ON AN EXACT SOLUTION OF THE RATE MATRIX OF QUASI-BIRTH-DEATH PROCESS WITH SMALL NUMBER OF PHASES

Rama Murthy Garimella
International Institute of Information Technology,
Hyderabad, India,
Email: rammurthy@iiit.ac.in

Rumyantsev Alexander
Institute of Applied Mathematical Research
of the Karelian Research Centre RAS;
Petrozavodsk State University,
Petrozavodsk, Russia
Email: ar0@krc.karelia.ru

KEYWORDS

QBD Process, Matrix-Analytic Method, Exact Solution, Internet-of-Things

ABSTRACT

A new method of obtaining exact solution for the rate matrix R in the Matrix-Analytic method in case of the phase state of dimension two is proposed. The method is based on symbolic solution of the determinental polynomial equation, and obtaining a linear matrix equation for the unknown rate matrix R by Cayley–Hamilton theorem. The method is applied to analyze the Energy-Performance tradeoff of an Internet-of-Things device. A new randomized regime switching scheme is proposed, which, as it is shown by means of numerical experiment, provides significant decrease of energy consumption of the system under study.

INTRODUCTION

Univariate polynomial equations naturally arise in many branches of mathematics. An exact symbolic solution of such an equation (in terms of arithmetic operations and radicals) is known to exist for polynomials of order less or equal to four. Moreover, the nonexistence of such a solution for a polynomial equation of order greater or equal to five was established in Abel–Ruffini theorem. The solvability concept of an arbitrary monic polynomial was provided by Galois group theory.

A generalization of polynomial equations in which the coefficients and the argument are matrices (matrix polynomial equations) has been studied in a number of works, and the detailed theory of uni-variate matrix polynomial equations was developed [15]. In the research area of Queueing Theory, the matrix quadratic equation

$$R^2A^{(2)} + RA^{(1)} + A^{(0)} = 0 \quad (1)$$

was first used to find a solution of a QBD process (by means of Complex Analysis-based method) in late 60's. Wallace [34] and Evans [10] showed that in the case of QBD process, matrix geometric solution exists for the equilibrium distribution where the rate matrix R is a minimal nonnegative solution of matrix quadratic equation (1). M. Neuts generalized the result to arbitrary $G/M/1$-type Markov processes and showed [24] that R is the minimal non-negative solution of matrix power series equation

$$\sum_{i=0}^{\infty} R^i A^{(i)} = 0. \quad (2)$$

Thus, the analysis is essentially reduced to obtaining the matrix R. In general, a few iterative procedures, claimed to be numerically stable, are used [24, 18, 6] (see also the comparison of iterative procedures [17]). Alternatively, the spectral decomposition-based methods were suggested [8, 21] (which required eigenanalysis of a matrix polynomial), some of them utilizing special structure of the model to dramatically decrease the computational complexity [9].

As opposed to probabilistic-based iterative procedures, the Jordan canonical form representation of rate matrix R has been suggested as a closed form/analytic solution of (2). H.R. Gail et al. [13] suggested Spectral Analysis method based on Jordan canonical form to analyze G/M/1- and M/G/1-type Markov chains. G. Rama Murthy [26] successfully proposed method for computing the Jordan form representation of R (i.e computing eigenvalues and generalized eigenvectors of R). He thus showed the relationship between classical Complex Analysis method and the iterative procedures.

However, an explicit formula for R in terms of the given matrices $A^{(0)}, A^{(1)}, A^{(2)}$ for the QBD process is, to the best of our knowledge, not available in general. The explicit expression for R was obtained for a number of special cases, in particular, when either $A^{(0)}$, or $A^{(2)}$ is a rank-one matrix [18, 14]. A natural question that remained was whether there are other cases where R can be expressed explicitly. In this paper we answer such a question when $A^{(0)}, A^{(1)}, A^{(2)}$ and, consequently, R are 2x2 matrices.

We apply the proposed algorithm of the exact solution of (1) to suggest a new simple approach to improve energy efficiency of battery-powered Internet-of-Things (IoT) devices. IoT is an intensively studied field, where energy efficiency is one of the dominant areas. The realization of many IoT applications relies on the wireless network of small battery-powered devices (such as wireless sensors, transmitter nodes, actuators etc.), with a typical battery lifetime varying from days to several years.

Proceedings 31st European Conference on Modelling and
Simulation ©ECMS Zita Zoltay Paprika, Péter Horák, Kata Váradi,
Péter Tamás Zwierczyk, Ágnes Vidovics-Dancs, János Péter Rádics (Editors)
ISBN: 978-0-9932440-4-9/ ISBN: 978-0-9932440-5-6 (CD)

Given that, a typical IoT device is expected to work in energy saving mode, though providing a required quality of service (QoS). Recent research covers various aspects of energy-efficient IoT network design, such as efficient network architecture [31, 32], implementation of efficient routing/clustering algorithms [7, 5, 29, 33] (see also a recent survey [1]). Our application is focused on peer-to-peer type wireless networks, which consist of basic low-price low-powered devices (nodes) which use no (or only basic) centralized management and possess strict energy consumption restrictions. Under this assumptions, implementation of the aforementioned sophisticated routing schemes and network architecture seems irrelevant. Instead, we propose a simple randomized regime switching scheme, which, once implemented at each node, provides significant decrease of energy consumption of the system under study.

We also note, that the proposed randomized management approach may be applied to systems, where cost effectiveness and service elasticity is important, such as high-performance and cloud-based computing systems [22], as well as teletraffic systems (e.g. on-demand content servers), where the operational cost (e.g. energy cost, or cloud service cost), as well as the system speed, is to be adopted to the working conditions. The approach is suitable for heavy load conditions, since the centralized management (which could become a bottleneck once implemented) is unnecessary.

This research paper is organized as follows. First, we present an algebraic approach on obtaining the matrix R. Next, we apply this approach to solve the optimization problem related to Energy-Performance tradeoff in the field of Internet-of-Things. We illustrate the approach with simulation results.

EXACT SOLUTION FOR RATE MATRIX OF A QBD PROCESS

The QBD process is a continuous time Markov process $\{(X(t), Y(t)), t \geqslant 0\}$ with countable state space $E := \{(0, j), j = 1, \dots, m_0, (i, j), i \geqslant 1, j = 1, \dots, m\}$, where the *phase* variable $Y(t)$ may take one of m (or m_0 for boundary states) values and *level* variable $X(t)$ is increased/decreased by at most one at each transition. The state space E can be partitioned into *levels* with level $n \geqslant 1$ being the subset $\{(n, j), j = 1, \dots, m\} \subset E$. The infinitesimal generator matrix of a QBD process has the following block-tridiagonal representation [18]

$$Q = \begin{pmatrix} A^{0,0} & A^{0,1} & 0 & 0 & \dots \\ A^{1,0} & A^{1,1} & A^{(0)} & 0 & \dots \\ 0 & A^{(2)} & A^{(1)} & A^{(0)} & \dots \\ 0 & 0 & A^{(2)} & A^{(1)} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (3)$$

where $A^{(i)}, i = 0, 1, 2$ are square matrices of order m, $A^{0,0}$ is a square matrix of order m_0 and $A^{1,0}, A^{0,1}$ are possibly rectangular matrices. Note that the matrix $A := A^{(0)} + A^{(1)} + A^{(2)}$ necessarily satisfies the balance condition

$$A\mathbf{1} = 0, \quad (4)$$

where $\mathbf{1}$ is the vector of ones of corresponding dimension. Recall also that $A^{(0)} \geqslant 0, A^{(2)} \geqslant 0$ componentwise, and $A^{(1)}_{i,j} \geqslant 0$ for $i \neq j$, whereas $A^{(1)}_{i,i} \leqslant 0$. Note that it readily follows, that for any $i = 1, \dots, m$,

$$|A^{(1)}_{i,i}| = \sum_{j=1}^{m} \Big[A^{(0)}_{i,j} + A^{(2)}_{i,j} + (1 - \delta_{i,j}) A^{(1)}_{i,j} \Big], \quad (5)$$

where $\delta_{i,j}$ is the Kronecker delta function.

To establish the stability conditions, it is necessary to find a solution of the following system

$$\begin{cases} \alpha A &= 0 \\ \alpha \mathbf{1} &= 1. \end{cases} \quad (6)$$

where the stochastic vector α may be interpreted as the distribution of the phase $Y(t)$ at high levels $X(t)$. Given that, the stability follows from the Neuts ergodicity condition, see [16]: the QBD process $\{X(t), Y(t)\}, t \geqslant 0$ is positive recurrent iff

$$\rho := \alpha A^{(0)} \mathbf{1} / \alpha A^{(2)} \mathbf{1} < 1. \quad (7)$$

Provided (7) holds, there exists a vector of limiting probabilities $\pi = (\pi_0, \pi_1, \dots)$, which consists of the vectors π_k of equilibrium probabilities of level $k \geqslant 0$ (i.e. $\pi_k = (\pi_{k,1}, \dots, \pi_{k,m}), k \geqslant 1$), and is the unique solution of the following system

$$\begin{cases} \pi Q &= 0 \\ \pi \mathbf{1} &= 1. \end{cases} \quad (8)$$

It is shown in [24], that the vector π can be level-wise computed by the celebrated matrix-geometric solution

$$\pi_k = \pi_{k-1} R, \quad k \geqslant 1, \quad (9)$$

where the square matrix R of order m is the minimal nonnegative solution of the system (1), which exists under the stability assumptions. The initial vectors π_0, π_1 are obtained by the following linear system of equations

$$(\pi_0, \pi_1) \begin{pmatrix} A^{0,0} & A^{0,1} \\ A^{1,0} & A^{1,1} + RA^{(2)} \end{pmatrix} = 0, \quad (10)$$

$$\pi_0 \mathbf{1} + \pi_1 (I - R)^{-1} \mathbf{1} = 1. \quad (11)$$

Thus, obtaining the matrix R is crucial for performance analysis of the stable system.

Intuitively, the probability $\pi_k \mathbf{1}$ of the system persistence at a particular level $k \geqslant 0$ is related to the speed of the geometrical decrease in (9). Indeed, the Spectral radius $\eta = sp(R)$ (the greatest eigenvalue of R) is the so-called *caudal characteristic* of the QBD. It is true, that since the system is stable, then η is real, positive, and $\eta < 1$ (which is related to the Perron–Frobenius theorem, see [18]). Thus, the probabilistic behavior of the QBD is defined by the values ρ, and η, for details see [18].

Note that in general the equation (1) has to be solved numerically. However, under some restrictions on the matrices $A^{(0)}$ and $A^{(2)}$, there is a possibility to obtain the explicit solution. The following theorems provide an algebraic formula for the matrix R of a discrete-time Markov chain (an extension to a continuous-time process is straightforward).

Theorem 1 *[18] Assume that $A^{(2)} = cr$, where c is a column vector, and r is a row vector s.t. $r\mathbf{1} = 1$. Then $G = \mathbf{1}r$.*

Note that the matrix G is a minimal nonnegative solution of a matrix equation $G = A_2 + A_1 G + A_0 G^2$, and the matrix R can be easily obtained by the well-known duality $R = A_0(I - A_1 - A_0 G)^{-1}$ [19].

Theorem 2 *[18] Assume that $A^{(0)} = cr$, where c is a column vector, and r is a row vector s.t. $r\mathbf{1} = 1$. Then $R = c\xi$, where $\xi = r(I - A^{(1)} - \eta A^{(2)})^{-1}$ and $\eta = \xi c$, where $\eta = \mathrm{sp}(R)$.*

Note that in the latter theorem the value η has to be computed in advance, see [18].

However, the restriction of the Theorems 1, 2 is the requirement of matrices $A^{(0)}$ or $A^{(2)}$ to be of rank one. A general exact solution of the matrix quadratic equation (1) can be obtained by obtaining the Jordano canonical form. This approach is based on the following lemma.

Lemma 1 *The polynomial matrix $A(\xi) := A^{(0)} + \xi A^{(1)} + \xi^2 A^{(2)}$ allows the following factorization:*

$$A(\xi) \equiv (\xi I - R)(\xi A^{(2)} + R A^{(2)} + A^{(1)}). \qquad (12)$$

The proof follows by direct expansion of the r.h.s. of (12) and substitution of (1). This result reported in [11] was first applied to QBD processes in [26].

Remark 1 *It follows from Lemma 1, that $\det A(\xi) = \det(\xi I - R)\det(\xi A^{(2)} + R A^{(2)} + A^{(1)})$. Thus, the eigenvalues of R are the zeroes of the determinantal polynomial $\det A(\xi)$. Note that there are exactly m zeroes of the determinantal polynomial $\det A(\xi)$ which are strictly inside unit circle and these are the eigenvalues of R (provided $\eta < 1$), see e.g. [26]. Hence, the remaining $r \leqslant m$ zeroes are the zeroes of $\det(\xi A^{(2)} + R A^{(2)} + A^{(1)})$, and are outside the unit circle (a more detailed discussion of the number and location of eigenvalues of R for the case $\eta \leqslant 1$ can be found in [23]).*

The following lemma proved in [26] enables determination of left eigenvectors of rate matrix, R.

Lemma 2 *u is a left eigenvector of rate matrix R corresponding to eigenvalue ξ if and only if*

$$u A(\xi) = \mathbf{0}. \qquad (13)$$

Thus, using Lemma 1 and Lemma 2, if rate matrix R is diagonalizable, it can be obtained by the spectral representation

$$R = T D T^{-1}, \qquad (14)$$

where D is the diagonal matrix of m eigenvalues inside the unit disk, and columns of T consist of right eigenvectors of rate matrix. A more general discussion of Jordan canonical form method of obtaining the rate matrix R is provided in [27]. It should be noted that the suggested approach is a closed form method, in contrast to widely used iterative numerical procedures.

EXACT ALGEBRAIC FORMULA FOR R OF THE QBD PROCESS WITH TWO PHASES

Let $m = 2$, that is, the phase state of the QBD process has exactly two states at each level. Observe that the determinantal polynomial $\det A(\xi)$ is of degree four. Note also, that $\xi = 1$ is always the root of $\det A(\xi)$ by the balance condition (4). Rewrite

$$\det A(\xi)/(\xi - 1) = a_3\xi^3 + a_2\xi^2 + a_1\xi + a_0. \qquad (15)$$

Denote by $\xi_i, i = 1, 2, 3$ the zeroes of the polynomial (15). By the Remark 1, two of these zeroes are inside the unit disk, denote them ξ_1, ξ_2. Consider the characteristic polynomial of the matrix R in a monic form

$$\det(\xi I - R) = (\xi - \xi_1)(\xi - \xi_2) = \xi^2 + b_1\xi + b_0, \qquad (16)$$

where $b_0 = \xi_1\xi_2 = \det R$, $b_1 = -(\xi_1 + \xi_2) = -\mathrm{Trace}(R)$ are (real) scalars, since R is a nonnegative matrix. By Perron–Frobenius theorem, $\eta = \mathrm{sp}(R)$ is simple, real eigenvalue, and $\eta \in (0, 1)$, provided (7) holds. W.o.l.o.g. let $\xi_1 = \eta$. Then ξ_2 is also real, which makes ξ_3 also real. Given that, it is easy to obtain

$$b_1 = \frac{a_2}{a_3} + \xi_3, \qquad (17)$$

$$b_0 = -\frac{a_0}{a_3\xi_3}. \qquad (18)$$

Hence, it is necessary to obtain ξ_3. Note that it can be done by the celebrated trigonometric solution of the cubic equation (15) by the following substitutions:

$$p = \frac{3a_3a_1 - a_2^2}{9a_3^2}, \quad q = \frac{2a_2^3 - 9a_3a_2a_1 + 27a_3^2a_4}{27a_3^3}. \qquad (19)$$

Then [35]

$$\xi_3 = -\frac{a_2}{3a_3} + 2\sqrt{-p}\cos\left(\frac{1}{3}\cos^{-1}\left(\frac{q}{2p\sqrt{-p}}\right)\right). \qquad (20)$$

By Cayley-Hamilton theorem, we obtain

$$R^2 = -b_1 R - b_0 I, \qquad (21)$$

which, by substitution into (1), leads to the following system of linear equations:

$$R\left[A^{(1)} - b_1 A^{(2)}\right] - b_0 A^{(2)} + A^{(0)} = \mathbf{0}. \qquad (22)$$

Thus, if $A^{(1)} - b_1 A^{(2)}$ is invertible, then

$$R = \left[b_0 A^{(2)} - A^{(0)}\right] \left[A^{(1)} - b_1 A^{(2)}\right]^{-1}. \quad (23)$$

Lemma 3 *The matrix $A^{(1)} - b_1 A^{(2)}$ is invertible.*

Proof: Since the eigenvalues are distinct, then R is diagonalizable, hence, R has a spectral representation (14), which is equivalent to

$$R = \eta E_1 + \xi_2 E_2, \quad (24)$$

where E_1, E_2 are the so-called residue matrices, with $E_1 + E_2 = I$. Note, that by Perron–Frobenius theorem, $E_1 \geqslant \mathbf{0}$ componentwise [18], which provides a componentwise inequality

$$R \geqslant \xi_2 E_1 + \xi_2 E_2 = \xi_2 I. \quad (25)$$

Recall, that from (4), the equality $A(1)\mathbf{1} = \mathbf{0}$ holds. Considering the expansion (12) at $\xi = 1$ provides $(I - R)(A^{(2)} + RA^{(2)} + A^{(1)})\mathbf{1} = \mathbf{0}$. However, since by (16) $\det(I - R) = b_0 \neq 0$, then the matrix $(I - R)$ is nonsingular, hence, $(A^{(2)} + RA^{(2)} + A^{(1)})\mathbf{1} = \mathbf{0}$, which provides (considering the non-trivial case $A^{(2)} \neq \mathbf{0}$)

$$|A^{(1)}_{i,i}| > \sum_{j=1}^{m} \left[\xi A^{(2)}_{i,j} + (RA^{(2)})_{i,j} + (1 - \delta_{i,j}) A^{(1)}_{i,j}\right], \quad (26)$$

for any real $\xi \in (-1, 1)$. Now taking $\xi = \eta$ in (26) and noting, that (25) provides $RA_2 \geqslant \xi_2 A_2$ componentwise, yields that $[A^{(1)} - b_1 A^{(2)}]$ is a strictly diagonally dominant matrix and hence is nonsingular. ■

Thus, the suggested procedure for exact computation of the rate matrix R is as follows.

1. Obtain the maximal eigenvalue ξ_3 by (20).
2. Obtain b_1 by (17) and b_0 by (18).
3. Obtain R by (23).

Alternatively, ξ_1, ξ_2 can be computed using Lemma 1. Further, left eigenvectors can be computed by Lemma 2. Finally, R is obtained by (14). Note that this approach can be generalized to an arbitrary $G/M/1$-type Markov process.

Now we consider the case $\xi_2 = 0$, i.e. the rate matrix R is singular. Thus, following the suggested procedure, we obtain $b_0 = 0$ and $b_1 = -\eta$. Hence $R = [-A^{(0)}][A^{(1)} + \eta A^{(2)}]^{-1}$, which corresponds to the known result presented by Theorem 2.

Remark 2 *Note that algebraically the matrices R and $A^{(0)}$ have the same rank. Since $\det A(\xi)$ is the polynomial of power $2m$, and eigenvalues of R are zeroes of the aforementioned polynomial, in the following cases (provided $\eta < 1$) the zeroes can be explicitly computed, and thus the rate matrix R can be theoretically computed in closed form:*
i) $m \leqslant 4$, if $A^{(0)}$ is rank-one matrix,
ii) $m \leqslant 3$ if $A^{(0)}$ is of rank two.
We continue with $m = 2$ and a full rank $A^{(0)}$ below.

RANDOM SWITCHING FOR POWER SAVING

Now we turn to a practical application of the result. Heavy restrictions on energy consumption of a battery-powered IoT device result in typical application of asynchronous information transmission/receive modes with various data rates, and equipment of the device with lightweight, low consuming logic [2]. These aspects lead to the following queueing model of a single IoT service device. Consider a queueing system with a renewal input flow of customers arriving into an (unbounded) First-Come-First-Served queue. The i.i.d. interarrival times are exponentially distributed with rate $\lambda > 0$. Each customer requires an exponentially distributed (with unit rate) amount of work to be done (say, information to be transmitted). The single server operates two speed modes (call them high and low), with rates $\mu_2 > \mu_1 > 0$. The server may switch the speed only at the arrival/departure epochs (asynchronously). Denote $c_0 < c_1 < c_2$ the energy consumption per unit time in idle (no customers in the system)/low/high modes. In order to preserve energy, the server implements the following random switching policy:

- at the task arrival epoch, given the current mode is low, switch to high mode with probability (w.p.) p_1, or remain low w.p. $1 - p_1$;
- at the task departure epoch, given the current mode is high, switch to low mode w.p. p_2, or remain high w.p. $1 - p_2$.

Let $\nu(t) \in \{0, 1 \dots\}$ be the number of customers, and $m(t) \in \{1, 2\}$ be the mode of the system at time $t \geqslant 0$. Then the following Markov process

$$\{(\nu(t), m(t)) \in \{0, 1, \dots\} \times \{1, 2\},\ t \geqslant 0\} \quad (27)$$

is a continuous-time QBD process, with $\nu(t)$ being the *level*, and $m(t)$ being the *phase* at time t.

The infinitesimal generator matrix of the process (27) has the form (3), where we define the matrices explicitly:

$$A^{(0)} = \begin{pmatrix} (1 - p_1)\lambda & p_1 \lambda \\ 0 & \lambda \end{pmatrix}, \quad (28)$$

$$A^{(1)} = \begin{pmatrix} -\lambda - \mu_1 & 0 \\ 0 & -\lambda - \mu_2 \end{pmatrix}, \quad (29)$$

$$A^{(2)} = \begin{pmatrix} \mu_1 & 0 \\ p_2 \mu_2 & (1 - p_2)\mu_2 \end{pmatrix}, \quad (30)$$

$$A^{0,0} = -\lambda I, \quad A^{0,1} = A^{(0)}, \quad (31)$$

$$A^{1,1} = A^{(1)}, \quad A^{1,0} = A^{(2)}. \quad (32)$$

To establish the stability criterion of the process (27), we construct the matrix

$$A := A^{(0)} + A^{(1)} + A^{(2)} = \begin{pmatrix} -p_1 \lambda & p_1 \lambda \\ p_2 \mu_2 & -p_2 \mu_2 \end{pmatrix}. \quad (33)$$

Solving the system (6) and using (28)-(30), the condition (7) provides

$$\alpha_1 \mu_1 + \alpha_2 \mu_2 > \lambda,$$

where

$$\alpha_1 = \frac{p_1\lambda}{p_1\lambda + p_2\mu_2},\ \alpha_2 = \frac{p_2\mu_2}{p_1\lambda + p_2\mu_2},$$

which is equivalent to

$$\lambda p_1(\lambda - \mu_2) + \mu_2 p_2(\lambda - \mu_1) < 0. \quad (34)$$

Intuitively, the condition (34) indicates a negative drift of the service process of the system under heavy load, with respect to the mode switching intensity.

Note that $p_1 = 1$ and $p_2 = 0$ corresponds to a classical $M/M/1$ service system working at the speed μ_2 (referred below as classical system), where the stability condition (34) reduces to the celebrated $\rho := \lambda/\mu_2 < 1$. Now let $\mathrm{E}\nu_0$ be the average number of customers, and $\mathrm{E}\mathcal{E}_0$ be the average energy consumption per unit time in the classical system in stationary regime. It can be readily seen, that

$$\mathrm{E}\nu_0 = \frac{\rho}{1-\rho}, \quad (35)$$

$$\mathrm{E}\mathcal{E}_0 = c_0(1-\rho) + c_2\rho, \quad (36)$$

where, recall, $1-\rho$ is the stationary idle probability of the classical system (for details on the aforementioned classical results see [3]). We may consider $\mathrm{E}\nu_0$ as the QoS parameter of the classical system.

Now we turn to the original two-mode system, i.e. we consider the non-trivial case $p_1, p_2 > 0$. Provided (34) holds, we define the matrix R following the steps of the suggested procedure of exact computation, where the coefficients in the polynomial (15) are obtained as follows:

$$\begin{aligned} a_3 &= \mu_1\mu_2(1-p_2),\\ a_2 &= -\mu_1(\lambda+\mu_2) - \lambda\mu_2(1-p_2),\\ a_1 &= \lambda(\lambda+\mu_1) + \lambda\mu_2(1-p_1),\\ a_0 &= -\lambda^2(1-p_1). \end{aligned}$$

After obtaining R from the equation (23), the equations (9)–(11) provide the stationary system state probabilities. Straightforward manipulation leads to the following system for π_1:

$$\begin{cases} \pi_1\left(\frac{1}{\lambda}A^{(2)} - R^{-1}\right)A^{(0)}\mathbf{1} &= 0,\\ \pi_1\left(\frac{1}{\lambda}A^{(2)} + (I-R)^{-1}\right)\mathbf{1} &= 1. \end{cases} \quad (37)$$

Then the value π_0 is obtained as follows:

$$\pi_0 = \frac{1}{\lambda}\pi_1 A^{(2)}. \quad (38)$$

The obtained solution allows to evaluate the average number of customers in the system as the QoS measure:

$$\mathrm{E}\nu_{p_1,p_2} = \pi_1(I-R)^{-2}\mathbf{1}, \quad (39)$$

where by notation ν_{p_1,p_2} we stress the dependence on the mode switching probabilities. The average energy consumption may be obtained as follows:

$$\mathrm{E}\mathcal{E}_{p_1,p_2} = \pi_0 c_0\mathbf{1} + \pi_1(I-R)^{-1}(c_1, c_2)^T, \quad (40)$$

where the transposed vector (c_1, c_2) is the column-vector of energy consumption in each mode. Then the following optimization problem can be solved: minimization of the average energy consumption in stationary regime, provided the controlled QoS decrease:

$$\begin{cases} \mathrm{E}\mathcal{E}_{p_1,p_2} & \to_{p_1,p_2>0} \min,\\ \mathrm{E}\nu_{p_1,p_2} & \leqslant (1+\varepsilon)\mathrm{E}\nu_0, \end{cases} \quad (41)$$

for some $\varepsilon > 0$.

Now we present the numerical investigation of the optimization problem (41), using the energy and speed values from a real-world example. We describe the ATmega328/P controller [4], widely used in applications. This low power consuming micro-controller is capable of dynamical voltage/frequency setup, which allows to significally reduce the power consumption. We use the values of power consumption for given frequency/voltage configuration extracted from [4] as follows: 0.6 mW for 1 MHz/2 V active regime, 26 mW for 8 MHz/5 V active regime, and 0.08 mW for 2 V idle regime. We assume, that $\rho < 1$, that is, the system working at 8 MHz is stable.

Note, however, that the high frequency regime is relatively expensive. Now we present two scenarios: (i) the system is stable at the low regime (with input rate λ_1); (ii): the system is unstable at the low regime (with input rate λ_2). Thus, we have the following settings:

$$\begin{aligned} &\mu_1 = 1,\ \mu_2 = 8,\\ &c_0 = 0.08,\ c_1 = 0.6,\ c_2 = 26,\\ &\lambda_1 = 0.5,\ \lambda_2 = 2.5. \end{aligned}$$

Following the procedure of obtaining the exact solution, we numerically solve the optimization problem (41) and obtain approximate optimal values p_1, p_2, as well as approximate optimal energy consumption. We vary the QoS degradation $1+\varepsilon \in (1.1, 30)$ and plot the obtained values. We perform simulation with R package [25].

It is easy to see from Fig. 1, that, as expected, if the low regime is stable, the system tends to an M/M/1 system working at low speed, with the appropriate growth of performance degradation.

On the other hand, if the system at low frequency is unstable (see Fig. 2), the system oscillates around the stability border, and the consumption tends to some average value w.r.t. the switching probabilities.

CONCLUSIONS AND FUTURE WORK

We have presented the algebraic approach for obtaining the exact value of rate matrix R in case the phase state of the QBD process has only two phases. Note, that, albeit being simplistic, such a QBD process may be used to model quite a number of recent applications, such as the IoT devices, high-performance and cloud-based servers (where processors use the Dynamic Voltage and Frequency Scaling technology), telecommunications (with two types of service, such as the Invite and non-Invite messages in SIP protocol [12]). For such types of applications, it is crucial to obtain the cost effectiveness

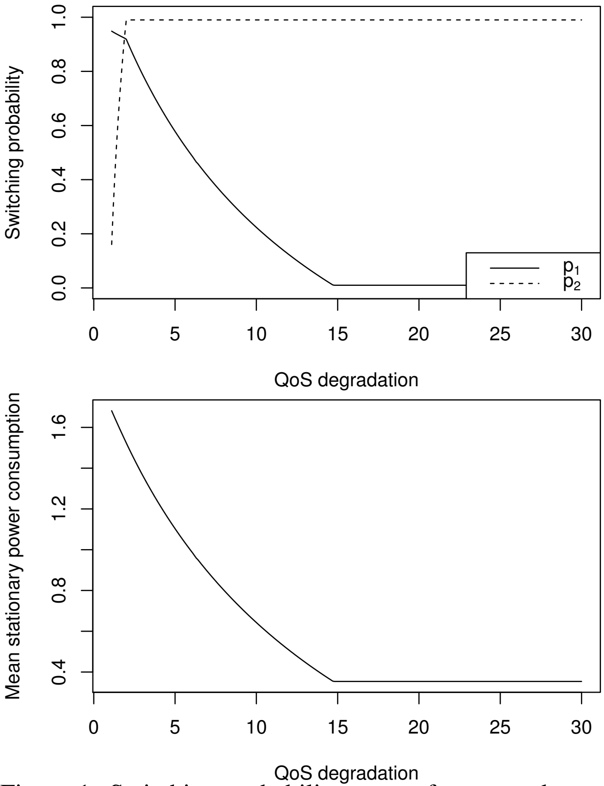

Figure 1: Switching probability vs. performance degradation (upper); average energy consumption per unit time vs. performance degradation in a system being stable at low level (input rate $\lambda_1 = 0.5, \mu_1 = 1$).

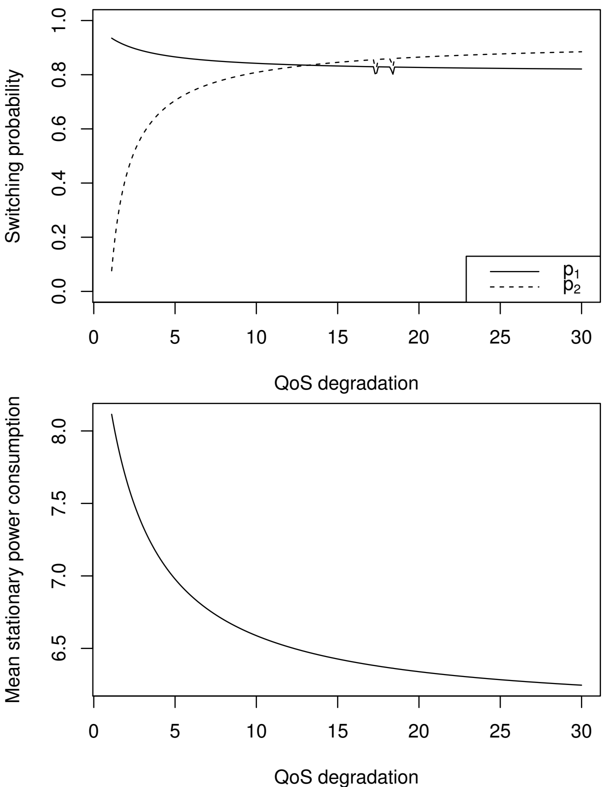

Figure 2: Switching probability vs. performance degradation (upper); average energy consumption per unit time vs. performance degradation in a system being unstable at low level (input rate $\lambda_2 = 2.5, \mu_1 = 1$).

(such as energy efficiency in the IoT) without implementation of a sophisticated control due to a significant overhead induced by such a control. We proposed a randomized scheme which may be implemented in such systems to obtain the desired cost effectiveness under controlled QoS degradation, when possible. We leave the field test of the proposed scheme for future research.

We note, that most of the results of this research paper can be generalized to arbitrary $G/M/1$-type Markov process (with two states at each level). In this case, the rate matrix R is a solution of the matrix power series equation (2). However, a detailed study is a topic of a separate discussion [28].

ACKNOWLEDGEMENTS

The authors thank Prof. Evsey Morozov for his helpful comments, that significantly improved the paper. The research of AR is partially supported by RF President's Grant No. MK-1641.2017.1, and by RFBR, grants No.15-07-02341, 15-07-02354, 16-07-00622, 15-29-07974.

REFERENCES

[1] Abbas, Z., Yoon, W.: A Survey on Energy Conserving Mechanisms for the Internet of Things: Wireless Networking Aspects. Sensors. 15, 24818–24847 (2015)

[2] Augustin, A., Yi, J., Clausen, T., Townsley, W.M.: A study of LoRa: Long range & low power networks for the internet of things. Sensors. 16, 1466 (2016)

[3] Asmussen, S. Applied Probability and Queues. 2003.

[4] ATmega328/P Datasheet Complete Rev.: Atmel-42735B-ATmega328/P_Datasheet_Complete-11/2016

[5] Bagula, A., Abidoye, A.P., Zodi, G.-A.L.: Service-Aware Clustering: An Energy-Efficient Model for the Internet-of-Things. Sensors. 16, 9–0 (2016)

[6] Bini, D.A., Latouche, G., Meini, B.: Solving matrix polynomial equations arising in queueing problems. Linear Algebra and its Applications. 340, 225–244 (2002)

[7] Chelloug, S.A.: Energy-Efficient Content-Based Routing in Internet of Things. Journal of Computer and Communications. 03, 9–20 (2015)

[8] Daigle, J. N., Lucantoni, D. M.: Queueing systems having phase-dependent arrival and service rates. Numerical Solution of Markov Chains, 161–202. Marcel Dekker Inc., New York (1991)

[9] Do, T.V., Chakka, R.: An efficient method to compute the rate matrix for retrial queues with large number of servers. Applied Mathematics Letters. 23, 638—643 (2010)

[10] Evans, R. V.: Geometric Distribution in some Two-Dimensional Queueing Systems, Operations Research. 15, 830–846 (1967)

[11] Gantmacher, Felix (1959), Theory of matrices, AMS Chelsea publishing

[12] Gaidamaka, Y.V.: Model with threshold control for analyzing a server with an SIP protocol in the overload mode. Automatic Control and Computer Sciences. 47, 211–218 (2013)

[13] Gail, H.R., Hantler, S.L., Taylor, B.A.: Spectral Analysis of M/G/1 and G/M/1 Type Markov Chains. Advances in Applied Probability. 28, 114 (1996)

[14] Gillent, F. and Latouche, G.: Semi-explicit solutions for M/PH/1-like queuing systems. European Journal of Operational Research. 13(2), 151–160, (1983)

[15] Gohberg, I., Lancaster, P., Rodman, L.: Matrix polynomials. Society for Industrial and Applied Mathematics, Philadelphia (2009).

[16] He, Q.-M.: Fundamentals of Matrix-Analytic Methods. Springer, New York (2014)

[17] Hung, T.T., Do, T.V.: Computational aspects for steady state analysis of QBD processes. Periodica Polytech. Ser. Electr. Eng. 44, 179–200 (2001)

[18] Latouche, G. and Ramaswami, V.: Introduction to Matrix Analytic Methods in Stochastic Modeling. ASA–SIAM, Philadelphia (1999)

[19] Latouche, G.: A note on two matrices occurring in the solution of quasi-birth-and-death processes. Stochastic Models. 3, 251–257 (1987)

[20] Low-Power Long Range LoRa Technology Transceiver Module. Datasheet No. DS50002346B, 2015.

[21] Mitrani, I., Chakka, R.: Spectral expansion solution for a class of Markov models: Application and comparison with the matrix-geometric method. Performance Evaluation. 23, 241–260 (1995)

[22] Mukherjee, D., Dhara, S., Borst, S., van Leeuwaarden, J.S.H.: Optimal Service Elasticity in Large-Scale Distributed Systems. ArXiv e-prints. 1703.08373, (2017)

[23] Naoumov, V., Samouylov, K. and Gaidamaka, Yu.: Multiplicative solutions of finite Markov chains. RUDN, Moscow (2015) (in Russian)

[24] Neuts, M.F. Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach. The Johns Hopkins University Press, Baltimore (1981).

[25] R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. `http://www.r-project.org/`

[26] Rama Murthy, G.: Transient and equilibrium analysis of computer networks: Finite memory and matrix geometric recursions, PhD. Thesis, Purdue University, West Lafayette (1989)

[27] Rama Murthy, G., Kim, M., Coyle, E. J.: Equilibrium analysis of skip-free Markov chains: Non-linear Matrix Equations. Communications in Statistics – Stochastic Models. 4, 547–571 (1991)

[28] Rama Murthy, G. and Rumyantsev, A.: On an exact solution of the rate matrix of G/M/1-type Markov process with small number of phases. Manuscript in preparation

[29] Rani, S., Talwar, R., Malhotra, J., Ahmed, S., Sarkar, M., Song, H.: A Novel Scheme for an Energy Efficient Internet of Things Based on Wireless Sensor Networks. Sensors. 15, 28603–28626 (2015)

[30] Sanchez-Iborra, R., Cano, M.-D.: State of the Art in LP-WAN Solutions for Industrial IoT Services. Sensors. 16, 708 (2016)

[31] Sarwesh, P., Shet, N.S.V., Chandrasekaran, K.: Energy Efficient Network Design for IoT Healthcare Applications. In: Bhatt, C., Dey, N., and Ashour, A.S. (eds.) Internet of Things and Big Data Technologies for Next Generation Healthcare. pp. 35–61. Springer International Publishing, Cham (2017)

[32] Sarwesh P, N. Shekar V. Shet, Chandrasekaran K: Energy efficient network architecture for IoT applications. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT). pp. 784–789 (2015)

[33] Vellanki, M., Kandukuri, S.P.R., Razaque, A.: Node Level Energy Efficiency Protocol for Internet of Things. Journal of Theoretical and Computational Science. 3, (2016)

[34] Wallace, V. L.: The Solution of Quasi Birth and Death Processes Arising from Multiple Access Computer Systems, PhD Thesis, University of Michigan (1969)

[35] Zwillinger, D. CRC Standard Mathematical Tables and Formulae, 31st Edition. CRC, Boca Raton, 2003, 910 pages.

AUTHOR BIOGRAPHIES

RAMA MURTHY GARIMELLA is an Associate Professor at the International Institute of Information Technology, Hyderabad, India. He is a member of Eta Kappa Nu, Phi Kappa Phi (Honor Societies in USA), IEEE Computer Society, Computer Society of India, Senior Member of Association for Computing Machinery (ACM), and Fellow of IETE (Institution of Electronics and Telecommunication Engineers). His research interests include Wireless Sensor Networks, Adhoc Wireless Networks, Multi-Dimensional Neural Networks, Performance Evaluation.

RUMYANTSEV ALEXANDER received his PhD. from Petrozavodsk State University. He is now a researcher in the Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences. His research interests include Stochastic Processes, Queueing Systems, High-Performance and Distributed Computing, Multi-Core and Many-Core Systems.