, z) = E[e^{-sS^i} z^{M^i}] = E[E[e^{-sS^i} z^{M^i} \mid M^i]] = E[(zf_i(s))^{M^i}]$, which then yields the following functional equation

$$w_i(s, z) = P_i(zf_i(s)) \,. \tag{2}$$

The mean and variance of S^i are hence

$$E[S^i] = m_i t_i \text{ and } \sigma_{S^i}^2 = \sigma_{M^i}^2 t_i^2 + m_i \sigma_{T^i}^2 \tag{3}$$

where σ_X^2 denotes the variance of a random variable X.

To be able to follow the process over several state changes we define the matrix $R(s, z) = (r_{ij}(s, z))$ where we also include the probability of the next state in the expression (2)

$$r_{ij}(s, z) = E[e^{-sS^i} z^{M^i} \mathbf{1}_{\{I_{k+1}=j\}} \mid I_k = i] = P_i(zf_i(s)) q_{ij} \tag{4}$$

or

$$R(s, z) = diag(P_i(zf_i(s)))Q \tag{5}$$

where $\mathbf{1}_{\{I_{k+1}=j\}}$ is 1 if the next state of the modulating Markov Chain is $j \in \Omega$, and zero otherwise, while $diag(\lambda_i)$ is the diagonal matrix with (diagonal) elements λ_i, $i = 1, \ldots, N$. Hence, the process defined by the pair $\{S, I\}$ constitutes a ordinary Markov RP with LST of the generator matrix $R(s, 1) = (r_{ij}(s, 1)) = diag(P(f_i(s))) \cdot Q$.

Let $X^i = V_1^i + V_2^i + \ldots + V_{M^i}^i$ denote the total volume of bits arrived while being in state i. Then, the envelope process $R(s, 1)$ defines the corresponding fluid model where the rate r_i in state i is taken to be the ratio between the mean volume of state i, and the mean soujourn time in that state, i.e.,

$$r_i = \frac{E[X^i]}{E[S^i]} = \frac{v_i}{t_i} \,. \tag{6}$$

If the number of arrivals in state i, M^i, is geometrically distributed, i.e., $p_i(m) = (1-p_i)p_i^{m-1}$ for $m = 1, 2, \ldots$, then the modulated RP will be an ordinary Markov RP with generator Matrix

$$F(t) = diag(G_i(t))\hat{Q} \tag{7}$$

where $\hat{Q} = (\hat{q}_{ij})$ is the transition matrix for the associated Markov Chain given by

$$\hat{q}_{ij} = \begin{cases} p_i & \text{for } j = i \\ (1-p_i)q_{ij} & \text{for } j \neq i \end{cases} \tag{8}$$

Due to the memoryless property of the geometrical distribution, it is sufficient for this particular case to consider the process at packet arrival instants, where p_i represents the probability that there is no state change between two succeeding arrivals, while $1 - p_i$ gives the probability that the state changes, and q_{ij} is the conditional probability that the next state is j, given that state i is left.

If we also assume that the inter-arrival times in state i have negative exponential distribution, i.e., $G_i(t) = 1 - e^{-\lambda_i t}$, then the arrival process is an ordinary Markov Process (MP). In this case, the envelope process defined by the LST of the generator matrix (7) will be a MP, since the time spent in a particular state will have a negative exponential distribution. For Markov RP, several results are available in the literature, and we refer to [6] and [7] for an update of some interesting results.

B. Steady state distributions

In the analysis the steady state distribution of the state variable I_k will play an important role when considering a modulated RP. At jump instances the state variable is governed by the transition matrix Q and if we take $\pi_i = \lim_{k\to\infty} P(I_k = i)$ to be the steady state probabilities for $i = 1, \ldots, N$, and let $\Pi = (\pi_i)$ be the corresponding row vector, the steady state distribution is determined by the following equations

$$\Pi = \Pi \cdot Q \quad \text{ and } \quad \Pi \cdot e = 1 \tag{9}$$

where e is a column vector with ones.

If we consider the state variable at an arbitrary time, say J_t at time t, then the corresponding steady state distribution $p_i = \lim_{t\to\infty} P(J_t = i)$ is found by scaling the probabilities π_i (at state jump instances) by the mean length of a period between jumps, $E[S^i] = m_i t_i$ giving the following steady state distribution

$$p_i = \frac{m_i t_i}{C}\pi_i \quad \text{ with } \quad C = \sum_{i=1}^{N} \pi_i m_i t_i = \Pi \cdot diag(m_i t_i) \cdot e \tag{10}$$

We also define the marginal distributions of time between packets (independent of the states). Over a long period the number of arrivals that see the process in state i with CDF $G_i(t)$ is proportional to $\pi_i m_i$ while the total number of arrivals in the same period is proportional to $\sum_{i=1}^{N} \pi_i m_i$. Hence, the probability that we have an arrival in state i is given by $\eta_i = \frac{\pi_i m_i}{\sum_{i=1}^{N} \pi_i m_i}$. If we take T_{avg} as the average random variable then the CDF $G_{avg}(t) = P(T_{avg} \leq t)$ will be given by

$$G_{avg}(t) = \frac{\sum_{i=1}^{N} \pi_i m_i G_i(t)}{\sum_{i=1}^{N} \pi_i m_i}\,. \tag{11}$$

From (11) we find the moments of T_{avg} to be

$$E[(T_{avg})^k] = \frac{\sum_{i=1}^{N} \pi_i m_i t_i^{(k)}}{\sum_{i=1}^{N} \pi_i m_i}\,. \tag{12}$$

Similar for the packet length, if we take V_{avg} as the *average* packet length distribution, the CDF $U_{avg}(t) = P(V_{avg} \leq t)$ is given by

$$U_{avg}(t) = \frac{\sum_{i=1}^{N} \pi_i m_i U_i(t)}{\sum_{i=1}^{N} \pi_i m_i}\,. \tag{13}$$

As above the moments of U_{avg} are found from (13) as

$$E[(V_{avg})^k] = \frac{\sum_{i=1}^{N} \pi_i m_i v_i^{(k)}}{\sum_{i=1}^{N} \pi_i m_i}\,. \tag{14}$$

In fact it is also possible to derive (11) with somehow different arguments where $G_{avg}(t)$ should be a weighted sum of the $G_i(t)$'s, ($G_{avg}(t) = \sum_{i=1}^{N} a_i G_i(t)$) and we need to determine the a_i's. The main idea is that two succeeding arrivals should see the same average distribution. By considering the process over a long time the probability that an arrival in state i is the last before state change is $\frac{1}{m_i}$ and that it is not is $1 - \frac{1}{m_i}$. If an arrival is the last before state change the next state is j by probability q_{ij} (with arrival distribution $G_j(t)$). For two succeeding distributions to be equal we therefore must have $G_{avg}(t) = \sum_{i=1}^{N} a_i G_i(t) = \sum_{i=1}^{N}(1 - \frac{1}{m_i})a_i G_i(t) + \sum_{i=1}^{N} \frac{1}{m_i} a_i \sum_{j=1}^{N} q_{ij} G_j(t)$. Rearranging this requires $\sum_{j=1}^{N}(-\frac{a_j}{m_j} + \sum_{i=1}^{N} \frac{a_i}{m_i} q_{ij})G_j(t) = 0$ or $-\frac{a_j}{m_j} + \sum_{i=1}^{N} \frac{a_i}{m_i} q_{ij} = 0$ for $j = 1, \cdots, N$. Hence, $a_j = \alpha \pi_j m_j$ for some constant α, and therefore $a_j = \eta_j$ for $j = 1, \cdots, N$.

C. The equilibrium Modulated RP

If we consider a modulated RP which has reached equilibrium and start to observe the process at a certain time, say $t = 0$ and observe the process from there, then the state variable J_t will be in steady state. The time to the next packet arrival, and the number of arrivals to the next state change, denoted by $\tilde{T^i}$ and $\tilde{M^i}$, will be distributed as the residuals of the random variables of T^i and M^i, i.e.,

$$\tilde{G}_i(t) = P(\tilde{T^i} \leq t) = \frac{1}{t_i}\int_0^t (1 - G_i(\tau))d\tau\,; \tag{15}$$

$$\tilde{p_i}(m) = P(\tilde{M^i} = m) = \frac{1}{m_i}\sum_{j=m+1}^{\infty} p_i(j)\,. \tag{16}$$

Furthermore, the LST and GF of the residual distributions may be found from (15) and (16), leading to $\tilde{f}_i(s) = \frac{1-f_i(s)}{st_i}$ and $\tilde{P}_i(z) = \frac{1-P_i(z)}{(1-z)m_i}$, respectively. Observe that the residual distribution of the number of arrivals in a particular state is the probability of having exactly m arrivals before a state changes, when picking a certain arrival interval at random.

For example, the residual sojourn time in state i, denoted as $\tilde{S}^i$, can be expressed as the sum of the residual inter-arrival time for the first packet, which is distributed as $\tilde{T}^i$, and then inter-arrival times of the remaining $\tilde{M}^i$ packets, which are distributed according to T^i, i.e.:

$$\tilde{S}^i = \tilde{T}^i + T_1^i + T_2^i + \ldots + T_{\tilde{M}^i}^i \tag{17}$$

where the all the T_k^i, all are independent and distributed according to T^i.

The joint transform of the time duration $\tilde{S}_i$ and the number of arrivals $1 + \tilde{M}^i$ until the next state change is given as

$$w_i(s, z) = z\tilde{f}_i(s)\tilde{P}_i[(zf_i(s))]$$

which by applying the relations for $\tilde{f}_i(s)$ and $\tilde{P}_i(z)$ gives

$$\tilde{w}_i(s, z) = \frac{z}{t_i m_i}\hat{w}_i(s, z) \tag{18}$$

where

$$\hat{w}_i(s, z) = \frac{(1 - f_i(s))(1 - P_i(zf_i(s)))}{s(1 - zf_i(s))} \tag{19}$$

The LST of the residual sojourn time in a given state i is now easily found from (18) by taking $z = 1$, giving $\tilde{w}_i(s, 1) = \frac{1-P_i(f_i(s))}{sm_i t_i}$ as expected. To be able to follow the process over the first state changes we also define the matrix $\tilde{R}(s, z) = (\tilde{r}_{ij}(s, z))$ where we include the probability of the next state in the expression (18)

$$\tilde{r}_{ij}(s, z) = E[e^{-s\tilde{S}}z^{1+\tilde{M}}\mathbf{1}_{\{I_1=j\}} \mid J_0 = i] = \tilde{w}_i(s, z)q_{ij} \tag{20}$$

or

$$\tilde{R}(s, z) = diag(\tilde{w}_i(s, z))Q = zdiag(\frac{1}{t_i m_i})\hat{R}(s, z) \tag{21}$$

where the matrix $\hat{R}(s, z)$ can also be written as

$$\hat{R}(s, z) = diag(\hat{w}_i(s, z))Q \tag{22}$$

and where $\tilde{S} = \tilde{S}^i$ (given by (17)) and $\tilde{M} = \tilde{M}^i$ are the duration and number of arrivals in this particular state where the state variable at time $t = 0$ is i; that is $J_0 = i$.

The data volume arrived during the initial state may be written as $\tilde{X}^i = V_0^i + V_1^i + V_2^i + \ldots + V_{\tilde{M}^i}^i$ where all the V^i's are distributed according to $U_i(x) = P(V^i \leq x)$ as defined in subsection II-A.

We may now combine some of the results discussed above and consider a modulated RP in equilibrium over several state changes. By combining the results above for the initial state II-C and for the normal state II-A we obtain the following theorem.

Theorem 1. *Consider a modulated RP in equilibrium and observe the process from a random point taken to be $t = 0$ and let Y_k and L_k be the time and numbers of arrivals up to the k'th state change. Defining the matrix*

$$\begin{aligned} &R^k(s, z) = (r_{ij}^k(s, z)) \quad \textit{where} \\ &r_{ij}^k(s, z) = E[e^{-sY^k}z^{L^k}\mathbf{1}_{\{I_{k+1}=j\}} \mid J_0 = i] \end{aligned} \tag{23}$$

we then have

$$R^k(s, z) = \tilde{R}(s, z) \cdot R(s, z)^{k-1} \tag{24}$$

where $\tilde{R}(s, z)$ is given by (21) and $R(s, z)$ is given by (4).

Proof. We have $Y^k = \tilde{S}_1 + S_2 \ldots + S_k$ and $L^k = 1 + \tilde{M}_1 + M_2 \ldots + M_k$ where $\tilde{S}_1$ and $1 + \tilde{M}_1$ are the time and number of arrivals to the first state change for the initial period and S_l and M_l are the time and number of arrivals for period l; $l = 2, \cdots, k$. By inserting in (23) we then obtain (24). □

D. Packet counts and index of dispersion

To study the behavior of modulated RP over longer time periods, we want to find the distribution of the number of arrivals up to a certain point in time. To do this we first find the distribution of the number of arrivals in the last period which includes that point in time. We therefore first consider the process in a given time interval. Suppose that the state is $I_k = i$ and define the time up to the l'th arrival within this state as

$$S_l^i = T_1^i + T_2^i + \ldots + T_l^i\,. \tag{25}$$

Similarly, let $\hat{N}_t^i$ denote the number of arrivals that occur up to time t, assuming that at time $t = 0$ the state is $I_k = i$ and there is no state change in the interval $(0, t)$. Therefore, the event $\{\hat{N}_t^i = l\}$ equals that of $\{S_l^i > t, S_{l+1}^i \leq t, M^i > l\}$. We hence have

$$\begin{aligned} &P(\hat{N}_t^i = l) = P(S_l^i > t, S_{l+1}^i \leq t, M^i > l) = \\ &(P(S_l^i \leq t) - P(S_{l+1}^i \leq t))P(M^i > l)\,. \end{aligned} \tag{26}$$

By defining the z-transform $\hat{H}^i(t, z) = E[z^{\hat{N}_t^i}]$ and by taking the Laplace transform $\hat{G}^i(s, z) = \int_{t=0}^{\infty} e^{-st}\hat{H}^i(t, z)dt$ and using the fact that the Laplace transform of the convolution $P(S_l^i \leq t)$ equals $\frac{f_i(s)^l}{s}$ then this give $\hat{G}^i(s, z) = \frac{1-f_i(s)}{s}\sum_{l=0}^{\infty}(zf_i(s))^l P(M^i > l) = \frac{(1-f_i(s))(1-P_i(zf_i(s)))}{s(1-zf_i(s))}$. Hence by (19) we have

$$\hat{G}^i(s, z) = \hat{w}_i(s, z) \tag{27}$$

Similarly, the initial period, $k = 0$ have to be treated somewhat differently due to the fact that the time to the first arrival is given by the residual time and the residual numbers of arrivals. By assuming $J_0 = i$ we define the time up to the l'th arrival in that period

$$\tilde{S}_l^i = \tilde{T}^i + T_1^i + \ldots + T_l^i \tag{28}$$

where $\tilde{T}^i$ is distributed according to residual arrival time. As above we let $\tilde{N}_t^i$ be the number of arrivals in the initial period ($k = 0$) up to a time t without any state changes in the interval $(0, t)$. The event $\{\tilde{N}_t^i = l\}$ equals that of $\{\tilde{S}_{l-1}^i > t, \tilde{S}_l^i \leq t, \tilde{M}^i > l - 1\}$ leading to

$$\begin{aligned} &P(\tilde{N}_t^i = l) = P(\tilde{S}_{l-1}^i > t, \tilde{S}_l^i \leq t, \tilde{M}^i > l - 1) = \\ &(P(\tilde{S}_{l-1}^i \leq t) - P(\tilde{S}_l^i \leq t))P(\tilde{M}^i > l - 1)\,. \end{aligned} \tag{29}$$

As above, we define the z-transform $\tilde{H}^i(t, z) = E[z^{\tilde{N}_t^i}]$ and the Laplace transform $\tilde{G}^i(s, z) = \int_{t=0}^{\infty} e^{-st}\tilde{H}^i(t, z)dt$. By using the fact that the Laplace transform of the convolution $\tilde{P}(S_l^i \leq t)$ equals $\tilde{f}_i(s)\frac{f_i(s)^{l-1}}{s}$ we find $\tilde{G}^i(s, z) = z\tilde{f}_i(s)\frac{1-f_i(s)}{s}\sum_{l=0}^{\infty}(zf_i(s))^l P(\tilde{M}^i > l)$. Since $\sum_{l=0}^{\infty} z^l P(\tilde{M}^i > l) = \frac{1}{1-z} - \frac{1-P_i(z)}{m_i(1-z)^2}$ we obtain

$$\tilde{G}^i(s,z) = z\frac{(1-f_i(s))^2}{t_i s^2}\left(\frac{1}{1-zf_i(s)} - \frac{1-P_i(zf_i(s))}{m_i(1-zf_i(s))^2}\right) \quad (30)$$

We may now state the following theorem on the distribution of the number of arrivals up to a certain time t.

Theorem 2. *Consider a modulated RP in equilibrium and let N_t be the number of arrivals up to time t and let $P(N_t = n)$ be the corresponding distribution and let $H(t,z) = E[z^{N_t}]$ be the z-transform. Then the Laplace transform $G(s,z) = \int_{t=0}^{\infty} e^{-st}H(t,z)dt$ is given by the following matrix expressions*

$$\begin{aligned} G(s,z) = &\tfrac{1}{s} + \tfrac{z-1}{Cs^2}\Pi \cdot diag(m_i \tfrac{1-f_i(s)}{1-zf_i(s)}) \cdot e - \\ &\tfrac{z}{Cs}\Pi \cdot diag(\tfrac{1-f_i(s)}{1-zf_i(s)}\hat{w}_i(s,z)) \cdot e + \\ &\tfrac{z}{C}\Pi \cdot \hat{R}(s,z) \cdot [I - R(s,z)]^{-1} \cdot \hat{R}(s,z)) \cdot e \end{aligned} \quad (31)$$

where Π is the steady state distribution at state jumps given by (9), the constant C is given in (10), $\hat{w}_i(s,z)$ is given by (19) and the matrices $R(s,z)$ and $\hat{R}(s,z)$ are given by (4) and (22) respectively.

Proof. We first observe to have $N_t = 0$ we must have the residual time $\tilde{T}^i > t$ and hence,

$$P(N_t = 0) = \sum_{i=1}^{N} p_i P(\tilde{T}^i > t) \quad (32)$$

The condition $N_t = n$ when $n > 0$ may be attained by either having no state changes up to t or on ore more state changes, e.g., $k \geq 1$. For the latter case, we condition on the elapsed time $Y^k = y$ and arrivals $L^k = l$ up to the k'th state change. Then the numbers of arrivals in the remaining interval of length $t - y$ has to be $n - l$ (to have $N_t = n$). By integrating and summing over all possible combinations of arrivals in the two intervals and summing over all $k \geq 1$, and then multiplying by the initial state probabilities and summing over all states both at time $t = 0$ and t, one gets the following expression

$$\begin{aligned} &P(N_t = n) = \textstyle\sum_{i=1}^{N} p_i P(\tilde{N}_t^i = n) + \\ &\textstyle\sum_{k=1}^{\infty}\sum_{i=1}^{N}\sum_{i_0=1}^{N} p_{i_0}\sum_{l=0}^{n}\int_{y=0}^{t} P(\hat{N}_{t-y}^i = n-l) \\ &d_y P(Y^k \leq y, L^k = l, I^k = i \mid J_0 = i_0) \, . \end{aligned} \quad (33)$$

Now we take the transforms of the expressions (32) and (33). Laplace transform of $P(N_t = 0)$ yields $P \cdot diag(\frac{1}{s} - \frac{1-f_i(s)}{t_i s^2}) \cdot e$. Similarly, the transforms of the second term $\sum_{i=1}^{N} p_i P(\tilde{N}_t^i = n)$ is $P \cdot diag(\tilde{G}^i(s,z)) \cdot e$. Finally the third convolution part yields the sum $\sum_{k=1}^{\infty} P \cdot R^k(s,z) \cdot diag(\hat{w}_i(z,s) \cdot e$. Using $R^k(s,z)$ given by (24) yields $\frac{z}{C}\Pi \cdot \hat{R}(s,z) \cdot [I - R(s,z)]^{-1} \cdot \hat{R}(s,z)) \cdot e$ where we also have used that $diag(\hat{w}_i(z,s) \cdot e = \hat{R}(s,z)) \cdot e$. Collecting and substituting for P in terms of the steady state jump probabilities Π and inserting for $\tilde{G}^i(s,z)$ by (30) we then obtain (31). □

To find the Laplace transform of the first and second moments of N_t turn out to beneficial to rewrite the expression for $G(s,z)$ above by taking $\hat{w}_i(s,z) = \frac{1}{sz}(1 + \frac{z-1}{1-zf_i(s)})(1 - P_i(zf_i(s))$. This leads to the following simplification

$$\begin{aligned} &G(s,z) = \tfrac{1}{s} + \tfrac{z-1}{Czs^2}\Pi \cdot diag(m_i) \cdot e + \\ &\tfrac{(z-1)^2}{Czs^2}\{\Pi \cdot diag(\tfrac{m_i}{1-zf_i(s)} - \tfrac{1-P_i(zf_i(s))}{1-zf_i(s)}) \cdot e + \\ &\Pi \cdot diag(\tfrac{1-P_i(zf_i(s))}{1-zf_i(s)}) \cdot Q \cdot [I - diag(P_i(zf_i(s))) \cdot Q]^{-1} \cdot \\ &diag(\tfrac{1-P_i(zf_i(s))}{1-zf_i(s)}) \cdot e\} \, . \end{aligned} \quad (34)$$

By (34) we find that the first moment is proportional to the length of the interval, while for the variance the general result is found in terms of Laplace transforms. We state the result in the following theorem.

Theorem 3. *For the mean and variance of N_t we have the following expressions*

$$E[N_t] = t\frac{\Pi \cdot diag(m_i) \cdot e}{C} \quad (35)$$

$$\begin{aligned} &\textstyle\int_{t=0}^{\infty} e^{-st}Var[N_t]dt = -\tfrac{\Pi \cdot diag(m_i) \cdot e}{Cs^2} - \tfrac{2(\Pi \cdot diag(m_i) \cdot e)^2}{C^2 s^3} + \\ &\tfrac{2}{Cs^2}\{\Pi \cdot diag(\tfrac{m_i}{1-f_i(s)} - \tfrac{1-P_i(f_i(s))}{1-f_i(s)}) \cdot e + \\ &\Pi \cdot diag(\tfrac{1-P_i(f_i(s))}{1-f_i(s)}) \cdot Q \cdot [I - diag(P_i(f_i(s))) \cdot Q]^{-1} \cdot \\ &diag(\tfrac{1-P_i(f_i(s))}{1-f_i(s)}) \cdot e\} \, . \end{aligned} \quad (36)$$

Proof. These results follow directly from (34) by differentiating with respect to z to first and second order, and then finding the Laplace transform of $E[N_t]^2$. □

For large t both the mean and variance will grow with rate proportional to t, it is therefore natural to introduce the *index of dispersion of counts* (IDC) as the ratio between variance and mean

$$I_t = \frac{Var[N_t]}{E[N_t]} \, . \quad (37)$$

By Tauberian arguments, it is possible to obtain the asymptotic expansion of $Var[N_t]$ for large t. The method used is to expand the inverse matrix $[I - diag(P_i(f_i(s)))]^{-1}$ in terms of its adjoint matrix and the determinant, and then also expand both $\frac{m_i}{1-f_i(s)} - \frac{1-P_i(f_i(s))}{1-f_i(s)}$ and $\frac{1-P_i(f_i(s))}{1-f_i(s)}$ for small s. We find the following asymptotic expansions for large t. The technical details is given in a separate annex.

Theorem 4. *The variance and IDC have the following asymptotic expressions for large t*

$$Var[N_t] = At + B + O(t^{-1}) \quad (38)$$

$$I_t = \frac{A}{D} + \frac{B}{D}t^{-1} + O(t^{-2}) \quad (39)$$

where A and B are constants given in terms of model parameters and are derived in Appendix A and $D = E[N_1] = \frac{\Pi \cdot diag(m_i) \cdot e}{C}$.

III. Fitting two state models to recorded traces

One of the main difficulties when applying general source models like modulated RP is to set the model's parameters based on real measurements (traces). In our case, for each state it is necessary to find three distributions, namely for the inter-arrival times, the packet size, and the number of arrivals, in addition to the transition matrix of the modulating Markov Chain. Instead, some important parameters, like the overall statistical moments and autocorrelation of key variables or the dispersion index, are quite easy to estimate by exploiting the (supposed) ergodic nature of the involved stochastic processes. Hence, by estimating a set of parameters and requiring that these measurements match the corresponding analytical expressions (derived from the model), we obtain a set of equations that may be solved for the model parameters.

We observe that the number of statistical distributions that need to be estimated for an N-state modulated RP based on the recorded traces grows linearly with N, while the size of the transition matrix is equal to N^2.

Clearly, the larger the number of model's parameters to be estimated, the larger the required data set, and the more noisy the resulting model. It is therefore convenient to simplify the model by fixing some of the model's parameters and estimating the remaining from the available data traces. The difficulty is to choose the right balance between model complexity, and its capability to capture the more important features of the actual source process.

To set the model parameters, we consider two classes of measurements, namely:

- short-time scale: values that are meaningful over short time scales, like mean and variance of the inter-arrival times, packet size, and number of arrivals processes,
- long-time scale: values describing the packet generation process over longer time scales, like packet counts or index of dispersion.

In the following we detail the proposed simplified source model.

A. Two-state model with negative exponential arrival distribution and geometrical distribution of arrivals

The simplest (non-trivial) model consists of two modulating states, $\Omega = \{1, 2\}$, with single-parameter distributions for the inter-arrival times, the packet size and the number of arrivals in a state. More specifically, we choose the inter-arrival times to be negative exponentially distributed random variables with mean t_1 and t_2 for state $I_k = 1$ and $I_k = 2$, respectively, while the number of arrivals in each state is modeled as a geometrically distributed random variable with mean m_1 and m_2, respectively while the packet size is exponentially distributed. Hence, the model is fully described by the parameter set $\{t_1, t_2, m_1, m_2\}$. For this case the Laplace transform (36) of the variance of N_t is invertible and we find the following IDC

$$\begin{aligned} I_t = 1 + 2\frac{m_1^2 m_2^2 (t_1 - t_2)^2}{(m_1 t_1 + m_2 t_2)^2 (m_1 + m_2)} - \\ 2\frac{m_1^3 t_1 m_2^3 t_2 (t_1 - t_2)^2}{t(m_1 t_1 + m_2 t_2)^3 (m_1 + m_2)} (1 - e^{-t\frac{m_1 t_1 + m_2 t_2}{m_1 t_1 m_2 t_2}}) \end{aligned} \quad (40)$$

when $t \to \infty$, (40) gives

$$I_\infty = 1 + 2\frac{m_1^2 m_2^2 (t_1 - t_2)^2}{(m_1 t_1 + m_2 t_2)^2 (m_1 + m_2)} . \quad (41)$$

Moreover if we know the IDC for a particular time t_0 and if we take $F^{-1}(y)$ as the inverse function of $F(x) = \frac{1-e^{-x}}{x}$ we may solve for the exponent in (40) leading to

$$\frac{m_1 t_1 + m_2 t_2}{m_1 t_1 m_2 t_2} = \frac{1}{t_0} F^{-1}\left(\frac{I_\infty - I_{t_0}}{I_\infty - 1}\right) . \quad (42)$$

Suppose that from trace measurements we have estimated the four parameters i.e. the mean, square coefficient of variation, and IDC at infinity and at a particular time t_0. We may then set up the following four equations to determine the model parameters $\{t_1, t_2, m_1, m_2\}$

$$\begin{aligned} &\tfrac{m_1 t_1 + m_2 t_2}{m_1 + m_2} = a = E[T_{avg}] \\ &2\tfrac{m_1 m_2 (t_2 - t_1)^2}{(m_1 t_1 + m_2 t_2)^2} = b = \tfrac{Var[T_{avg}]}{E[T_{avg}]^2} - 1 \\ &2\tfrac{m_1^2 m_2^2 (t_2 - t_1)^2}{(m_1 t_1 + m_2 t_2)^2 (m_1 + m_2)} = c = I_\infty - 1 \\ &\tfrac{m_1 t_1 + m_2 t_2}{m_1 t_1 m_2 t_2} = d = \tfrac{1}{t_0} F^{-1}(\tfrac{I_\infty - I_{t_0}}{I_\infty - 1}) \end{aligned} \quad (43)$$

where we assume that the parameters on the right hand side $\{E[T_{avg}], Var[T_{avg}], I_\infty, I_{t_0}\}$ all are known by direct measurements. Fortunately, (43) yields quadratic equations by substituting $x_1 = m_1 t_1$ and $x_2 = m_2 t_2$ for which we find the following solutions

$$\begin{aligned} x_1 &= \tfrac{1}{4b^2 d\eta} (\Delta + (2\eta + b(\eta - 2))\sqrt{\Delta}) \\ x_2 &= \tfrac{1}{4b^2 d\eta} (\Delta - (2\eta + b(\eta - 2))\sqrt{\Delta}) \\ m_1 &= \tfrac{1}{4ab^2 d\eta} (\Delta + (2\eta - b(\eta + 2))\sqrt{\Delta}) \\ m_2 &= \tfrac{1}{4ab^2 d\eta} (\Delta - (2\eta - b(\eta + 2))\sqrt{\Delta}) \end{aligned} \quad (44)$$

where

$$t_1 = \frac{x_1}{m_1} \text{ and } t_2 = \frac{x_2}{m_2} . \quad (45)$$

and

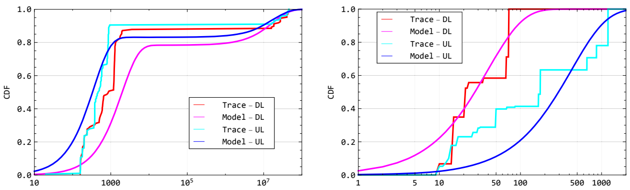
$$\Delta = 4\eta^2 + 4\eta(\eta - 2)b + (\eta - 2)^2 b^2 \ \ and \ \ \eta = acd . \quad (46)$$

B. Modeling examples

To test the proposed method we have recorded some traces from M2M devices. Below we have applied the method for three types of M2M sources namely

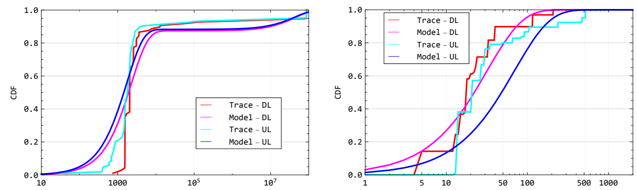
- Single electricity meter.
- Concentrator which aggregates measurements from different electricity meters.
- Parking meter.

The source types we consider are taken from a real M2M network based on recorded traces for both uplink (UL) and downlink (DL). All the traces are recorded within on week time of raw data. Based on the traces, different time-series are constructed and analysed. In the examples below we mainly concentrate on the packet arrival process. The measured parameters are: mean and coefficient of variation of the arrival times, and index of dispersion at two points in time, one very large and the second at 10 s. The estimated parameters are given in Tab. I. We observe that the average time between packets is relatively long, e.g., in the range of one hour for the electricity meter, while for the parking meter the mean is around a couple of minutes. We have large values, i.e., in the 10-25 range, for both the square coefficient of variation and the dispersion index for large time. The index of dispersion at



(a) Inter Arrival Time [ms] (b) Packet Length [Bytes]

Fig. 2. CDF of the inter arrival time and packet length obtained from the model and based on direct estimation from traces for single Electric Meter type of source.



(a) Inter Arrival Time [ms] (b) Packet Length [Bytes]

Fig. 3. CDF of the inter arrival time and packet length obtained from the model and based on direct estimation from traces for Electric Meter from concentrator type of source.

10 s is less than 10 for all cases. From these parameters and by manual inspection of the recorded time-series we conclude that these traffic sources are very bursty in nature.

TABLE I
MEASURED PARAMETERS FOR THE SOURCES

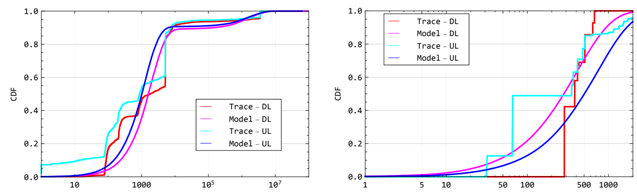
	$E[T_{avg}]$ [s]	$\frac{Var[T_{avg}]}{E[T_{avg}]^2}$	I_∞	$I_{10\text{sec}}$	$E[V_{avg}]$ [Bytes]	$\frac{Var[V_{avg}]}{E[V_{avg}]^2}$
ElMeter P2P UL	2969.73	10.8499	12.4174	9.63390	425.168	1.27135
ElMeter P2P DL	3812.49	8.2299	10.0329	7.12250	39.5970	0.44533
ElMeter Cons. UL	5025.31	16.0387	15.2734	5.07427	69.5619	3.53885
ElMeter Cons. DL	5387.97	14.8912	14.7908	4.22270	32.0204	1.64640
Parking UL	162.055	20.0424	25.3383	6.89966	728.636	10.42150
Parking DL	188.133	17.1263	22.8702	4.94596	406.867	0.10329

TABLE II
CALCULATED MODEL PARAMETERS FOR THE SOURCES.

	t_1 [s]	m_1	t_2 [s]	m_2
ElMeter P2P UL	0.36137	6.86923	17597.4	1.39443
ElMeter P2P DL	0.58994	5.76723	17596.7	1.59488
ElMeter Cons. UL	1.73002	8.09070	42825.4	1.07525
ElMeter Cons. DL	2.27016	7.89402	42826.4	1.13559
Parking UL	1.26545	13.6396	1717.15	1.41026
Parking DL	1.95609	12.5222	1721.01	1.52089

The estimated two-state model parameters are shown in Tab. II. We observe that the resulting calculated parameters give a typical ON/OFF pattern with two very distinct states. In the first state, the inter arrival time is in the range of one second and the mean number of packets is estimated around 10 packets, while in the second mode inter-packet time is large and the mean number of packet arrivals in this state is less than two for all the sources. Hence, for all the cases we have a typical ON/OFF behaviour, with a burst of approximately ten packets and then very long time between bursts.

The corresponding distributions of the inter-arrival times and packet lengths for the different cases are shown in Fig. 2, Fig. 3 and Fig. 4. Compared to the estimated distributions, the CDF based on the model is smooth. However, for all the three cases, even if the match is not perfect, the similarity is evident. We see that the two modes manifest themselves in the form of the CDF with most of the arrivals occurring in the first mode with the relative small inter arrival times, while in mode two the time between packets will be several hours for the



(a) Inter Arrival Time [ms] (b) Packet Length [Bytes]

Fig. 4. CDF of the inter arrival time and packet length obtained from the model and based on direct estimation from traces for single Parking Meter type of source.

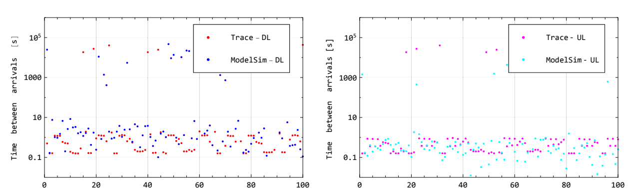
Electrical Metering sources. Instead, for the Parking source, the corresponding time between arrivals is typically half an hour. For the packet length distributions we have also added the curves based on negative exponential approximations with the same mean as the empirical estimate ones.

Fig. 5, Fig. 6 and Fig. 7 show time-series from the traces and from simulations, using the two-state model with the calculated parameters given in Tab. II. Also these figures confirm that the two-state model fits quite well with the traces, which clearly show that most of the inter-arrival times are quite small. Then there are a few observations with large time between arrivals and we clearly see that the model recreates similar behaviour.

A natural follow up of the plain descriptive comparison between estimated and recorded CDFs and time series, would be to perform a more in-depth analysis of the differences based on appropriate metrics. This is, however, outside the scope of this paper, but could be part of a possible extension.

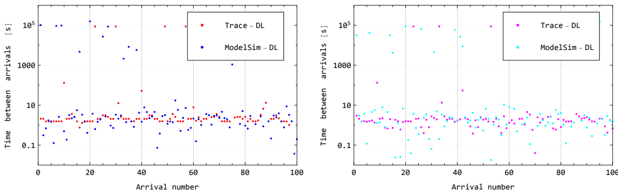
IV. CONCLUSION

In this paper, we have proposed a generic packet-level model, which is an alternating renewal process with different



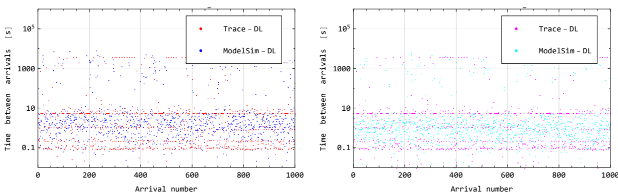
(a) Downlink (b) Uplink

Fig. 5. Simulated and trace time series of the inter arrival times for single Electric Meter type of source.



(a) Downlink (b) Uplink

Fig. 6. Simulated and trace time series of the inter arrival times for Electric Meter from concentrator type of source.



(a) Downlink (b) Uplink

Fig. 7. Simulated and trace time series of the inter arrival times for single Parking Meter type of source.

numbers of arrivals in each state. After that all the arrivals for a particular state have arrived, the process changes its state, which corresponds to different distributions of the inter-arrival times and of the number of arrivals in the state. We derived the most important performance measures for such processes, as the marginal arrival distributions, and the corresponding statistical moments, and we analysed the distribution of the packet count in a given time interval. The variance of the packet count is found in terms of Laplace transform, and asymptotic behaviour is found for large times. For two- and three-state models, explicit results are derived.

Finally, we also proposed a two-state model with exponentially distributed packet inter-arrival times and geometrically distributed number of arrivals in a state, for a total of four model parameters. We suggested a method to determine these parameters from quantities estimated from traffic traces. The method was tested against three types of M2M traces. The analysed traces showed that some of the M2M traffic sources are very bursty in nature, like ON/OFF traffic with approximately ten packets before a very long OFF period without any packet generation. The proposed model provided reasonably similar stochastic characteristics, though without achieving a perfect match.

The main question that we leave for future study is whether such a simple two-state model is sufficient to capture the characteristics of the actual M2M traffic sources that may impact on the design and performance of wider systems, thus greatly simplifying the study of techniques to include M2M traffic into next-generation networks.

APPENDIX A
ASYMPTOTIC EXPANSION OF $Var[N_t]$ FOR LARGE t

We first expand the different parts of the Laplace transform of $Var[N_t]$ in (36) for small s. We denote σ^2 and γ^3 as the variance and 3'rd central moment respectively. We find

$$P_i(f_i(s)) = 1 - m_i t_i s + \frac{1}{2}u_i^{(2)}s^2 - \frac{1}{6}u_i^{(3)}s^3 + o(s^3) \quad (47)$$

where

$$\begin{aligned} u_i^{(2)} &= t_i^2(\sigma_{M_i}^2 + m_i^2) + \sigma_{T_i}^2 m_i \\ u_i^{(3)} &= t_i^3(\gamma_{M_i}^3 + m_i^3) + t_i\sigma_{T_i}^2(3\sigma_{M_i}^2 + 3m_i^2 - 3m_i) + \gamma_{T_i}^3 m_i \end{aligned} \quad (48)$$

and

$$\frac{1 - P_i(f_i(s))}{1 - f_i(s)} = m_i - \frac{1}{2}w_i^{(1)}s + \frac{1}{12}w_i^{(2)}s^2 + o(s^2) \quad (49)$$

where

$$\begin{aligned} w_i^{(1)} &= t_i(\sigma_{M_i}^2 + m_i^2 - m_i) \\ w_i^{(2)} &= t_i^2(2\gamma_{M_i}^3 - 3\sigma_{M_i}^2 + 2m_i^3 - 3m_i^2 + m_i) + \\ &\quad \sigma_{T_i}^2(3\sigma_{M_i}^2 + 3m_i^2 - 3m_i) \end{aligned} \quad (50)$$

and

$$\frac{m_i}{1 - f_i(s)} - \frac{1 - P_i(f_i(s))}{(1 - f_i(s))^2} = \frac{1}{2}\nu_i^{(1)} - \frac{1}{6}\nu_i^{(2)}s + o(s) \quad (51)$$

where

$$\begin{aligned} \nu_i^{(1)} &= \sigma_{M_i}^2 + m_i(m_i - 1) \\ \nu_i^{(2)} &= t_i(\gamma_{M_i}^3 - 3\sigma_{M_i}^2 + m_i(m_i - 1)(m_i - 2)) \end{aligned} \quad (52)$$

From the expression of the Laplace transform of the variance given by (36) we see that the hard part to is to expand $[I - diag(P_i(f_i(s)))Q]^{-1}$ for small s. We use the notion of adjoint matrices and use the result $adjA \cdot A = A \cdot adjA = I \cdot detA$ for non-singular NxN matrix, or $A^{-1} = \frac{adjA}{detA}$. Expanding we have $[I - diag(P_i(f_i(s)))Q] = I - Q + s\,diag(m_i t_i) \cdot Q - \frac{1}{2}s^2 diag(u_i^2) \cdot Q + o(s^2)$. Similar we expand both the adjoint and the determinant $adj[I - diag(P_i(f_i(s)))Q] = H_0 + sH_1 + s^2H_2 + o(s^2)$ and $det[I - diag(P_i(f_i(s)))Q] = b_0 + sb_1 + s^2b_2 + s^3b_3 + o(s^3)$. It follows that $b_0 = 0$ since $det[I - Q] = 0$. By expanding the of identity for adjoint matrices we obtain the following equations to determine H_0

$$\begin{aligned} &H_0[I - Q] = [I - Q]H_0 = 0 \\ &H_1[I - Q] + H_0 diag(m_i t_i)Q = \\ &[I - Q]H_1 + diag(m_i t_i)QH_0 = b_1 \end{aligned} \quad (53)$$

The first equation gives $H_0 = a_0 L$ where a_0 is a constant and $L = e \cdot \Pi$. By pre- or post-multiplying the second equation by the matrix L gives $a_0 L \cdot diag(m_i t_i) \cdot Q \cdot L = b_1 L$ giving $a_0 C = b_1$ with $C = \Pi \cdot diag(m_i t_i) \cdot e$, and hence $H_0 = \frac{b_i}{C}L$. Expanding to second order of the inverse by using the expression $\frac{adj[I - diag(P_i(f_i(s)))Q]}{det[I - diag(P_i(f_i(s)))Q]}$, we finally find the following expansion for the inverse

$$\begin{aligned} &Q[I - diag(P_i(f_i(s)))Q]^{-1} = \frac{1}{s}\frac{L}{C} + \frac{1}{b_1}B_1 - \frac{b_2}{b_1}\frac{L}{C} + \\ &s(\frac{1}{b_1}B_2 - \frac{b_2}{b_1^2}B_1 + [\frac{b_2^2}{b_1^2} - \frac{b_3}{b_1}]\frac{L}{C}) + o(s) \end{aligned} \quad (54)$$

where we have defined $B_i = QH_i$ and further H_i and b_i is the i'th coefficients in the expansion of $adj[I - diag(P_i(f_i(s)))Q]$ and $det[I - diag(P_i(f_i(s)))Q]$ respectively. The sought constants A and B are now the coefficients of s^{-2} and s^{-1} in the expansion of the Laplace transform (36) above. We first observe that the coefficient of s^{-3} vanishing as expected. We find

$$\begin{aligned} A = & \frac{\Pi \cdot diag(\nu_i^{(1)}) \cdot e}{C} - \frac{\Pi \cdot diag(m_i) \cdot e}{C} \\ & + 2\frac{\Pi \cdot diag(m_i) \cdot B_1 \cdot diag(m_i) \cdot e}{b_1 C} - 2\frac{b_2}{b_1}\left(\frac{\Pi \cdot diag(m_i) \cdot e}{C}\right)^2 \\ & - 2\frac{(\Pi \cdot diag(m_i) \cdot e)(\Pi \cdot diag(w_i^{(1)}) \cdot e)}{C^2} \end{aligned} \tag{55}$$

and

$$\begin{aligned} B = & 2\frac{\Pi \cdot diag(m_i) \cdot B_2 \cdot diag(m_i) \cdot e}{b_1 C} - 2\frac{b_2}{b_1}\frac{\Pi \cdot diag(m_i) \cdot B_1 \cdot diag(m_i) \cdot e}{b_1 C} \\ & + \frac{\Pi \cdot diag(m_i) \cdot B_1 \cdot diag(w_i^{(1)}) \cdot e}{b_1 C} + \frac{\Pi \cdot diag(w_i^{(1)})) \cdot B_1 \cdot diag(m_i) \cdot e}{b_1 C} \\ & - \frac{1}{3}\frac{\Pi \cdot diag(\nu_i^{(2)}) \cdot e}{C} + 2[(\frac{b_2}{b1})^2 - \frac{b_3}{b_1}](\frac{\Pi \cdot diag(m_i) \cdot e}{C})^2 \\ & + 2\frac{b_2}{b_1}\frac{(\Pi \cdot diag(m_i) \cdot e)(\Pi \cdot diag(w_i^{(1)})) \cdot e}{C^2} \\ & + \frac{1}{3}\frac{(\Pi \cdot diag(m_i) \cdot e)(\Pi \cdot diag(w_i^{(2)})) \cdot e}{C^2} + \frac{1}{2}(\frac{\Pi \cdot diag(w_i^{(1)}) \cdot e}{C})^2 \end{aligned} \tag{56}$$

For the general case, the expansion of the determinant and the adjoint of $[I - diag(P_i(f_i(s)))Q]$ will be the hard part to find.

Appendix B
Explicit expressions for a two and three state models

The results given in (55) and (56) above require the matrices B_1 and B_2 as the first and second order expansion of the ajoint matrix $[I - diag(P_i(f_i(s)))Q]$ as well as b_1,b_2 and b_3; the three first coefficient for the determinant. For general number of states analytical expression is hard to find unless for small value of number of states. Below we give explicit expressions for the two and three state cases. For N=2 we find the following expression for the constant A

$$\begin{aligned} A = & \frac{1}{(m_1 t_1 + m_2 t_2)^3}((m_1 + m_2)^2(m_1\sigma_{T_1}^2 + m_2\sigma_{T_2}^2) + \\ & (t_1 - t_2)^2(m_2^2\sigma_{M_1}^2 + m_1^2\sigma_{M_2}^2)) \end{aligned} \tag{57}$$

For N=3 the expressions are far more technical with several more parameters. We take the Q-matrix as follows

$$Q = \begin{pmatrix} 0 & q_{12} & 1 - q_{12} \\ 1 - q_{23} & 0 & q_{23} \\ q_{31} & 1 - q_{31} & 0 \end{pmatrix} \tag{58}$$

and define the following auxiliary parameters

$$\begin{aligned} r_1 &= 1 - q_{23}(1 - q_{31}) \\ r_2 &= 1 - q_{31}(1 - q_{12}) \\ r_3 &= 1 - q_{12}(1 - q_{23}) \end{aligned} \tag{59}$$

and the steady state probabilities at state jumps is then

$$\pi_1 = \frac{r_1}{r_1 + r_2 + r_3}, \pi_2 = \frac{r_2}{r_1 + r_2 + r_3}, \pi_3 = \frac{r_3}{r_1 + r_2 + r_3} \tag{60}$$

We find the following expression for the constant A

$$\begin{aligned} & A = \frac{1}{(r_1 m_1 t_1 + r_2 m_2 t_2 + r_3 m_3 t_3)^3} \\ & \Big((r_1 m_1 + r_2 m_2 + r_3 m_3 r_3)^2(r_1 m_1 \sigma_{T_1}^2 + r_2 m_2 \sigma_{T_2}^2 + r_3 m_3 \sigma_{T_3}^2) + \\ & r_1\sigma_{M_1}^2\big(r_2 m_2(t_1 - t_2) + r_3 m_3(t_1 - t_3)\big)^2 + \\ & r_2\sigma_{M_2}^2\big(r_1 m_1(t_2 - t_1) + r_3 m_3(t_2 - t_3)\big)^2 + \\ & r_3\sigma_{M_3}^2\big(r_1 m_1(t_3 - t_1) + r_2 m_2(t_3 - t_2)\big)^2 + \\ & \gamma_{12} m_1^2 m_2^2(t_1 - t_2)^2 + \gamma_{13} m_1^2 m_3^2(t_1 - t_3)^2 + \gamma_{23} m_2^2 m_3^2(t_2 - t_3)^2 + \\ & 2 m_1 m_2 m_3\big(r_1 m_1 \delta_1(t_1 - t_2)(t_1 - t_3) + \\ & r_2 m_2 \delta_2(t_2 - t_3)(t_2 - t_1) + r_3 m_3 \delta_3(t_3 - t_1)(t_3 - t_2)\big)\Big) \end{aligned} \tag{61}$$

where we have defined the parameters

$$\gamma_{ij} = r_i r_j(2 - r_i - r_j) \quad for \quad i, j = 1, 2, 3 \tag{62}$$

and

$$\delta_i = 1 - r_i - \prod_{j=1, j\neq i}^3 (1 - r_j) \quad for \quad i = 1, 2, 3 \tag{63}$$

Observe that we get the two state solution if two of the states have equal mean and variance of their arrival distribution. E.g. if we have $t_2 = t_3 (= t_2^*)$ and $\sigma_{T_2}^2 = \sigma_{T_3}^2 (= \sigma_{T_2^*}^2)$ we obtain the result for $N = 2$ by defining the following weighted mean and variance

$$m_2^* = \frac{r_2}{r_1} m_2 + \frac{r_3}{r_1} m_3 \tag{64}$$

$$\begin{aligned} \sigma_{M_2^*}^2 = & \frac{r_2}{r_1}\sigma_{M_2}^2 + \frac{r_3}{r_1}\sigma_{M_2}^2 + \frac{r_2(2 - r1 - r2)}{r_1^2} m_2^2 + \\ & \frac{r_3(2 - r1 - r3)}{r_1^2} m_3^2 + 2\frac{r_2 + r_3 - r_2 r_3 - r_1}{r_1^2} m_2 m_3 \end{aligned} \tag{65}$$

References

[1] M. Z. Shafiq and et al., "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1960–1973, Dec. 2013.

[2] N. Nikaein and et al., "Simple traffic modeling framework for machine type communication," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, Aug. 2013, pp. 1–5.

[3] D. R. Cox, *Renewal theory*, ser. Methuen's monographs on applied probability and statistics. Methuen and Co. Ltd, 1962.

[4] O. Al-Khatib and et al., "Traffic modeling for machine-to-machine (M2M) last mile wireless access networks," in *Proceedings of the 2014 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014, pp. 1199–1204.

[5] E. Grigoreva and et al., "Coupled Markovian arrival process for automotive machine type communication traffic modeling," May 2017.

[6] A. M. Kshirsagar and Y. P. Gupta, "Asymptotic values of first two moments in Markov renewal processes," *Biometrika*, vol. 54, no. 3-4, pp. 597–603, 1967.

[7] J. J. Hunter, "On the moments of Markov renewal processes," *Advances in Applied Probability*, vol. 1, no. 2, pp. 188–210, 1969.