Performance Evaluation 109 (2017) 8–22

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

Whittle-networks with signals

Thu-Ha Dao-Thi [a], J.M. Fourneau [b,*], Minh-Anh Tran [c]

[a] Institute of Mathematics, VAST, Hanoi, Viet Nam
[b] DAVID, Univ. Versailles St-Quentin, France
[c] LACL, Univ. Créteil, France

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form 2 October 2016
Accepted 19 December 2016
Available online 3 January 2017

Keywords:
Queueing networks
Signals product form steady-state distribution

ABSTRACT

We present an extension of Whittle networks with customers and signals as defined by Gelenbe in his seminal paper on negative customers. Customers are queued and served according to the balance rules defined for Whittle networks while signals are not queued but interact with the customers at their arrival. We consider various types of signals: they can delete customers, flush out a queue or change the class of a customer. We prove that all these networks have a product form steady-state distribution as soon as the signals obey a balance rule consistent with the service balance defined by Whittle.

1. Introduction

Whittle networks have two interesting properties: their steady-state distributions (when they exist) have a product form solution and they are insensitive [1]. The insensitivity property implies that the probability distribution does not really depend on the service time distribution but only on the first moment of this distribution. Thus, we can replace an arbitrary distribution for the service times by an exponential distribution with the same mean which gives an easier Markov chain to analyze. Such a simplification of the model was the key property to analyze flows on Internet (see for instance [2–7]). Clearly, flows do not have an exponential duration but some load balancing mechanisms are very similar to the balance equation for Whittle networks and it explains the success of Whittle networks as modeling tool for flows over Internet.

In [8], to put more emphasis on the model defined by Whittle in [9], Serfozo has introduced Whittle network process as a class of queueing networks with state dependent transition probabilities which obey a balance property (for a formal definition, see Definition 1). It was assumed that the services follow exponential distributions. Let $\vec{x}$ be the vector of number of customers in the queues and let $\phi_i(\vec{x})$ be the service rate at queue i when the network state is $\vec{x}$. Let $\vec{x} - e_j$ be the state $\vec{x}$ with one customer less at queue j.

Definition 1. The service rates $\phi_i(\vec{x})$, functions of the state vector $\vec{x}$, are balanced if

$$\phi_i(\vec{x})\phi_j(\vec{x} - e_i) = \phi_j(\vec{x})\phi_i(\vec{x} - e_j), \quad \forall i, j : x_i > 0, x_j > 0.$$

The balance property has a very simple physical interpretation. Given a state $\vec{x}$, and $\langle \vec{x}, \vec{x} - e_{i_1}, \vec{x} - e_{i_1} - e_{i_2}, \dots, 0\rangle$ a direct path from state $\vec{x}$ to state $\vec{0}$, the balance property says that the following expression, which is the product of the service rates along the path, does not depend on the considered path:

$$\Phi(\vec{x}) = \frac{1}{\phi_{i_1}(\vec{x})\phi_{i_2}(\vec{x} - e_{i_1}) \dots \phi_{i_m}(\vec{x} - e_{i_1} - \dots - e_{i_{m-1}})}. \quad (1)$$

* Corresponding author.
E-mail addresses: dttha@math.ac.vn (T.-H. Dao-Thi), Jean-Michel.Fourneau@uvsq.fr (J.M. Fourneau), thu-ha.dao-thi@m4x.org (M.-A. Tran).

http://dx.doi.org/10.1016/j.peva.2016.12.001
0166-5316/

In a Whittle network, the service rate functions $\phi_i(\vec{x})$ are exclusively characterized by the balance function Φ:

$$\phi_i(\vec{x}) = \frac{\Phi(\vec{x} - e_i)}{\Phi(\vec{x})}, \quad i = 1, \ldots, N, x_i > 0. \tag{2}$$

The most famous example of Whittle networks is a network of Processor Sharing queues. Another well-known example is a network with constant capacity C, with one class of customer, and a scheduler for queue i such that $\phi_i(x) = \frac{Cx_i}{\sum_j x_j}$.

The balance property was used in [6] to define balance fairness for a scheduler. A data network is considered as a set of links shared by some competing flows. Balance fairness implies insensitivity in the sense that the stationary distribution does not depend on any traffic characteristics except the traffic intensity on each route. This insensitivity property does not hold in general for well-known allocations such as max–min fairness (where the rates of individual flows are made as equal as possible) or proportional fairness (utility-based allocations). It was shown in [4], that the balance fairness property is an accurate approximation for max–min fairness and proportional fairness, two properties used to define scheduler for flows. In [2], the Erlang model and many other well-known models (such as Processor Sharing) which are all useful for studying communication networks, were shown to satisfy the balance property defined by Whittle, thus enhancing the potential applications of Whittle networks. More recently, the balance fairness property was also proved to be an accurate bound for greedy rate or alpha fairness [7].

Here, we consider networks with multiple classes of customer and several types of signals. It is worth pointing out that in a Whittle network of single-class queues, service rates are balanced with respect to queues. In this paper, we consider Whittle networks of multi-class queues where system state $\vec{x}$ takes into account not only the customer number in each queue, but the customer number $x_i^{(k)}$ of each class k in every queue i. The service rates are then functions of these variables and the balance property is defined with respect to classes.

Depending on the context, we shall consider the balance property only within each queue (Definition 2) or over the whole network (Definition 3).

Definition 2. The service rates $\phi_i^{(k)}(\vec{x}_i)$ are called balanced within queue i if there exists a balance function $\Phi_i(\vec{x}_i)$ such that

$$\phi_i^{(k)}(\vec{x}_i) = \frac{\Phi_i(\vec{x}_i - e_i^{(k)})}{\Phi_i(\vec{x}_i)} \quad \forall k. \tag{3}$$

Definition 3. The service rates $\phi_i^{(k)}(\vec{x})$ are called balanced over the whole network if there exists a global balance function $\Phi(\vec{x})$ such that

$$\phi_i^{(k)}(\vec{x}) = \frac{\Phi(\vec{x} - e_i^{(k)})}{\Phi(\vec{x})} \quad \forall k, \forall i. \tag{4}$$

Remark that Definition 2 is less general than Definition 3. The methods we use to analyze the network will depend on the way the network is balanced (i.e. balance over the network, or balance within a queue) and the type of signals considered in the model.

The theory of queues with signals has received considerable attention since the seminal paper on positive and negative customers [10] published by Gelenbe more than 20 years ago. Traditional queueing networks models are used to represent contention among customers for a set of resources. Customers move from server to server. They wait for service. They do not interact among themselves. In a network of queues with signals (also denoted as a G-network of queues) signals interact at their arrival into a queue with customers already backlogged. Furthermore, customers are allowed to change into signals at the completion of their service. Signals are never queued. At their arrival, they try to interact. If they fail, they disappear immediately. Despite this deep modification of the classical model, G-networks still preserve the product form property for the steady-state distribution of some Markovian queueing networks. G-networks have been used to model many artificial or biological systems. Cognitive Packet Networks and Random Neural Networks are the most well-known applications (see [11] for a bibliography).

The first type of signal was introduced as negative customers in [10]. A negative customer deletes a positive customer at its arrival at a non empty queue. Positive customers are usual customers which are queued and receive service or are deleted by negative customers. Under typical assumptions for Markovian queueing networks (Poisson arrivals for both types of customers, exponential service times for positive customers, Markovian routing of customers, open topology, independence) Gelenbe proved that such a network has a product form solution for its steady-state behavior when the chain is ergodic. Note that the open topology is mandatory because in a closed queueing network with negative customers, the positive customers disappear and all the queues are empty at steady-state with probability one.

It must be clear that the results are more complex than Jackson's networks. The G-networks flow equations exhibit some uncommon properties: they are neither linear as in closed queueing networks nor contracting as in Jackson queueing networks. Therefore the existence of a solution had to be proved [12] by new techniques from the theory of fixed point equation.

G-networks had also motivated many new important results in the theory of queues. As negative customers lead to customer deletions, the original description of quasi-reversibility by arrivals and departures does not hold anymore and a new version had been proposed by Chao and his co-authors in [13]. A different approach, based on Stochastic Process Algebra,