Towards an Analysis of Traffic Shaping and Policing in Fog Networks Using Stochastic Fluid Models

Jiaojiao Jiang, Longxiang Gao, Jiong Jin, Tom H. Luan, Shui Yu,
Dong Yuan, Yong Xiang, and Dongfeng Yuan*

ABSTRACT

This paper gives models and analytic techniques for studying shaping and policing data traffic in fog networks. The traffic in these networks is expected to be highly diverse and bursty, and regulation will be required as an integral part of congestion control. We generalize the Leaky Bucket model to shape and police traffic source for rate-based congestion control in high-speed fog networks. In particular, the Markov modulated fluid sources reflect the bursty characteristics of data traffic. To measure the performance of the model in shaping and policing traffic, we derive four performance metrics. The experimental results show that with proper design the Leaky Bucket model effectively controls a 4-way trade-off between throughput, loss probability, delay and burstiness of data traffic. Numerical results also reveal that the model performance is sensitive to certain traffic source characteristics.

ACM Reference Format:
Jiaojiao Jiang, Longxiang Gao, Jiong Jin, Tom H. Luan, Shui Yu, Dong Yuan, Yong Xiang, and Dongfeng Yuan. 2017. Towards an Analysis of Traffic Shaping and Policing in Fog Networks Using Stochastic Fluid Models. In *Proceedings of ACM Woodstock conference (Mobiquitous'17).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Fog computing has recently emerged as a new technology to enable the smooth data delivery and real-time data processing between fog services and mobile users [14]. The Fog computing extends Cloud services by deploying Fog servers at the physical proximity of mobile users, *e.g.*, parks and shopping malls. The fog network was initiated by Cisco to enable the fog computing technology at the edge of the network [2]. The main characteristics of fog network include ubiquity, decentralized management and cooperation [16].

*Jiaojiao Jiang and Jiong Jin are with the School of Software and Electrical Engineering, Swinburne University of Technology, Australia. Longxiang Gao, Shui Yu and Yong Xiang are with the School of Information Technology, Deakin University, Australia. Tom H. Luan is with the National Key Laboratory of Integrated Networks Services, Xidian University, China. Dong Yuan is with the School of Electrical and Information Engineering, University of Sydney, Australia. Dongfeng Yuan is with the School of Information Science and Engineering, Shandong University, China.

MobiQuitous 2017, November 7–10, 2017, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-5368-7/17/11…$15.00
https://doi.org/10.1145/3144457.3144496

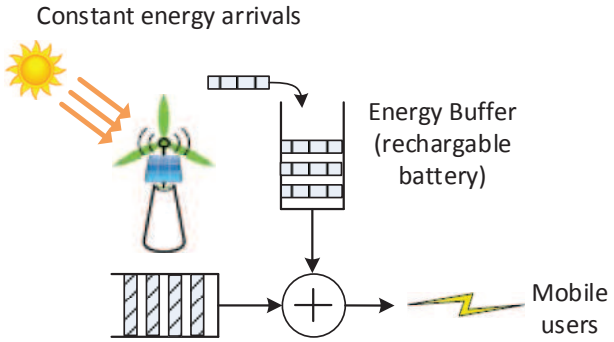

Figure 1: Structure of a fog node.

The explosively growing demand for ubiquitous mobile devices has led to a significant increase in data traffic policing and energy consumption by fog networks. Many papers have been devoted to address energy consumption issues [6–8]. For example, in a recent work [3], researchers developed an analytical framework on the energy buffer using a G/G/1(/N) queuing model. The model considers a random energy charging process of a general energy arrival pattern (e.g., energy from heterogeneous green energy sources) and a general discharging process that fits different mobile applications. However, the high diversity and burstiness of data traffic are not considered in current works. The inherent flexibility and high burstiness of data traffic of fog networks, such as transmitting videos and voice, make the shaping and policing of traffic control problems of such networks very critical.

In general, a node in fog networks is mainly composed of an energy buffer (such as rechargeable battery) and a data buffer as shown in Fig 1. The energy buffer stores the energy charged at a steady rate; with adequate energy storage, data traffic in the data buffer could be transmitted and consume energy. The data buffer therefore varies from time to time and is fundamental to the data transmission policing and shaping.

In this paper, we adopt and generalized the Leaky Bucket stochastic fluid model [12] to shape and police bursty data traffic. In particular, the involved Markov modulated fluid sources reflect the bursty characteristics of data traffic in fog networks. The Leaky Bucket corresponds to a counter, which is incremented each time a cell is generated by the source and is decremented periodically with a suitable leaky rate. The Leaky Bucket can be analyzed as a G/D/1/N queue with finite waiting room N and a suitable arrival process. Each active virtual connection has its own counter. Fig. 2 portrays the Leaky Bucket model on a fog node. In order to analyze

the performance of the Leaky Bucket model, we derive four different performance metrics. The metrics are all derived in explicit formulas:

- Average throughput ($\langle\lambda\rangle$) is the average data or information units transmitted through a fog node. In particular, we obtain the dependence of $\langle\lambda\rangle$ on other data source characteristics.
- Loss probability (P_L) computes the probability of data failed to be transmitted due to the extremely high burstiness of data traffic.
- Average delay ($\widehat{delay}$) measures the average number of time units required to transmit a given data buffer.
- Squared coefficient of variation (SCV) calculates the burstiness of output/transmitted data rate.

We analyze each of these metrics and the relationship among them. We notice that throughput $\langle\lambda\rangle$ and loss probability P_L depend on energy arrival rate r (see Fig. 2) and the total buffer of data buffer and energy buffer, $B = B_D + B_E$. To achieve a desired value of $\langle\lambda\rangle$ or P_L, the total buffer B (rather than individual B_D or B_E) needs to increase with energy arrival rate r. The average delay $\widehat{delay}$ and squared coefficient of variation (SCV) depends on energy arrival rate r and energy buffer size B_E (see Fig. 2). To achieve a desired value of $\widehat{delay}$ or SCV, the energy buffer size B_E needs to increase with r. Furthermore, We theoretically and experimentally show that the traffic source parameters provide a four-way trade-off between average throughput $\langle\lambda\rangle$, data cell loss probability P_L, average data delay $\widehat{delay}$ and burstiness of output rate SCV.

The rest of this paper is organized as follows. The analysis of the Leakey Bucket model is presented in Section 2. We derive the four performance metrics in Section 3. Experimental results of the model performance are presented in Section 4. Section 5 concludes the remarks of this paper.

2 STOCHASTIC FLUID MODEL OF DATA BUFFERING

In this section, we generalize the Leaky Bucket mechanism to model traffic shaping and policing in fog networks. The "Leaky Bucket algorithm" [12, 15], and its performance have been analyzed world wide [9, 13, 17, 18]. In general, the Leaky Bucket algorithm is characterized by its bucket depth or threshold (B_D) and its leak rate (B_E) as shown in Fig. 2. It can be analyzed as a G/D/1/N queue with finite waiting room N and a suitable arrival process. The basic idea is that a certain amount of fluid is added to the bucket contents each time a cell enters the network. The bucket leaks at a constant rate set equal to the cell rate as agreed for that particular connection. If cells are sent at a higher rate than the leak rate, the level of the fluid inside the bucket will rise until a certain critical level (the bucket limit) is exceeded. Then it is concluded that the connection violates the agreed rate and the cell is discarded. For discarded cells no fluid is added. All cells for that connection will be blocked until the bucket level has dropped below the limit. Hence, Leaky Bucket algorithm has been commonly used in monitoring peak and average traffic rates of variable bitrate (VBR) services. Through setting the average traffic rates according to the Leaky Bucket algorithm and thereafter we can control traffic.

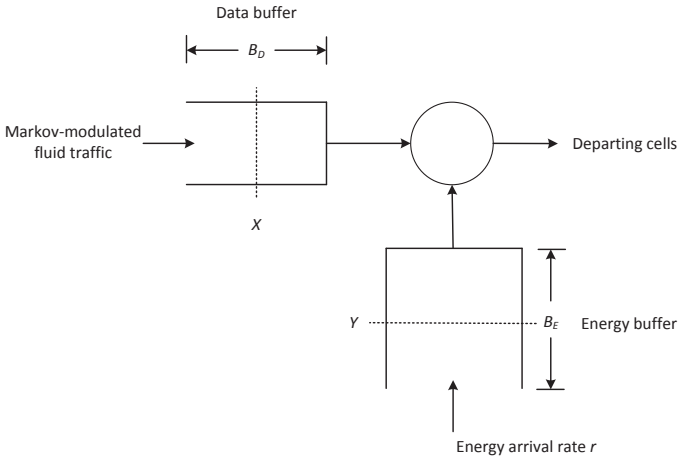

Figure 2: Fluid-flow analysis.

2.1 The Stochastic Fluid Model

In this subsection, we carry out a general analysis of leaky bucket algorithm from which performance metrics can be readily derived. Fig. 2 portrays a fluid analysis version of the data traffic. The energy arrival rate and the energy bucket size are r and B_E respectively, and the data buffer size is B_D. The incoming bursty data traffic is modeled as an N-state Markov modulated fluid flow (see Fig. 3). Transition from state i to state j is governed by a Markov chain with a rate parameter μ_{ij} for $1 \le i, j \le N$, and the input rate of the fluid data at state i is λ_i. The underlying continuous time Markov chain of the fluid data for its state transition can be represented by an $N \times N$ infinitesimal generating matrix M, defined by

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1j} & \cdots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2j} & \cdots & \mu_{2N} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mu_{i1} & \mu_{i2} & \cdots & \mu_{ij} & \cdots & \mu_{iN} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mu_{N1} & \mu_{N2} & \cdots & \mu_{Nj} & \cdots & \mu_{NN} \end{bmatrix}, \quad (1)$$

where, the diagonal rate parameter μ_{ii} is given as

$$\mu_{ii} = -\sum_{j=1, j\neq i}^{N} \mu_{ij}, \ 1 \le i \le N. \quad (2)$$

Hence, the steady-state probability π_i that the Markov chain is in state i can then be derived by solving the following equation

$$\boldsymbol{\pi} M = 0, \quad (3)$$

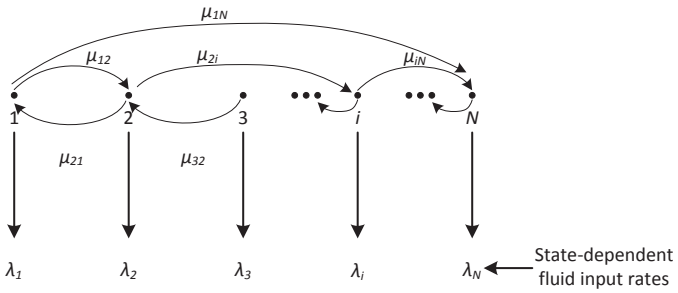

Figure 3: Markov-modulated fluid data model.

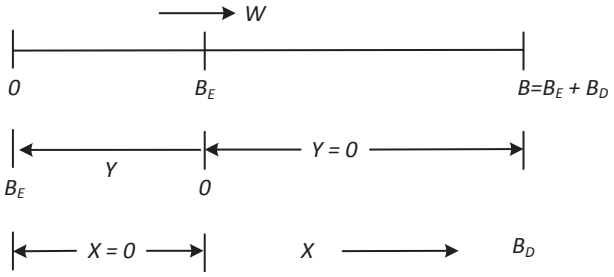

Figure 4: "Virtual buffer" variable W_t.

where, the row vector $\boldsymbol{\pi}$ is of the N steady state probabilities

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_i & \cdots & \pi_N \end{bmatrix}$$

and a normalization condition $\sum_{i=1}^{N} \pi_i = 1$.

Assume two random variables X_t and Y_t are the contents of the data buffer and the energy buffer at time t, where $0 \le X_t \le B_D$ and $0 \le Y_t \le B_E$. The objective is now to find the steady-state occupancy statistics of the two buffers, X_t and Y_t for the data and energy buffers respectively. As soon as the buffer occupancy statistics of X_t and Y_t are known, the cell loss and other performance metrics can be found. In practice, however, the statistics of X_t and Y_t cannot be found separately since these two random variables are highly related to each other. Instead, we apply the method in [4], and define an equivalent "virtual" queue representation of X_t and Y_t jointly. Note the following two facts from the Leaky Bucket technique:

(1) The data buffer can be occupied only if the energy queue is empty ($Y_t = 0$); otherwise the data cells would be transmitted, one per energy buffer ($X_t > 0$). Hence,

$$Y_t = 0, \ X_t > 0. \tag{4}$$

(2) Conversely, the energy buffer can be occupied only if the data buffer is empty ($X_t = 0$). If data cells were queued, they would each capture a unit of energy and be transmitted ($Y_t > 0$). Hence,

$$X_t = 0, \ Y_t > 0. \tag{5}$$

Combining Eqs. (4) and (5), we obtain

$$X_t Y_t = 0. \tag{6}$$

Therefore, instead of finding the individual statistics of X_t or Y_t directly, we shall find the joint probability first from which the data cell loss is then derived.

Following the approach of [4], by combining the two buffers of X_t and Y_t together, we now define the following single "virtual buffer" random variable W_t

$$W_t \equiv X_t - Y_t + B_E. \tag{7}$$

Given the ranges of X_t and Y_t and the condition in Eqs. (4) and (5), by defining B as the sum of the two buffer sizes,

$$B \equiv B_D + B_E. \tag{8}$$

We note that W_t has the following properties:

(1) $0 \le W_t \le B$.
(2) In the range $0 \le W_t \le B_E$,

$$X_t = 0, \ W_t = B_E - Y_t. \tag{9}$$

(3) In the range $B_E \le W_t \le B = B_E + B_D$,

$$Y_t = 0, \ W_t = X_t + B_E. \tag{10}$$

Once the statistics of W_t is found, we can find the statistics of X_t and Y_t using the second and the third property above. The range of W_t and its relation to Y_t are diagrammed in Fig. 4. Define the joint probability of w,

$$F_i(w) = \Pr\left[W_t \le w, \ S = i\right], \tag{11}$$

at time t at state i of the input data Markov chain, $1 \le i \le N$. Note from the definition of W_t that this implies

$$F_i(B) = \pi_i. \tag{12}$$

Using the approach in [1], the solution for $F_i(w)$ is given by the solutions to the following set of differential equations,

$$(\lambda_i - r)\frac{F_i(w)}{dw} = \sum_{j=1}^{N} \mu_{ji} F_j(w), \ 1 \le i \le N. \tag{13}$$

It is straightforward to obtain the governing differential equations:

$$\frac{d\boldsymbol{F}(w)}{dw} D = \boldsymbol{F}(w) M, \tag{14}$$

where

$$\boldsymbol{F}(w) = \begin{bmatrix} F_1(w) & F_1(w) & \cdots & F_N(w) \end{bmatrix},$$

M is the matrix of Eq. (1), and D is $N \times N$ diagonal matrix

$$D \equiv \mathrm{Diag}\left[\lambda_i - r\right]. \tag{15}$$

Let scalar $F(w)$ be the sum of state probability $F_i(w)$,

$$F(w) = \sum_{i=1}^{N} F_i(w). \tag{16}$$

Through the analysis in [10], the solution to Eq. (14) can be found as follows:

$$d\boldsymbol{F}(w) = \sum_{j=1}^{N} a_j \boldsymbol{\Phi}_j e^{z_j w}, \tag{17}$$

where, $(z_j, \boldsymbol{\Phi}_j)$ is the (eigenvalue, eigenvector) pair satisfying the eigenvalue equation

$$z_j \boldsymbol{\Phi}_j D = \boldsymbol{\Phi}_j M, \ 1 \le j \le N. \tag{18}$$

The a_j's, $1 \leq j \leq N$, are constants to be determined by invoking N boundary conditions. Since $\boldsymbol{\pi} M = 0$, one eigenvalue of Eq. (18) must be zero. Calling this eigenvalue z_1, its associated eigenvector $\boldsymbol{\Phi}_1 = \boldsymbol{\pi}$. Hence, Eq. (17) can be simplified to

$$\boldsymbol{F}(w) = a_1 \boldsymbol{\pi} + \sum_{j=2}^{N} a_j \boldsymbol{\Phi}_j e^{z_j w}. \tag{19}$$

We now need N boundary conditions to find the unknown constants a_j, $1 \leq j \leq N$. To establish these, we note that some of the Markov chain states must be underload or "downward" states for those states i, $\lambda_i < r$; the others must be overload or "upward" states for those States i, $\lambda_i > r$. Hence, all N states of the Markov chain modulating the data arrival rate can be divided into two disjoint sets,

$$S_U \equiv \{i \in N | \lambda_i - r > 0\} : \text{upward states}; \tag{20}$$

and

$$S_D \equiv \{i \in N | \lambda_i - r < 0\} : \text{downward states}. \tag{21}$$

Consider an arbitrary downward state, $i \in S_D$. Since, the data arrival rate is less than the energy arrival rate in these states, $\lambda_i < r$, the energy is tending to collect. The data queue is tending to empty ($X_t \to 0$) and the energy queue is tending to fill ($Y_t \to B_E$). Hence,

$$W_t = X_t - Y_t + B_E \to 0 \tag{22}$$

For these states then, the probability the "virtual queue" is full tends to zero, or we can write

$$\Pr[W_t = B, S = i] = 0, i \in S_D \tag{23}$$

As Eq. (23) is the PDF (Probability density Function) of W, we note that $\Pr[W_t = B, S = i]$ is equal to the difference between the CDFs (Cumulative Distribution Function) $\Pr[W \leq B, S = i] - \Pr[W_t \leq B^-, S = i]$, *i.e.*, $F_i(B) - F_i(B^-)$. Note that B^- is our notation for the largest value less than B in the PDF of W_t. Given Eq. (12), we thus have

$$F_i(B^-) = \pi_i, \ i \in S_D. \tag{24}$$

Consider an arbitrary upward state, $i \in S_U$. For these states, with $\lambda_i > r$, the energy buffer tends to empty ($Y_t \to 0$), the data buffer tends to fill ($X_t \to B_D$), and

$$W_t = X_t - Y_t + B_E \to B_D + B_E = B. \tag{25}$$

For these states, then, the probability that the virtual queue is empty must be zero. We thus have

$$F_i(0^+) = \Pr[W_t \leq 0^+, S = i] = 0, i \in S_U. \tag{26}$$

Note that 0^+ is our notation for the smallest value larger than 0 in the PDF of W_t.

Since a state is either an downward state or upward state, Eqs. (24) and (26) provide the necessary N boundary conditions from which to find the N unknown constants a_j, $1 \leq j \leq N$, of Eq. (19). The N equations to be solved for the N unknown a_j could be written out in scalar form as follows:

(1) $i \in S_D$:

$$F_i(B^-) = \pi_i = a_1 \pi_i + \sum_{j=2}^{N} a_j \Phi_{ji} e^{z_j B^-}. \tag{27}$$

(2) $i \in S_U$:

$$F_i(0^+) = 0 = a_1 \pi_i + \sum_{j=2}^{N} a_j \Phi_{ji}. \tag{28}$$

Having found the distributions of W_t, we now can extract from it the distributions of the energy buffer content Y_t and the data buffer content X_t. Specifically,

$$\Pr[\text{data buffer full}] = 1 - F(B), \tag{29}$$

$$\Pr[\text{energy buffer full}] = F(0), \tag{30}$$

$$\Pr[Y_t \leq y, S = i] = w_i - F_i(B_E - y), \ 0 \leq y \leq B_E, \tag{31}$$

and

$$\Pr[X_t \leq x, S = i] = F_i(x + B_E), \ 0 \leq x \leq B_D, \tag{32}$$

where, $F_i(\cdot)$ and $F(\cdot)$ are defined in Eqs. (11) and (16), respectively.

2.2 Example of On-Off Data Sources

We now apply this analysis to the simplest example, the basic on-off data source (see Fig. 5), where the inter-arrival rates are 0 and R_p in the OFF state and the ON state respectively. This could be representative of a voice data source or an image data source, depending on the choice of parameters. Also assume the transition rate from OFF state to ON state is α and from ON state to OFF state is β. In this model, note that the OFF state refers to state 1 and ON state to state 2 in the analysis. Notice that $\lambda_1 = 0$, $\lambda_1 = R_p$, and the continuous time Markov chain generating matrix and the steady state probability row vector are as follows

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix},$$

and

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix}.$$

Here, there is only one eigenvalue z to be found. Through basic computation, we have

$$z = -\frac{\alpha + \beta}{R_p - r}(1 - \rho), \tag{33}$$

where

$$\rho = \frac{R_p}{r}\left(\frac{\alpha}{\alpha + \beta}\right) = \frac{R_p p}{r}, \text{ with } p \equiv \frac{\alpha}{\alpha + \beta}. \tag{34}$$

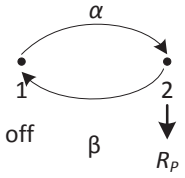

Figure 5: Simple on-off fluid data source.

For the single eigenvector $\mathbf{\Phi} = [\Phi_1, \Phi_2]$ appearing in Eq. (28), we have

$$\Phi_1/\Phi_2 = (R_p - r)/r = \left(\frac{R_p}{r} - 1\right). \tag{35}$$

Letting $\Phi_2 = 1$ arbitrarily, we get, for this example,

$$\boldsymbol{F}(x) = a_1\boldsymbol{\pi} + a_2\left[\left(\frac{R_p}{r} - 1\right), 1\right]e^{zx}. \tag{36}$$

The two unknown constants a_1 and a_2 are found using the boundary conditions of Eqs. (27) and (28). In this simple example, with $N = 2$ states only, state one with $\lambda_1 = 0$ is the downward state and state two with $\lambda_2 = R_p$ is the upward state. We must thus have $0 < r < R_p$ for the fluid analysis to provide a stationary solution in this case. This implies that the single eigenvalue z of Eq. (33) will be negative if the parameter ρ defined by Eq. (34) is less than 1.

From Eq. (27), we have, using Eq. (36),

$$F_1(B^-) = \pi_1 = a_1\pi_1 + a_2\left(\frac{R_p}{r} - 1\right)e^{zB^-}, \tag{37}$$

with $\pi_1 = (1 - p) = \beta/(\alpha + \beta)$.

Similarly, from Eqs. (28) and (36), we have

$$F_1(0^+) = 0 = a_1\pi_2 + a_2, \tag{38}$$

where $\pi_2 = p = \alpha/(\alpha + \beta)$.

Solving Eqs. (37) and (38) simultaneously for a_1 and a_2, we get

$$a_1 = 1/\left[1 - \frac{\alpha}{\beta}\left(\frac{R_p}{r} - 1\right)e^{zB}\right], \tag{39}$$

and

$$a_2 = -\pi_2 a_1 = -a_1\alpha/(\alpha + \beta). \tag{40}$$

Using Eqs. (39) and (40) in Eq. (36), we finally get for the two probability distribution functions in this case of a single on-off data source,

$$F_1(w) = \pi_1\Delta(w)/\Delta(B), \tag{41}$$

and

$$F_2(w) = \pi_2(1 - e^{zw})/\Delta(B), \tag{42}$$

with

$$\Delta(w) \equiv 1 - \frac{\alpha}{\beta}\left(\frac{R_p}{r} - 1\right)e^{zw}, \tag{43}$$

where $\pi_1 = \beta/(\alpha + \beta)$, $\pi_2 = 1 - \pi_1 = \alpha/(\alpha + \beta)$, and z are given by Eq. (33).

All the performance parameters of interest can be obtained from $F_1(w)$ and $F_2(w)$. We shall obtain these after first developing performance parameters for the general N-state Markov chain model.

3 PERFORMANCE METRICS

We study the performance of the Leaky Bucket model as measured by the average throughput $\langle\lambda\rangle$, the cell loss probability P_L, the squared coefficient of variation of the output rate SCV, and the average data delay $\widehat{delay}$; we also investigate the trade-offs as functions of the model parameters B_D, B_E and r. The on-off source model is used for illustration and dependence of the metrics on model parameters are analyzed.

3.1 The Average Throughput

Consider the average throughput $\langle\lambda\rangle$. This can be calculated in a number of equivalent ways. One way is to say that this is the average load, less the cell loss when the data buffer is full. This could be written as

$$\langle\lambda\rangle = \sum_{i=1}^{N}\lambda_i\pi_i - \sum_{i=1}^{N}(\lambda_i - r)\Pr[X_t = B_D, S = i]. \tag{44}$$

The first term on the right-hand side is the average load, averaged over all N states. The second term is the average data cell loss, also averaged over the N states. However, note that the upward states only will be included in the sum since we have already seen that in the downward states the data buffer cannot be full and cells cannot be lost (see Eq. (23)). The probability $\Pr[X_t = B_D, S = i]$ that the data buffer is full with the system in overload state i, is readily calculated as

$$\begin{aligned}&\Pr[X_t = B_D, S = i]\\ &= \Pr[W_t = B, S = i]\\ &= \Pr[W_t \le B, S = i] - \Pr[W_t \le B^-, S = i]\\ &= \pi_i - F_i(B^-).\end{aligned} \tag{45}$$

We thus have, from Eq. (44),

$$\langle\lambda\rangle = \sum_{i=1}^{N}\lambda_i\pi_i - \sum_{i=1}^{N}(\lambda_i - r)\left[\pi_i - F_i(B^-)\right]. \tag{46}$$

This can be rewritten as follows,

$$\langle\lambda\rangle = r + \sum_{i=1}^{N}F_i(B^-)(\lambda_i - r). \tag{47}$$

An alternate approach is to focus on average energy throughput. Since every cell requires a unit of energy for transmission, the average energy throughput must equal average cell throughput. We thus have, immediately,

$$\langle\lambda\rangle = r - \sum_{i=1}^{N}(r - \lambda_i)F_i(0). \tag{48}$$

The second term here represents a mean cell "loss" due to a full energy buffer, since

$$\Pr[Y_t = B_E, S = i] = \Pr[W_t = 0, S = i]. \tag{49}$$

This implies that

$$\sum_{i=1}^{N}(\lambda_i - r)F_i(B^-) = \sum_{i=1}^{N}(\lambda_i - r)F_i(0). \tag{50}$$

Consider the on-off data source model in Fig. 5. Recall that $\lambda_1 = 0$, $\lambda_2 = R_p$, $\rho \equiv R_p p/r$, $p = \alpha/(\alpha + \beta)$ in this case. Then it is readily shown, using Eqs. (41) and (42) in either (47) or (48), that the normalized throughput for this example is given by

$$\langle\lambda\rangle/r = 1 - (1 - \rho)/\Delta(B), \tag{51}$$

where, $\Delta(B)$ is defined in Eq. (43).

3.2 Data Cell Loss Probability

Once the throughput has been calculated, it is straightforward to obtain the data cell loss probability P_L:

$$P_L = 1 - \langle\lambda\rangle / \sum_{i=1}^{N} \lambda_i \pi_i. \tag{52}$$

If we apply this analysis to the on-off data source (see Fig. 5). The performance metric of average throughput $\langle\lambda\rangle$ is the average load, *i.e.*, $\sum_{i=1}^{2} \lambda_i \pi_i$, less the load gets rejected when the data buffer is full, *i.e.*, $\sum_{i=1}^{2} (\lambda_i - r)\Pr[X_t = B_D, S = i]$. Given Eq. (25) and the properties of the virtual buffer W_t, the average throughput can finally be found in the following equation

$$\langle\lambda\rangle = r - \frac{(1-\rho)r}{\Delta(B)}. \tag{53}$$

The cell loss probability P_L, which is 1 minus the ratio of the average throughput $\langle\lambda\rangle$ to the average load $\sum_{i=1}^{2} \lambda_i \pi_i$, can be obtained in the following equation

$$P_L = \frac{1 - \Delta(B)}{\rho} = \frac{\alpha}{\rho\beta}\left(\frac{R_P}{r} - 1\right) e^{zB}, \tag{54}$$

where z is the eigenvalue defined in Eq. (33). Cell loss could happen only when $R_p > R$ and thus $z < 0$ in this range.

3.3 The Average Data Delay

Given data buffer size B_D and energy arrival rate r, we see that it requires B_D/r units of time for the data to transmit. Then, we can obtain the distribution of data delay by arriving data as follows

$$\Pr[delay \le t] = \frac{1}{\langle\lambda\rangle} \sum_{i=1}^{N} \lambda_i F_i(rt + B_E), \text{ if } t < \frac{B_D}{r}, \tag{55}$$

and

$$\Pr[delay = \frac{B_D}{r}] = \frac{r}{\langle\lambda\rangle}[1 - \sum_{i=1}^{N} F_i(B)], \tag{56}$$

where $F_i(\cdot)$ is defined in Eq. (11). Therefore, the average data delay can be obtained from

$$\widehat{delay} = \sum_{t} t \cdot \Pr[delay = t]. \tag{57}$$

Consider the on-off data source in Fig. 5. Recall that $\lambda_1 = 0$ and $\lambda_2 = R_p$ in this case. Using Eq. (41), we obtain

$$\Pr[delay \le t] = \frac{1}{\langle\lambda\rangle} R_p F_2(rt + B_E), \text{ if } t < \frac{B_D}{r}, \tag{58}$$

and

$$\Pr[delay = \frac{B_D}{r}] = \frac{r}{\langle\lambda\rangle}[1 - F_2(B)]. \tag{59}$$

3.4 The Squared Coefficient of Variation of Output

We finally consider the variability in the output rate (*i.e.*, the burstiness of output traffic). We particularly adopt the squared coefficient of variation [5, 11] to measure the variability. Let SCV denote the squared coefficient of variation

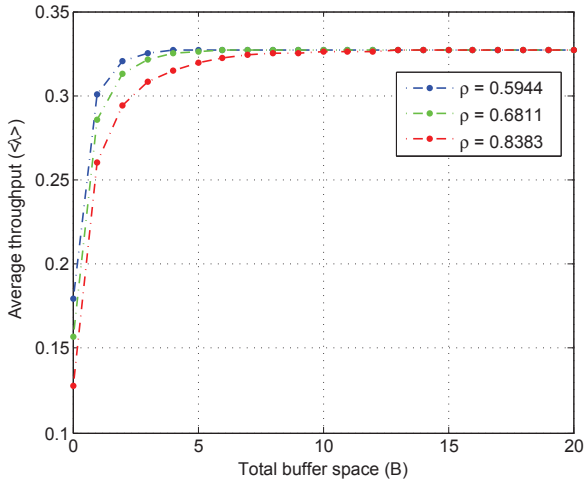

Figure 6: Average throughput as a function of energy arrival rate r and total buffer space B. $\alpha = 0.49$, $\beta = 1$, $R_p = 1$, $r = 0.55, 0.48, 0.39$ for $\rho = 0.59, 0.68, 0.84$.

of output rate. We, therefore, obtain the following formula for SCV,

$$SCV = \frac{\langle\lambda^2\rangle - \langle\lambda\rangle^2}{\langle\lambda\rangle^2}, \tag{60}$$

where $\langle\lambda\rangle$ is defined in Eq. (53) and $\langle\lambda^2\rangle$, which the average of the squared throughput, is defined as follows

$$\langle\lambda^2\rangle = r^2\left[1 - F(B_E)\right] + \sum_{i=1}^{N} \lambda_i^2 F_i(B_E). \tag{61}$$

Consider the on-off data source in Fig. 5. Recall that $\lambda_1 = 0$ and $\lambda_2 = R_p$ in this case. Using Eqs. (16) and (41), we obtain

$$\langle\lambda^2\rangle = r^2\left[1 - \frac{\left(1 - \frac{R_p\rho}{r}\right) + \rho e^{zB_E}\left(\frac{R_p}{r} - 1\right)}{\Delta(B)}\right]. \tag{62}$$

4 NUMERICAL INVESTIGATION

We study the performance of the Leaky Bucket model as measured by the four performance metrics derived in Section 3, including the average throughput $\langle\lambda\rangle$, data cell loss probability P_L, the average data delay $\widehat{delay}$, and the variation of output burstiness SCV. We also study the trade-offs as functions of the data buffer and energy buffer parameters B_D, B_E and r. The on-off data source model is used and the effect of the average data delay and variability of the on and off periods on the Leaky Bucket performance are also examined. The accuracy of the fluid model and the analytical results were validated in many other works [4, 9, 12, 13, 15, 17, 18], so we do not demonstrate it here.

In the exact Leaky Bucket model, the data source alternates between the on state and the off state. During the on state, whose mean duration is $1/\beta$ seconds, the source transmits data cells (which are with fixed length packets)

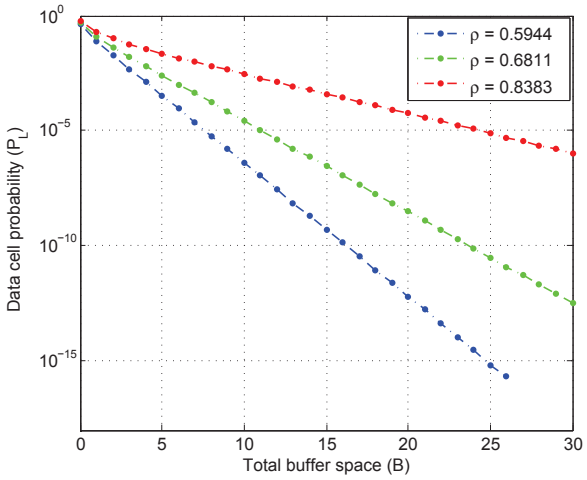

Figure 7: Loss probability as a function of energy arrival rate r and total buffer space B.

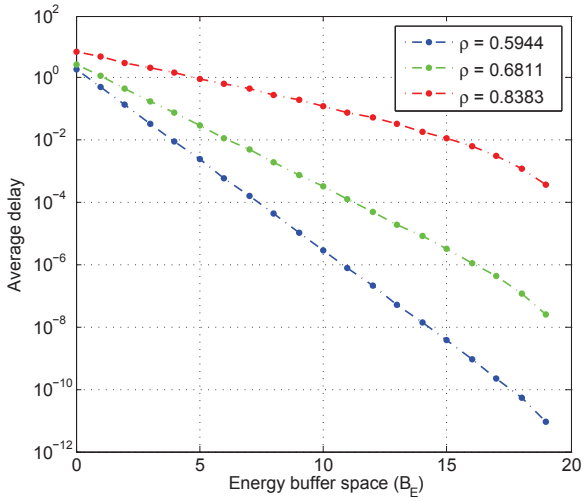

Figure 8: Average delay as a function of energy arrival rate r and energy buffer space B_E, with $B = 20$.

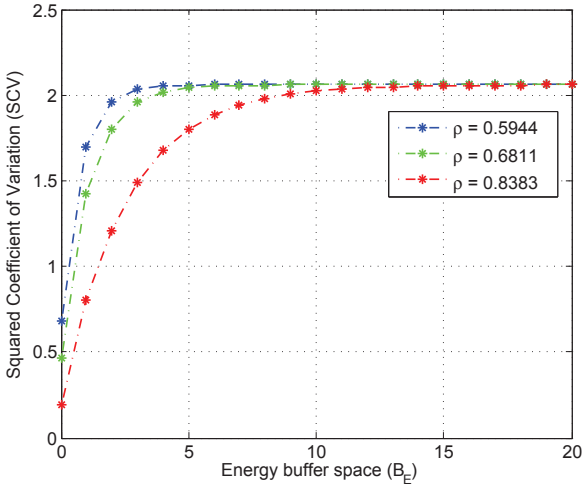

Figure 9: Squared coefficient of variation of output rate as a function of energy buffer space B_E, with $B = 20$.

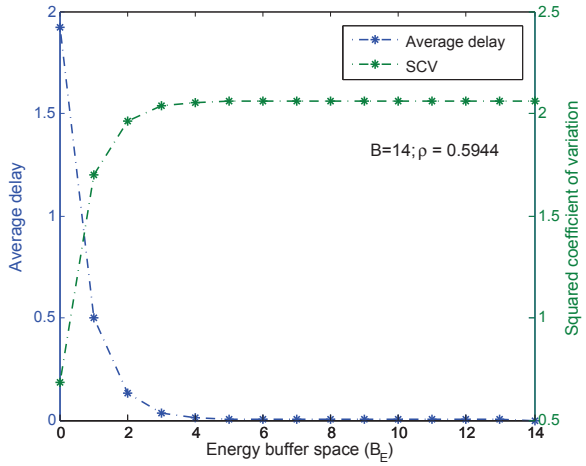

Figure 10: The tradeoff between data cell delay and burstiness of output rate controlled by the partitioning of B into B_D and B_E, with $B = 10, P_L = 1.5818 \times 10^{-6}, r = 0.55$.

at a constant rate of r cells per second. The energy arrives at the energy bucket at a constant rate of r energy buckets per second. In the following experiments, we set β and α to 1.0 and 0.49 respectively, $R_p = 1$, and r is chosen from $\{0.55, 0.48, 0.39\}$.

We first examine the effect of the total buffer size B ($= B_D + B_E$) on the average throughput $\langle\lambda\rangle$ and the data cell loss probability P_L. According to our analysis in Eqs. (51) and (54), we see that $\langle\lambda\rangle$ and P_L depend on the total buffer size B and the energy arrival rate r (or the energy traffic intensity $\rho = \frac{R_p}{r}\left(\frac{\alpha}{\alpha+\beta}\right)$). In our experiments, we increase the total buffer size B from 0 to 20, and let r varies in $\{0.55, 0.48, 0.39\}$ (or alternatively ρ varies in $\{0.59, 0.68, 0.84\}$). The results are shown in Fig. 6 and Fig. 7, respectively. Fig. 6 shows that as ρ approaches 1, the virtual buffer size B required to achieve a desired value of throughput $\langle\lambda\rangle$ increases greatly, especially when $B \leq 10$. Correspondingly, the total buffer size B required to achieve a desired value of P_L also increases sharply (see Fig. 7). This is a undesirable feature, because the larger requirement of B generally means larger mean data cell delay. However, as we will see below, the inherent flexibility in partitioning B into B_D and B_E provides a mean for achieving a reasonable compromise between mean data cell delay and variability while keeping the loss probability constant.

Then, we investigate the dependence of average data delay ($\widehat{delay}$) and the squared coefficient of variation (SCV) of output rate on energy arrival rate (r) and energy buffer size (B_E). From Eqs. (57) and (60), we know that $\widehat{delay}$ depends

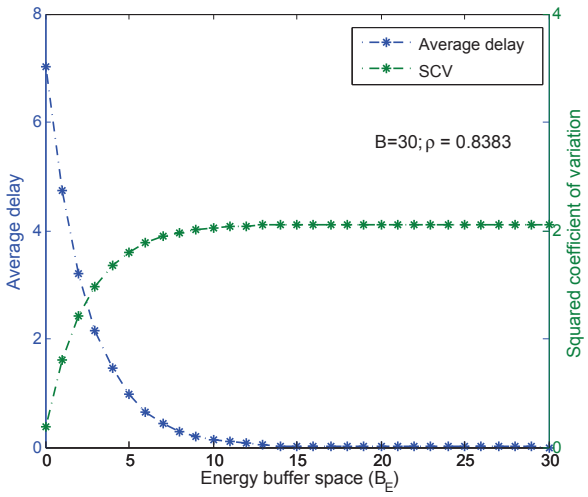

Figure 11: The tradeoff between data cell delay and burstiness of output rate controlled by the partitioning of B into B_D and B_E, with $B = 30, P_L = 1.0807\times10^{-6}, r = 0.39$.

on r, average throughput $\langle\lambda\rangle$, and B_E. We keep the same parameters as we used in Fig. 6 and Fig. 7, while set the maximum total buffer size B as 20 and let B_E increase from 0 to 20. The experimental results are shown in Fig. 8 and Fig. 9, respectively. From Fig. 8, we can see that, generally, the average delay decreases with the increase of energy buffer size B_E. However, as energy traffic intensity ρ approaches 1 (or alternatively r close to the energy source mean rate), the energy buffer size B_E required to achieve a desired value of delay $\widehat{delay}$ increases dramatically. From Fig. 9, we see that, the squared coefficient of variation (SCV) of output rate increases with energy buffer size B_E, and it approaches a steady SCV when B_E is large enough. This is because the greater B_E allows larger data transmission. Hence, the variation of output rate becomes larger. Furthermore, we again notice that the energy buffer size B_E required to achieve a desired value of SCV of output rate increases greatly with energy intensity ρ.

We finally investigate the tradeoff between average data cell delay $\widehat{delay}$ and the squared coefficient of variation of the output rate SCV. The results are displayed in Fig. 10 and Fig. 11 for r equal to 0.55 and 0.39, respectively. In Fig. 10, the total buffer size B is set to 14 units of information giving a loss probability of 1.5818×10^{-6}. As B_E increases from 0 to B, the average delay decreases from a maximum value of 1.98 units of time to 0, while SCV increases from 0.22 to a maximum value of the squared coefficient of variation of the uncontrolled source rate. Similar results are also observed in Fig. 11, where the loss probability is 1.0807×10^{-6}.

5 CONCLUSION

In this paper, we have generalized a stochastic fluid model to analyze the data traffic shaping and policing phase of rate-based congestion control of high-speed fog networks. We have shown how to generalize the well-known Leaky Bucket model police and shape traffic from diverse and bursty data sources. In analyzing the performance of the model, we derived explicit formulas for four different performance metrics, including average throughput, cell loss probability, average data delay and the variation of output rate. On the basis of the Leaky Bucket model and the derived metrics, we perform numerical evaluations with the aim of assessing the effectiveness of the Leaky Bucket mechanism. The experimental results have shown how the Leaky Bucket mechanism can be used to police and shape traffic from bursty data sources. It has been shown that the source parameters provide a four-way trade-off between average throughput, cell loss probability, average data delay and burstiness of output rate. Numerical results have also revealed the sensitivity of performance to source traffic assumptions.

REFERENCES

[1] David Anick, Debasis Mitra, and Man Mohan Sondhi. 1982. Stochastic Theory of a Data-Handling System with Multiple Sources. *Bell Labs Technical Journal* 61, 8 (1982), 1871–1894.
[2] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. 2012. Fog computing and its role in the internet of things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. ACM, 13–16.
[3] Lin X Cai, Yongkang Liu, Tom H Luan, Xuemin Shen, Jon W Mark, and H Vincent Poor. 2014. Sustainability analysis and resource management for wireless mesh networks with renewable energy supplies. *IEEE Journal on Selected Areas in Communications* 32, 2 (2014), 345–355.
[4] Anwar I Elwalid and Debasis Mitra. 1991. Stochastic fluid models in the analysis of access regulation in high speed networks. In *Global Telecommunications Conference, 1991. GLOBECOM'91.'Countdown to the New Millennium. Featuring a Mini-Theme on: Personal Communications Services*. IEEE, 1626–1632.
[5] Brian Everitt and Anders Skrondal. 2002. *The Cambridge dictionary of statistics*. Vol. 106. Cambridge University Press Cambridge.
[6] Longxiang Gao, Tom H Luan, Bo Liu, Wanlei Zhou, and Shui Yu. 2017. Fog computing and its applications in 5g. In *5G Mobile Communications*. Springer, 571–593.
[7] Longxiang Gao, Tom H Luan, Shui Yu, Wanlei Zhou, and Bo Liu. 2017. FogRoute: DTN-based data dissemination model in fog computing. *IEEE Internet of Things Journal* 4, 1 (2017), 225–235.
[8] Tom H Luan, Longxiang Gao, Zhi Li, Yang Xiang, Guiyi Wei, and Limin Sun. 2015. Fog computing: Focusing on mobile users at the edge. *arXiv preprint arXiv:1502.01815* (2015).
[9] Gerd Niestegge. 1990. The leaky bucketpolicing method in the ATM (Asynchronous Transfer Mode) network. *International Journal of Communication Systems* 3, 2 (1990), 187–197.
[10] Ben Noble, James W Daniel, et al. 1988. *Applied linear algebra*. Vol. 3. Prentice-Hall New Jersey.
[11] George F Reed, Freyja Lynn, and Bruce D Meade. 2002. Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and diagnostic laboratory immunology* 9, 6 (2002), 1235–1239.
[12] Mischa Schwartz. 1996. *Broadband integrated networks*. Vol. 19. Prentice Hall PTR New Jersey.
[13] Khosrow Sohraby and Moshe Sidi. 1994. On the performance of bursty and modulated sources subject to leaky bucket rate-based access control schemes. *IEEE Transactions on Communications* 42, 234 (1994), 477–487.
[14] Ivan Stojmenovic, Sheng Wen, Xinyi Huang, and Hao Luan. 2016. An overview of fog computing and its security issues. *Concurrency and Computation: Practice and Experience* 28, 10 (2016), 2991–3005.
[15] Jonathan S Turner. 1986. New direction in communication (or which way in the information age?). *Proc. Int. Zurich Sem. Digit. Communication* (1986), 25–32.

[16] Luis M Vaquero and Luis Rodero-Merino. 2014. Finding your way in the fog: Towards a comprehensive definition of fog computing. *ACM SIGCOMM Computer Communication Review* 44, 5 (2014), 27–32.

[17] Naoaki Yamanaka, Youichi Sato, and Ken-Ichi Sato. 1992. Performance limitation of leaky bucket algorithm for usage parameter control and bandwidth allocation methods. *IEICE Transactions on Communications* 75, 2 (1992), 82–86.

[18] Nanying Yin and Michael G Hluchyj. 1993. Analysis of the leaky bucket algorithm for on-off data sources. *Journal of High Speed Networks* 2, 1 (1993), 81–98.