An efficient algorithm for modelling and dynamic prediction of network traffic

Wenjie Fan and Hong Zhang

College of Information Science and Engineering,
Chengdu University,
Chengdu, 610106, China
and
Key Laboratory of Pattern Recognition and
Intelligent Information Processing,
Chengdu, 610106, China
Email: fanwenjie@cdu.edu.cn
Email: 945389781@qq.com

Kuan-Ching Li*

Department of Computer Science and Information Engineering (CSIE),
Providence University, Taiwan
and
Hubei University of Education, China

Email: kuancli@pu.edu.tw
*Corresponding author

Shunxiang Zhang

Anhui University of Science and Technology, China

Email: sxzhang@aust.edu.cn

Mario Donato Marino

School of Computing, Creative Technologies and Engineering,
Leeds Beckett University, UK
Email: M.D.Marino@leedsbeckett.ac.uk

Hai Jiang

Department of Computer Science,
Arkansas State University, USA
Email: hjiang@astate.edu

Comment [T1]: Author: Please confirm if the author's address is correct.

Comment [T2]: Author: Please provide full institutional address

Comment [T3]: Author: Please provide full institutional address

Abstract: Network node degradation is an important problem in internet of things given the ubiquitous high number of personal computers, tablets, phones and other equipments present nowadays. In order to verify the network traffic degradation as one or multiple nodes in a network fail, this paper proposes one algorithm based on product form results (PRF) for fractionally auto regressive integrated moving average (FARIMA) model, namely PFRF. In this algorithm, the prediction method is established by FARIMA model, through equations for queuing situation and average queue length in steady state derived from queuing theory. Experimental simulations were conducted to investigate the relationships between average queue length and service rate. Results demonstrated that, not only it has good adaptability, but has also achieved promising magnitude of 9.87 as standard deviation which shows its high prediction accuracy, given the low-magnitude difference between original value and the algorithm.

Keywords: prediction; product form results; PRF; FARIMA model; average length of queue.

Reference to this paper should be made as follows: Fan, W., Zhang, H., Li, K-C., Zhang, S., Marino, M.D. and Jiang, H. (xxxx) 'An efficient algorithm for modelling and dynamic prediction of network traffic', *Int. J. Computational Science and Engineering*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Wenjie Fan received his MS from the Sichuan Normal University, Chengdu, China. Presently, he is currently a Lecturer in the Chengdu University. His current research interests include networks, big data and real-time computing.

Hong Zhang received his MS in Computer Science and Technology from the Southwest Jiaotong University, Chengdu, China. Presently, he is currently an Associate Professor in the Chengdu University. His current research interests include big data, computer network and optimisation theory.

Kuan-Ching Li is a Professor in the Department of Computer Science and Information Engineering at the Providence University, Taiwan. He is a recipient of awards from Ministry of Science and Technology (MOST)/Taiwan and Nvidia, as also recipient of numerous chair, distinguished and chair professorships from universities in several countries. He is the Editor of several technical professional books published by CRC Press, IGI Global, McGraw-Hill and Springer. His topics of interest include networked computing, GPU computing, parallel software design, and performance evaluation and benchmarking. He is a senior member of the IEEE and a Fellow of the IET.

Shunxiang Zhang received his PhD in Computer Application Technology from the Shanghai University, China. Presently, he is the Department Chair and an Associate Professor at the Anhui University of Science and Technology. His main research interests include web mining, semantic search, and internet of things.

Mario Donato Marino is currently a Senior Lecturer in the Leeds Beckett University. He received his BE in Electrical Engineering with emphasis in Computer Engineering at the University of Sao Paulo in 1993, while MSc in Computer Engineering in 1996 and PhD in Computer Engineering, both from the same university. He has worked at the University of Sao Paulo as an Assistant Professor and University of Texas at Austin and University of Virginia as a researcher. He has received a best paper award on an international top conference and has co-authored several international articles in conferences, journals, and workshops which include computer architecture, microprocessor evaluation, systems, high-performance computing, distributed computing, parallel computing, and performance evaluation. He serves as an Associate Editor on the *International Journal of Embedded Systems* (*IJES* – Inderscience) and /has served a number of committees in conferences, workshops, and to the editorial board of the *International Journal of High Performance Computing and Networking* (*IJHPCN* – Inderscience). In 2005, he received an award on teaching while working at University of Sao Paulo. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and a member of the Association of Computer Machinery (ACM).

Hai Jiang is a Professor in the Department of Computer Science at Arkansas State University, USA. He received his BS from the Beijing University of Posts and Telecommunications, China, and MA and PhD from the Wayne State University, Detroit, MI, USA. He is a professional member of ACM and IEEE Computer Society. He has published one book and more than 70 papers in refereed journals and conference proceedings as well as several book chapters. He has been involved in 60 conferences and workshops as a Program/Workshop/Session Chair or as a program committee member. His current research interests include parallel and distributed systems, computer and network security, high performance computing and communication, and modelling and simulation.

This paper is a revised and expanded version of a paper entitled [title] presented at [name, location and date of conference].

Comment [T4]: Author: Please reduce biographical details of M.D. Marino to no more than 100 words.

Comment [T5]: Author: If a previous version of your paper has originally been presented at a conference please complete the statement to this effect or delete if not applicable.

1 Introduction

With the rapid increase on the aggregated bandwidth and requirements to the scale of interconnection, the effectiveness of nodes in networks is of huge importance on networks' performance, and thus, higher precision analysis is vital to the description of network flow, further analysis and design.

Given the importance of this problem, there are a number of researches related to the network interconnect and flow (You et al., 2016; Asakura, 2007; Divakaran and Chinnagounder, 2015; Su et al., 2016; Nakayama and Gang, 2013; Chuang and Jiang, 2006; Ye et al., 2015; Abraham and Rao, 2008; Todinov, 2012). For example, traffic prediction models are available, such as autoregressive integrated moving average model (ARIMA), as well a number of traffic prediction models available, such as fractionally auto regressive integrated moving average (FARIMA), artificial neural networks (ANN) and wavelet-based, with good performance and accuracy.

Recent investigations (Sahinoglu and Libby, 2005; Liu and Liu, 2002; Wei et al., 2013; Zhang et al., 2015) based on evaluation measurements have put forward the reliability of communication network. In Sahinoglu and Libby (2005), the reliability of the network was evaluated using probability expression, while in Liu and Liu (2002), the flow factors and connectivity of the network were considered. The prediction algorithm-based FARIMA model for breakdown (PFB) was put forward to depict the network traffic in Wei et al. (2013), whilst the prediction algorithm based on discrete-queue for FARIMA model (PDF) algorithm was proposed to accurately predict the network traffic in Zhang et al. (2015).

Modelling using queuing theory can support to build models of network traffic yet analyse average response of a parameterised network and their performance as well node effectiveness. The retrial queuing steady captain for Geo B/G/1 was proposed in Kim (2009), where the characteristic of the system is that, if the server is busy, customers will enter into the phase of retry until the completion of the service. The discrete GI/Geo multiple vacation/1 queue was studied in Zhao and Deng (2010).

Targeting the fundamental bandwidth constraints ubiquitously present due to the growth networks – such as IoT – as on abovementioned past investigations, this paper aims at the following contributions:

- Product form results (PRF) algorithm is proposed to approach the determination of average data packet and network bandwidth requirements.
- Takes into consideration the knowledge of the queuing theory to compute the node media access control (MAC) layer packet flow quantitatively, based on related discrete time mathematical methods.
- It is derived a mathematical model that analyses the steady average length of queue (Nolan and Panorska, 2001).
- The model aims to predict the network traffic behaviour by using a discrete time algorithm, to the best of our knowledge for the first time.
- The analysis allows the determination of the length of MAC layer in order to improve the performance of a network by using discrete time structure algorithm for network traffic prediction.
- As further demonstrated, compared to previous algorithms the proposed one is efficient because the PFRF algorithm fully considers the relationship between the average queue length of the nodes, the probability of the packets arriving at the nodes, and the steady state of the nodes.

Experimental results of the proposed algorithm are validated using simulation software, and show that the flow prediction algorithm based on the optimal product form solution can also be used to calculate the length of MAC layer, key information to improve the performance of the network by using the discrete time structure algorithm for network traffic prediction.

The remaining of this paper is organised as follows. Section 2 describes the background, important formulation and how the average queue length of data packets is determined. Section 3 describes the related work. PRF algorithm is described in Section 4, Section 5 describes the experiments and results, and finally, conclusions and future work are found in Section 6.

2 Background

Terminologies and definitions used throughout this paper are given in this section. The N discrete time queuing systems are connected to a network, whereas each system is labelled as node j, $j = 1, \ldots, N$. Like in single discrete time queuing system, it is referred in this paper that, the leaving of data packet will only happen at the beginning of time slot (n, n^+) at each node.

In addition, in the beginning of time slot (n, n^+), if one data packet finished its service at node i and then left, the packet will become one packet arrival of node j at that beginning of time slot with probability r_{ij}, $i, j = 1, \ldots, N$.

Furthermore, the arrival of data packet at each node from external network is represented by an event that happens at the end of time slot (n^-, n) when this event is consolidated. That is, each node in the network is a discrete time queuing system which is late and delayed entrance for it to enter. In this situation, the probability r_{ij}, $i = 1, \ldots, N$; $j = 0, 1, \ldots, n$ is called the transition probability of data packet between different nodes. These are the rules that form one routing of data packet in the network and satisfy the following condition:

$$\sum_{j=0}^{N} r_{ij} = 1, \quad i = 1, \ldots, N \tag{1}$$

This paper will only focus on the stable behaviour of network due to high complexity of queuing network modelling. In the following steps, we define some terms.

L_j represents the flow of data packet at any discrete time node j under stable situation, and the flow includes data that both are queuing and accepting services at node j, $j = 1, \ldots, N$. Analogously, L_j^+ indicates the flow of data packet at time n^+ in the node j on a steady state.

Random vectors $L_1, \ldots, L_N$ define the length of queue in a steady state of network at slot points. The random vectors $L_1^+, \ldots, L_N^+$ are called the length of queue at inner time slot to network.

If, at the beginning of time slot n, n^+, a data packet leaves the node j, then L_j and L_j^+ read and contain different values. L_j includes data packets that will leave the node j at the beginning of time slot n, n^+ not included in L_j^+. If vectors are re-ordered, so calculations are performed more conveniently as follows:

$$\begin{aligned} &L = (L_1, ..., L_N), \\ &L^+ = (L_1^+, ..., L_N^+), \\ &n = (n_1, ..., n_N), \end{aligned} \quad (2)$$

and

$$\begin{aligned} &L = n \Leftrightarrow L_1 = n_1, ..., L_N = n_N \\ &L^+ = \Leftrightarrow L_1^+ = n_1, ..., L_N^+ = n_N \end{aligned} \quad (3)$$

Comment [T6]: Author: Please confirm if the equation is correct.

2.1 Definitions

Next, we present some important definitions based on report (Tian et al., 2008) that are fundamental to understand the proposed algorithm.

Definition 1: At any given time, if time queuing system in the network has the length of queue in a steady state [the distribution of the length of queue in a steady state is showed in formula (4)], then this network owns PRF (Tian et al., 2008).

$$\begin{aligned} &\pi(n) = P\{L = n\} = \prod_{j=1}^{N} P(L_j = n_j) \\ &n^+(n) = P\{L^+ = n\} = \sum_{j=1}^{N} P\{L_j^+ = n_j\} \end{aligned} \quad (4)$$

Additionally, in the time queuing system, L_j and L_J^+ are the lengths of queue in a steady state when the node j is observed solely by the network.

For continuous time queuing system, there may be PRF where both data packet and service are found in Poisson process. Though, for discrete time queuing system, each node will arrive in batches according to a Bernoulli formulation, to make the system owning PRF.

Definition 2: In a discrete time queuing system which has delayed entrance for entering, each node will arrive in Bernoulli way. Moreover, at each end of the time slot (n^-, n), the quantity of node X is reciprocally independent, and all of them accept the same distribution.

$$a_i = P\{X = i\}, i = 0, 1, ... \quad (5)$$

And L_l and L_l^+ represent the number of data packets in the system at the bound of time slot and inner time slot, and A_l represents the number of data packets arrived at (l^-, l), D_l is the number of data packets left (l, l^+) after finished their services. Hence, the result follows:

$$\begin{aligned} &L_{l+1} = L_l - D_l + A_l, l \geq 0; \\ &L_{l+1}^+ = L_{l+1} - D_{l+1} = L_l^+ + A_l - D_{l+1}, l \geq 0 \end{aligned} \quad (6)$$

It is shown that, on the formula (6) D_l relies on $_l$ but is independent of all events before time slot mode l, which can be written as:

$$\begin{aligned} p^D(n, i) &= P\{D_l = i \mid L_k, k = 0, 1, ..., l-1; L_l = n\} \\ &= P\{D_l = i \mid L_l = n\}.i = 0, 1, ..., n \end{aligned} \quad (7)$$

$p^D(n, i)$ is called the leaving function of queuing system. For systems that arrive in Bernoulli way and their leaving functions follow formula (7), the lengths of $\{L_l \geq 0\}$ and $\{L_l^+ \geq 0\}$ are both Markov chains. At the same time, if the system reaches to a balance and L and L^+ represent the stationary length of queue at the bound of time slot and inner time slot, the distribution is noted as:

$$\pi_i = P\{L = i\}, \pi_i^+ = P\{L^+ = i\}, i = 0, 1, ... \quad (8)$$

Definition 3: A discrete time queuing system which arrives in Bernoulli way is reversible, if and only if there is a distribution of probability $\{b_i, i \geq= 0\}$, make:

$$\pi_{n+i} p^D(n+i, i) = b_i \pi_n^+, \quad n, i \geq= 0 \quad (9)$$

Proof: Formula (9) comes into existence when and only when $a_i = b_i$, since:

$$\begin{aligned} \pi_{n+i} p^D(n+i, i) &= P\{L_l = n+i\} P\{D_l = i \mid L_l = n+i\} \\ &= P\{D_l = i, L_l = n+i\} \\ &= P\{D_l = i, L_l = n+i\} \\ &= P\{D_l = i \mid L_l^+ = n\} \pi_n^+ \end{aligned}$$

Comparing this equation to formula (9), the result is:

$$P\{D_l = i \mid L_l^+ = n\} = b_i, \quad i = 0, 1, ... \quad (10)$$

shows that, although D_l relies on L_l, it has no relation with L_l. Also, we represent $\{b_i, i \geq 0\}$ as the distribution of a leaving data packet of time slot in a stable situation. Therefore, comparing to formula (9), the latter comes into existence only when $a_i = b_i$. Thus, for quasi reversible discrete time queuing system, leaving and arrival of data packet will experience the same process in batches at Bernoulli way.

2.2 Important theorem

Next, we present a theorem (Tian et al., 2008) that helps to understand why the proposed algorithm is important.

Theorem 1: Data packets arrive in batches at Bernoulli way and distributed in batches as a_i, $i = 0, 1, \ldots$, the discrete time queuing system is quasi reversible, when and only when $a_i = b_i$, $i = 0, 1, \ldots$, the leaving function is:

$$p^D(n, i) = \frac{\varepsilon(n-i)}{\theta(n)} a_i \quad (11)$$

Here, $\varepsilon(n)$ is a non-negative function and it meets the condition that:

$$\sum_{n=0}^{\infty} \varepsilon(n) < \infty$$

$\theta(n)$ is a positive function and its definition is as follows:

$$\theta(n) = \sum_{i=0}^{n} \varepsilon(n-i)a_i,\ n \ge 0 \tag{12}$$

Under above situations, the distribution of the stationary length of the queue can be expressed as:

$$\pi_n = c\theta(n), \quad \pi_n^+ = c\varepsilon(n), \quad n \ge 0 \tag{13}$$

c as the common normalisation factor.

Proof:

Assuming that the steady-state distribution of the length of packets queue exists on the time slot boundary, so the length in the time slot is $\{\pi_i^+, i = 0, 1, \cdots\}$, marked as:

$$\pi_i = \varpi(i), \pi_i^+ = \Psi(i), i = 0, 1, \cdots.$$

The formula (9) can be rewritten as:

$$p^D(n, i) = \frac{\Psi(n-i)}{\varpi(n)} b_i,\ i = 0, 1, \cdots \tag{14}$$

Because of $\sum_{i=0}^{n} p^D(n, i) = 1,$ we can get the formula:

$$\varpi(n) \sum_{i=0}^{n} \Psi(n-i) b_i,\ n \ge 0. \tag{15}$$

Assuming the formula (11) fits for $\{a_i,\ i = 0, 1 \cdots\}$ and formula (11) fits for $\{\Psi(n),\ n = 0, 1, \cdots\}$, The formula (11) can be rewritten as:

$$\begin{aligned} &\varpi(n+j-i)p^D(n+j-i, j)a_i = \Psi(n-i)a_j a_i \\ &\sum_{j=0}^{\infty}\sum_{i=0}^{n} \varpi(n+j-i)p^D(n+j-i, j)a_i \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{n} \Psi(n-i)a_j a_i \\ &= \varpi(n)\sum_{j=0}^{\infty} a_j = \varpi(n),\ n = 0, 1, \cdots. \end{aligned} \tag{16}$$

From the formula (17) and

$$\pi_n^+ = \sum_{i=0}^{\infty} \pi_{n+i} p^D(n+i, i),\ n = 0, 1, \cdots,$$

we can obtain:

$$\pi_n^+ = c\Psi(n),\ n = 0, 1, \cdots. \tag{17}$$

Therefore, the necessity and the sufficiency are proved.

2.3 Average queue length of data packets

To determine the average queue length of data packets, we assume that packets of nodes in the network have the first come first served (FIFO) strategy.

To start, we begin with modelling of the nodes, where the discrete time is based on the Geom/Geom/C queuing theory, following.

A data packet reaches one node with probability p ($0 < p < 1$) at head of each time slot (n, n^+). At the same time, no data packet reaches with probability $\bar{p} = 1 - p.$ All the packets' arrival constitutes a Bernoulli process with parameters p.

The beginning and ending of all packets' service occurs at the end of time slot (n^-, n), $n = 1, 2, \ldots$. The service time of packets are subject to be identically distributed, as:

$$P\{S = k\} = \bar{u}^{k-1}u,\ k = 1, 2, \ldots \tag{18}$$

There are parallel service nodes and queue with FIFO rules in the network. Arrival interval and service time are independent to each other.

According to the queuing theory, using L_n^+ to indicate the number of data packets when $t = n^+$ in instantaneous systems. Because the $\{L_n^+, n \ge 0\}$ is a neat – define MC before the term MC is showed (MC), there are countable state spaces $\Omega = \{0, 1, \ldots\}$, and the transition probability is:

$$p_{ij} = P\left\{L_n^+ = j \mid L_{n-1}^+ = i\right\},\ n \ge 1,\ i, j \in \Omega \tag{19}$$

When the number of customers in the system meets $i \le c - 1$, the state transition $i \to j(1 \le j \le i \le c - 1)$ behaves accordingly to two types of incompatible ways:

$$\begin{aligned} &p_{ij} = \bar{p}\binom{i}{j} u^{i-j}\bar{u}^{j} + p\binom{i}{j-1} u^{i-j+1}\bar{u}^{j-1}, \\ &1 \le j \le i \le c-1 \end{aligned} \tag{20}$$

The similar analysis is given:

$$p_{i0} = \bar{p}u^i,\ 0 \le i \le c-1, \tag{21}$$

When $i \ge c$, $L_{n-1}^+ = i$ contains on all nodes in the time slot (n^-, n) all busy, the transition probability does not depend on i at this time:

$$p_{i,i-c} = \bar{p}u^c,\ i \ge= c, \tag{22}$$

$$\begin{aligned} &p_{i,j} = \bar{p}\binom{c}{j} u^{c-j}\bar{u}^{j} + p\binom{c}{j-1} u^{c-j+1}\bar{u}^{j-1}, \\ &i \ge c,\ i-c \le j \le i, \end{aligned} \tag{23}$$

$$p_{i,i+1} = p\bar{u}^c,\ i \ge= c \tag{24}$$

Assuming that

$$\binom{c}{-1} = \binom{c}{c+1} = 0 \tag{25}$$

The transition probability in formula (23) can be unified represented as:

$$\begin{aligned} &p_{i,j} = \bar{p}\binom{c}{j} u^{c-j}\bar{u}^{j} + p\binom{c}{j-1} u^{c-j+1}\bar{u}^{j-1}, \\ &i \ge c,\ i-c \le j \le i+1 \end{aligned} \tag{26}$$

In short:

$$a_k = \overline{p}\binom{c}{k-1}u^{k-1}\overline{u}^{c-k+1} + p\binom{c}{k}u^k\overline{u}^{c-k}, \quad k = 0, 1, ..., c+1 \tag{27}$$

Obviously, when c services nodes are all busy, a_k is the probability when $L_{n-1}^+ - L_n^+ = k-1$ in a time slot (L_n^+ has more one packet than L_{n-1}^+ when $k = 0$, $L_{n-1}^+ - L_n^+ = 1$).

Using L^+ to indicate the stability limit of L_n^+, the distribution can be represented as:

$$\lambda_k = P\{L^+ = a_k\} = \lim_{n\to\infty} P\{L_n^+ = a_k\}, k \geq 0 \tag{28}$$

According to formula (27),

$$a(z) = \sum_{k=0}^{c+1} a_k z^k = (z\overline{p} + p)(uz + \overline{u})^c \tag{29}$$

Synthetically formula (28), (29) the system average queue length is:

$$\overline{\lambda_k} = \frac{Ka(z)}{(1-a(z))^2} \tag{30}$$

where K is pending constant factor, and its magnitude is set based on network packet format.

3 Related work

Compared to Zhang et al.'s (2015) report, which uses discrete queue to realise traffic prediction and is only suitable for small network data traffic, PFRF algorithm is suitable for all kinds of data traffic and not only small networks but medium to wide area ones as well.

Wei et al.'s (2013) investigation uses discrete time theory to get the network traffic prediction, which is only suitable for ad hoc network with slow-speed nodes, whilst PFRF algorithm works not only for these types of network nodes but also for high-speed nodes in ad hoc networks.

4 PFRF algorithm

In this work, the network traffic prediction algorithm based on the discrete time queuing theory fully considers factors such as node's MAC layer queue queuing and the arrival time of packets, providing reliable prediction about the network of the upcoming traffic. Therefore, this algorithm initialises the network yet to establish the balance relationship under stable state of the system, as well to calculate the average queue length of node i in the steady state through probability of the arrival business flow and corresponding generating formula (Chung et al., 2014). Finally, to select appropriate parameters to predict the network traffic, the algorithm proposed follows:

Step 1 The state parameter of node i and relevant links in Initialisation network at some time slot t.

Step 2 Circulate the first step. Establish the balance relations according to references (21) and formula (4) to (7):

$$(\lambda + \mu + w)p_{1,1} = \mu p_{1,2} + \theta p_{2,1} \tag{31}$$

$$(\lambda + \mu + w)p_{1,k} = \lambda p_{1,k-1} + \mu p_{1,k+1} + \theta p_{2,k}, k \geq 2 \tag{32}$$

$$(\eta + \theta)p_{2,1} = wp_{1,1} \tag{33}$$

$$(\eta + \theta)p_{2,k} = wp_{1,k} + \eta p_{2,k-1}, k \geq 1 \tag{34}$$

Step 3 Since data packets of node arrive in Bernoulli way, one node's data packet arrival probability at some time slot can be calculated as:

$$\overline{P} = P\{X = k\} = (1-p)p^k \tag{35}$$

Step 4 According to references (20), calculate the node's denial of service. Thus, the probability of data packets getting into congested queue is:

$$\delta = u(1-\overline{P})\overline{P}^{k-1} \tag{36}$$

Step 5 Calculate the transition probability of node i in bound of time slot s according to formula (7). It is:

$$p(n, n') = P\{L_{s+1} = n' \mid L_s = n\}, n, n' \in \Omega \tag{37}$$

Step 6 Calculate the best product result of node i according to formula (4),

$$\pi(n) = \prod_{j=1}^{N} \pi_j(n_j), \quad \pi^+(n) = \prod_{j=1}^{N} P\{L_j^+ = n_j\} \tag{38}$$

Step 7 Calculate the distribution on the steady length of queue of node i according to formulas (13) and (37).

$$\pi_n = \left(1-\frac{p}{q}\right)\left(\frac{p}{q}\right)^n, \quad n \geq 0; \qquad \pi_n^+ = \begin{cases} \dfrac{1}{\overline{p}}\left(1-\dfrac{p}{q}\right), & n = 0 \\ \dfrac{1}{\overline{p}}(1-q)\left(1-\dfrac{p}{q}\right)\left(\dfrac{p}{q}\right)^n & n \geq 1; \end{cases} \tag{39}$$

Step 8 Utilising formula (40) to choose truncation parameter M and Sample parameter N:

$$X(n) = \sum_{m=0}^{M-1} c(m)Z_\alpha(n-m), \quad X_n = \sum_{m=0}^{M-1} \pi_n, \quad n = 0, 1, ..., N-1 \tag{40}$$

Step 9 Utilising formula (9) to calculate coefficient $a(k)$, $k = 0, 1, \ldots, M + N - 1$ and utilise B is FFT algorithm to calculate the discrete Fourier transform $(m) = DM + N(c(m))$,

$$\begin{cases} c(0) = 1 \\ c(m) = \dfrac{\Gamma(m+d)}{\Gamma(d)\Gamma(m+1)} \end{cases} \quad (41)$$

Γ in it is gamma function.

Step 10 Generate $M + N$ numbers of steady distribution variable $Z(m)$ with standard symmetry α, and utilise B in FARIMA algorithm to calculate the discrete Fourier transform $\hat{Z}(m) = DM + N(Z(m))$ of cycle development $Z(m)$.

Step 11 Utilise B in FARIMA algorithm to calculate the inverse Fourier transform of product $(\hat{c}, \hat{Z})$, and

$$Y(t) = D_{M+N}^{-1}\left(\hat{c}\hat{z}(m)\right) \quad (42)$$

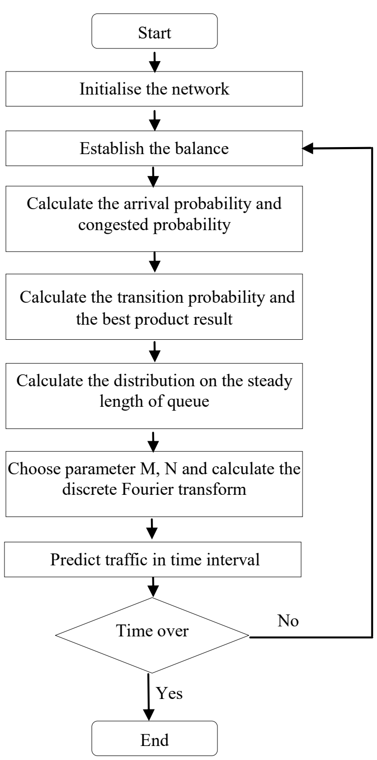
In this formula, $\hat{c}\hat{z}(m) = \hat{c}(m)\hat{z}(m)$, $Y(t)$ is the predicted flow;

Step 12 Make $t = t + 1$, skip to Step 1 until network gets into another condition;

Step 13 End.

The overall algorithm steps are shown in Figure 1.

Figure 1 Traffic prediction model algorithm



5 Experiments

In this section, the algorithm proposed and experiments to demonstrate it are discussed.

5.1 *Environment and methodology*

Simulation experiments were conducted in a personal computer with one 64-bit 2-core 3.2 GHz Intel processor, 8 GB memory, and Microsoft Windows 10 operating system. The software installed for analysis of this proposed algorithm is network simulator (NS) version 2.27 with 28 simulated nodes and mathematical analysis is MATLAB version 8.0.

In order to prove the rationality and accuracy of the research proposed in this paper, it is needed to establish FARIMA model of nodes in NS2, and based on fractional Brownian motion to generate network flow in point-to-point manner with cross shape.

The occurrence rate of average data packet on nodes should be 1,800 Kb/s, which is the mainstream speed in FARIMA model. To execute simulation for the first time, the proposed algorithm PFRF is used to simulate and predict network flow. For future steps, FBM algorithm is applied to monitor the network flow.

In this research, crossing data transmission is used as a method for data transmission. In experiments performed, the network data packet is abstracted into a node, from which the actual flow of network data can be simulated. To verify the effectiveness of the algorithm proposed, experiments above presented are carried out from the overall network traffic simulation.

Concurrently, result analysis of simulation in MATLAB which is the main mathematical analysis tool software and effectiveness evaluation of node in network are performed.

5.2 *Results*

Effective evaluations of network nodes are shown on Figure 2. In this same figure, the flow predicted by PFRF algorithm is close to the original flow. From formula (43), it is obtained as result the standard deviation 9.87 compared to the maximum throughput $4*10^4$ Kb, what shows the accuracy this proposed algorithm may achieve is favourable and promising.

$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - u)^2} \quad (43)$$

Analyses of main factors that may affect the flow performance are presented next. Assuming that ω = 0.3, θ = 0.7 and η = 0.6 which is mainstream parameter in FARIMA model. Figure 3 shows the relationships with the different flow arrival rate λ between average queue length $\bar{\lambda}$ and service rate μ. At the same time, it is possible to note a negative correlation between the average length of queue and the service rate. At small service rates, as larger λ, the smaller the average queue length and better the performance is. Nevertheless, at higher service rate, the curve is mutated.

Comment [T7]: Author: Please confirm if the equation is correct.

With the smaller λ, the smaller the average length of queue and better the performance is.

Figure 2 Comparison of PFRF prediction algorithm and actual flow (see online version for colours)

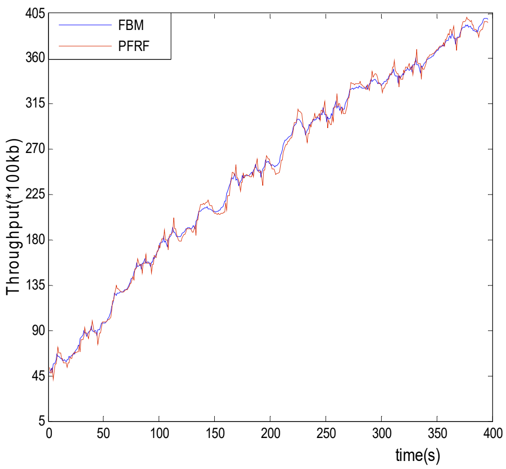
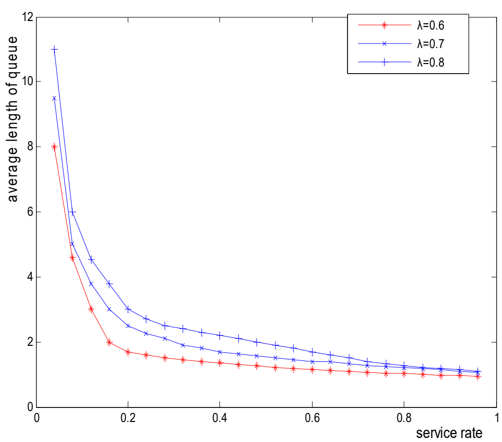


Figure 3 Comparison between average queue length and service rate with different traffic arrival rate λ (see online version for colours)



It can be seen from the experiments that the PFRF algorithm based on the PRF can effectively restrain the average queue length and improve the robustness of the network. Especially in ad hoc network, the improvements have a great help to improve the usability of the node.

Concurrently, assuming that ω = 0.3, θ = 0.7 and η = 0.6, it is shown in Figure 4 the relationships with different flows' arrival rate η between the average length of queue $\bar{\lambda}$ and service rate μ. It is possible to note that the average length of queue is increasing with the service rate increases to a maximum value, and decreases to a stabilised state at end. It shows that, if one node A is failed, the more flow reaches A, the greater impact to the system is.

Figure 4 Comparison between average queue length and service rate with different traffic arrival rate η (see online version for colours)

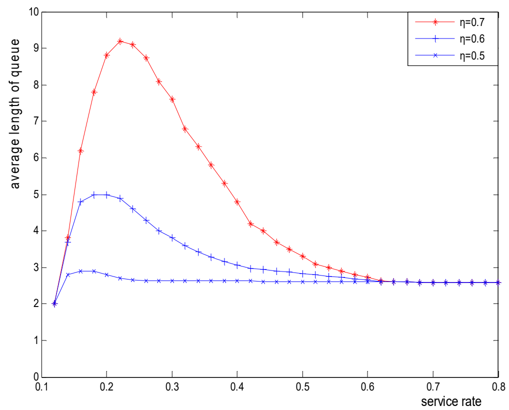
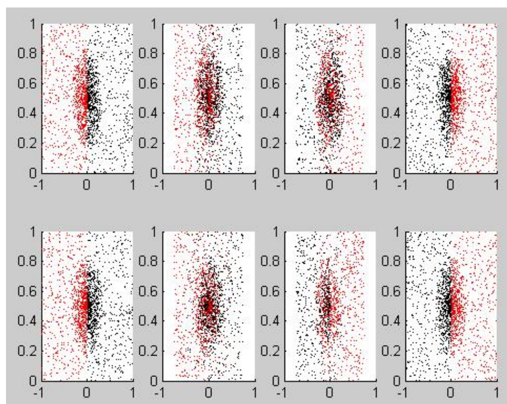


Figure 5 Comparison of actual network traffic and predicted by proposed algorithm (see online version for colours)



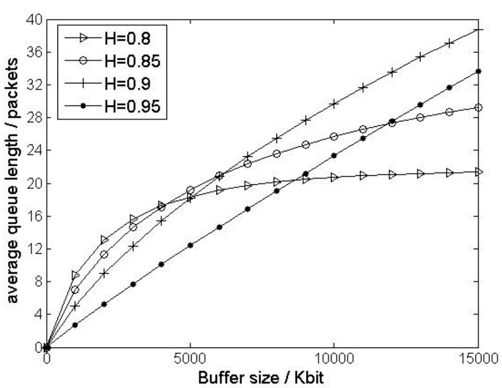
Simulation of network data packet abstracted as nodes as presented in Figure 5. The network for different times of the actual data flow is shown in upper half part of Figure 5, whereas the PFRF algorithm for the prediction of traffic is shown in the lower half. It can be observed that the predicted and the actual flow are found in corresponding moments.

Meanwhile, in order to investigate the key factors affecting PFRF methods, this work follow the experiment of generating traffic streams under different fractal H parameter values, studying the performance changes at nodes S as well. The variation relationship between average queue length and the buffer of the parameter values under different H are shown in Figure 6. In this figure, the buffer is increased with the increase of the average queue length. As the buffer is small and larger value of H, this corresponds that the curve of the average queue length is smaller. Next, as buffer turns to be large, the transition happens, the smaller the value of parameter *H*, the smaller the curve corresponding to average queue length.

However, traditional studies do not fully take into account the impact of network factors, emphasised on the impact of traffic-related properties, one-sided to receive the greater the worse performance conclusions. As shown in Figure 6, this phenomenon to reset the effect of finite buffer and truncation effect are explained. As the buffer is empty, the reset effect weaken the traffic flow is associated with the long front of the traffic flow. On the other hand, as the buffer is full and the subsequent arrival traffic is dropped, the truncation effect will weaken the long-related effects.

Through these experiments, we can see that the PFRF algorithm adjust the product solution application parameters in the different traffic flow data to suppress the length of queue and enhance the robustness of networks. So the PFRF algorithm has a very effective application in traffic prediction.

Figure 6 Relationship between average queue length and buffer size



5.3 Validation

The experiment results show that PFRF algorithm is efficient: the PFRF model based on fractal Brown motion FBM and ARMR process also have the ability to describe the long related and the underlying buffer queue, so it can be used in modelling and analysis the self-adaptive network traffic. The traffic prediction method based on PRF in this paper shown the discrete time node of the packet queuing system, used the length of queue at steady-state and pulled-into discrete time quasi reversible queuing theory, got the optimal conditions of PRF in the discrete time queue network. The experiments show that the PFRF model has an accurate flow prediction mechanism, which is suitable for network traffic modelling.

6 Conclusions and future work

It is presented in this paper relevant and constructive discussion based on discrete time, raising a new alternative to predict actual network flow PFRF when nodes in network lose effectiveness. This algorithm makes use of queuing theory and FARIMA model to investigate the conditions where queue of node data packet and average length of queue strike, and further analyse the main factors that influence function of flow through simulation. Results obtained could verify the effectiveness of PFRF algorithm.

As future research work, topics that include the combination of queuing system with start time and queuing system arriving in batches will be analytically studied, as well to establish a complete set of network function evaluation system. The prediction model proposed will also be considered to be coupled to studies on resource index and discovery system in distributed computing environments (Chung et al., 2012, 2013, 2014), where the predictions on the dynamic traffic are of fundamental importance to higher quality of service of distributed systems. As known, the network delay on network performance is of great importance, identifying main factors through the existing methods based on the traffic prediction to improve the performance of network is kept as aim of this research work. Another topic aimed is energy consumption and efficiency, which is fundamental and relevant in modern infrastructures. An investigation of the proposed algorithm on trade-offs in terms of network bandwidth and energy is also considered.

Acknowledgements

We acknowledge the support from the Scientific Research Fund of SiChuan Provincial Department of Science and Technology (2016GZ0113), Scientific Research Fund of Key Laboratory of Pattern Recognition and Intelligent Information Processing of Chengdu University (MSSB-2015-12), Scientific Research Fund of Sichuan Provincial Department of Science and Technology (2015GZ0333), Scientific Research Fund of State's Key Project of Research and Development Plan (2016YFB0800605).

References

You, W., Qian, K. and Qian, Y. (2016) 'Software-defined network flow table overflow attacks and countermeasures', *Int. J. of Soft Computing and Networking*, Vol. 1, No. 1, pp.70–81.

Asakura, Y. (2007) 'Requirements for transport network flow models used in reliability analysis', *Int. J. of Critical Infrastructures*, Vol. 3, Nos. 3/4, pp.287–300.

Divakaran, S. and Chinnagounder, C. (2015) 'A framework for topology-based traffic grooming with restoration in optical networks', *Int. J. of High Performance Computing and Networking*, Vol. 8, No. 4, pp.358–369.

Su, X., Zhang, D., Dai, S., Zhang, J. and Wang, D. (2016) 'Mobile traffic identification based on application's network signature', *Int. J. of Embedded Systems*, Vol. 8, Nos. 2/3, pp.217–227.

Nakayama, A. and Gang, P.L. (2013) 'Improved algorithm using generalised flows for an optimisation problem in a cash flow network', *Asian J. of Management Science and Applications*, Vol. 1, No. 1, pp.67–95.

Chuang, P-J. and Jiang, Y-T. (2006) 'Dynamic channel selection: an efficient strategy for balancing traffic in meshes', *Int. J. of Computational Science and Engineering*, Vol. 2, Nos. 1/2, pp.3–22.

Ye, X., Zhang, J., Liu, Y. and Su, J. (2015) 'Study on the measurement of international knowledge flow based on the patent citation network', *Int. J. of Technology Management*, Vol. 69, Nos. 3/4, pp.229–245.

Abraham, J.N. and Rao, K.S. (2008) 'Production planning through flow network optimisation and mixed integer linear programming models in a petroleum refinery', *Int. J. of Operational Research*, Vol. 3, No. 3, pp.315–335.

Todinov, M.T. (2012) 'Algorithms for minimizing the lost flow due to failed components in repairable flow networks with complex topology', *Int. J. of Reliability and Safety*, Vol. 6, No. 4, pp.283–310.

Sahinoglu, M. and Libby, D.L. (2005) 'Measuring availability indexes with small samples for component and network reliability using the Sahinoglu-Libby probability model', *IEEE Transactions on Instrumentation and Measurement*, Vol. 54, No. 3, pp.1283–1295.

Liu, A. and Liu, Y. (2002) 'Traffic performance analysis of network with unreliable components', *Chinese Journal of Electronics*, Vol. 30, No. 10, pp.1459–1462.

Wei, J., You, L. and Hu, J. (2013) 'A novel depict method of actual traffic performance', *Microelectronics & Computer*, Vol. 30, No. 3, pp.35–37.

Zhang, H., Hu, J. and Shen, Y. (2015) 'A new network traffic prediction algorithm based on discrete queue', *Journal of Computational Information Systems*, Vol. 11, No. 12, pp.4255–4261.

Kim, B. (2009) 'Tail asymptotics for the queue size distribution in a discrete-time Geo/G/1 retrial queue', *Queueing System*, Vol. 61, pp.243–254.

Zhao, D. and Deng, B. (   ) 'The stationary queue-length distribution for the repaired M/G/1 queue under N-policy with delayed startup-closedown based on the maximum entropy', *Mathematics in Practice and Theory*, Vol. 40, No. 18, pp.140–147.

Nolan, J. and Panorska, A. (2001) 'K estimation of stable spectral measures', *Math Computer Modeling*, Vol. 4, No. 8, pp.1113–1122.

Tian, N., Xu, X. and Ma, Z. (2008) *Discrete-time Queuing Theory*, Science Press, Beijing.

Tan, X., Huang, Y. and Jin, W. (2007) 'Modeling and performance analysis of self-similar traffic based on FBM', *2007 IFIP International Conference on Network and Parallel Computing – Workshops*, pp.543–548.

Tang, Y. and Tang, X.I. (2006) *Queuing Theory*, Science Press.

Chung, W-C., Lin, Y-H., Lai, K-C., Li, K-C. and Chung, Y (2012) 'A self-adaptive resource index and discovery system in distributed computing environments', *IJAHUC – International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 10, No. 2, Inderscience Publishers.

Chung, W-C., Hsu, C-J., Lai, K-C., Li, K-C. and Chung, Y-C. (2013) 'Direction-aware resource discovery in large-scale distributed computing environments', *The Journal of Supercomputing*, March, Vol. 66, No. 1, pp.229–248, Springer.

Chung, W-C., Hsu, C-J., Lai, K-C., Li, K-C. and Chung, Y-C. (2014) 'On maintenance of cooperative overlays in multi-overlay networks', *IET Communications*, October, Vol. 8, No. 15, pp.2676–2683.

Comment [T8]: Author: Please provide the issue number.

Comment [T9]: Author: Please cite the reference in the text or delete from the list if not required.

Comment [T10]: (1) Author: Please cite the reference in the text or delete from the list if not required.

(2) Author: Please provide the place of publication.

Comment [T11]: Author: Please provide the page numbers.