Fork and Join Queueing Networks with Heavy Tails: Scaling Dimension and Throughput Limit

Yun Zeng
The Ohio State University
Columbus, Ohio
zeng.153@osu.edu

Jian Tan
The Ohio State University
Columbus, Ohio
tan.252@osu.edu

Cathy H. Xia
The Ohio State University
Columbus, Ohio
xia.52@osu.edu

ABSTRACT

Parallel and distributed computing systems are foundational to the success of cloud computing and big data analytics. Fork-Join Queueing Networks with Blocking (FJQN/Bs) are natural models for such systems. While engineering solutions have long been made to build and scale such systems, it is challenging to rigorously characterize the throughput performance of ever-growing systems, especially in the presence of heavy-tailed delays. In this paper, we utilize an infinite sequence of FJQN/Bs to study the throughput limit and focus on regularly varying service times with index $\alpha > 1$. We introduce two novel geometric concepts - scaling dimension and extended metric dimension - and show that an infinite sequence of FJQN/Bs is throughput scalable if the extended metric dimension $< \alpha - 1$ and only if the scaling dimension $\leq \alpha - 1$. These results provide new insights on the scalability of a rich class of FJQN/Bs.

CCS CONCEPTS

• **Networks** → **Network performance analysis**; *Topology analysis and generation*;

KEYWORDS

Fork/join, queueing network, scalability, heavy tails, network dimension, throughput limit

ACM Reference Format:
Yun Zeng, Jian Tan, and Cathy H. Xia. 2018. Fork and Join Queueing Networks with Heavy Tails: Scaling Dimension and Throughput Limit. In *SIGMETRICS '18 Abstracts: ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems Abstracts, June 18–22, 2018, Irvine, CA, USA*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3219617.3219668

1 INTRODUCTION

Parallel and distributed computing systems are foundational to the success of cloud computing and big data analytics. Numerous large-scale analytics have been developed over distributed servers to achieve high performance. Parallel and distributed computing also exhibits itself in wireless sensor networks, in composite web services, in distributed stream computing, in distributed file systems, in MapReduce frameworks, in end-system multicast, etc.

SIGMETRICS '18 Abstracts, June 18–22, 2018, Irvine, CA, USA

ACM ISBN 978-1-4503-5846-0/18/06.
https://doi.org/10.1145/3219617.3219668

As the sizes of various parallel and distributed computing systems continue to grow, their throughput performance could degrade due to synchronization delays, processing time variations, or data storage, I/O, and bandwidth constraints. The problem has been well recognized in all kinds of distributed computing environments but the analysis is non-trivial. What further complicates the investigation is the presence of heavy-tailed processing times that have been widely documented therein. These heavy-tailed processing times can cause extremal delays that directly impact the synchronization and bring down the throughput. One critical issue concerns *throughput scalability*: can we properly design a parallel and distributed processing system in massive scale under heavy-tailed delays so that the throughput performance can be sustained? While practical engineering solutions have long been made to build and scale such systems, the mathematical foundations toward understanding the throughput performance of ever-growing systems remain rudimentary.

2 MODEL

The above parallel and distributed computing systems can be naturally modeled as fork-and-join queueing networks with blocking (FJQN/Bs). A FJQN/B, denoted by $N = (V, E)$, consists of a set of nodes V representing servers and a set of directed arcs E representing routing of jobs. Associated with each arc, there is a buffer of finite capacity for job storage between services.

Each node models a single server that adopts the First Come First Serve policy. Services are conducted in a fork-join manner: each service consumes exactly one job from every upstream buffer and generates exactly one job to every downstream buffer. A server is starved (blocked) if one of the upstream (downstream) buffers is empty (full). An idle server can schedule a service only when it is neither blocked nor starved. During the service, jobs remain in the upstream buffers. For simplicity, we consider a homogeneous setting where all buffers are of constant size $b < \infty$ and all service times are i.i.d. of distribution F_σ. In particular, we focus on the cases where F_σ is regularly varying with index $\alpha > 1$.

For a given FJQN/B, the throughput at node $v \in V$ is defined as the average number of service completions in a unit time in the long run. Under i.i.d. service times, the throughput is identical at every node, which is referred to as the network throughout and can be expressed as

$$\theta(N) = \left(\lim_{m \to \infty} \frac{\mathbb{E}\left[T_{m,v}(N)\right]}{m} \right)^{-1}, \tag{1}$$

where $T_{m,v}(N)$ denotes the m-th service completion time at node v.

To investigate the throughput limit, we utilize an infinite sequence of FJQN/Bs $\mathcal{N} = \{N_1, N_2, \ldots, N_i, \ldots\}$ to characterize the

way the system grows. Each $N_i = (V_i, E_i)$ is a finite-sized FJQN/B. This sequence $\mathcal{N}$ is said to be *throughput scalable* if the limit infimum of the network throughput is strictly positive.

3 PRELIMINARIES

Previous studies on scalability of FJQN/Bs either focus on special network structures or assume light-tailed service times. [4] discusses the throughput limit of an infinite tandem queueing network with blocking. [5] shows the linear growth of the maximum weighted path on a lattice. The work is extended by [3] to address the scalability of pattern grid, which applies to FJQN/Bs of lattice structures. [1] shows the scalability of a multicast tree under light-tailed service times and bounded degree of the tree. For generally structured networks, [6] presents necessary conditions for scalability when service times are either light-tailed or of Pareto distributions. In particular, [7] shows that

$$\limsup_{i\to\infty} D_i < \infty \text{ and } \limsup_{i\to\infty} L_i^* < \infty \tag{2}$$

is a necessary and sufficient condition for throughput scalability of FJQN/Bs under light-tailed service times, where D_i and L_i^* represent respectively the network degree and the minimum level of N_i.

However, the question remains on how to guarantee throughput scalability of arbitrarily structured FJQN/Bs under heavy-tailed service times. As illustrated by the following examples, when service times are regularly varying, Condition (2) is not enough to guarantee scalability. This observation motivates us to propose the concepts of network dimensions in Section 4 so as to determine scalability of FJQN/Bs with heavy tails.

Tandem Network: Consider a sequence of FJQN/Bs $\mathcal{N} = \{N_i\}_{i=1}^{\infty}$ where N_i is a tandem network with a single source and i downstream nodes as shown in Table 1(a). It is easily verified that Condition (2) holds and hence the system is scalable under light-tailed service times. However, in heavy-tailed scenarios, if the service times are regularly varying with index $\alpha < 2$, then the sequence will not be throughput scalable. The argument is based on last-passage percolation model and extreme value theory. In fact, the existence of the second moment of the service time distribution is known necessary for scalability of tandem networks [4].

Binary Tree Network: In comparison, consider a sequence $\mathcal{N} = \{N_i\}_{i=1}^{\infty}$ where N_i is a binary tree network with a root and i layers (see Table 1(j)). For such system, Condition (2) is again satisfied, which is enough to guarantee scalability if the service times are light-tailed. However, under regularly varying service times with any index $\alpha \in \mathbb{Z}^+$, the throughput is not scalable. This is mainly due to the exponential growth of the network size which imposes an exponential decay on the throughput. Similar discussions appear in [2].

Lattice Network: Consider a sequence $\mathcal{N} = \{N_i\}_{i=1}^{\infty}$ where N_i is a d-dimensional lattice network with i nodes on each side (see Table 1(e)(f)). For this system, any $\alpha < d+1$ will make the sequence not scalable. Meanwhile, we can show that any $\alpha > d+1$ will make the sequence scalable. This sufficient condition is based on bounding the throughput by the growth of lattice animals and the results in [5].

From the above examples, we observe that in addition to the network degree and the minimum level, the throughput limit under heavy-tailed service times depends on complicated characterizations of how the network scales. This motivates the propositions of network dimensions in Section 4.

4 CHARACTERIZATION OF SCALING

The following common topological concepts are needed to characterize network topology. Let $G=(V,E)$ be the undirected counterpart of $N=(V,E)$. The distance of two nodes u, v in G, denoted as $dis(u,v)$, is the minimum number of edges among all undirected paths connecting u and v. The diameter of a graph G, denoted as $\Delta(G)$, is the maximum of the distance of any pair of nodes in the graph, i.e. $\Delta(G) = \max\{dis(u,v)\big|\forall u, v \in V\}$. The diameter of a network N is the diameter of its undirected counterpart G.

In the rest of this section, we introduce the scaling dimension as a way to characterize the growth of the most critical part of the sequence by a function of network size and diameter, and we introduce the extended metric dimension as a way to map networks onto lattices.

4.1 Scaling Dimension

Consider an infinite sequence of FJQN/Bs $\mathcal{N}=\{N_i\}_{i=1}^{\infty}$ under Condition (2). Let $\Omega\left(\bar{\mathcal{I}}, \bar{\mathcal{N}}\right)$ be the collection of $\left(\bar{\mathcal{I}}, \bar{\mathcal{N}}\right)$ satisfying the following:

1) $\bar{\mathcal{I}} = \{i_n\}_{n=1}^{\infty}$ is a sequence of strictly increasing natural numbers;

2) $\bar{\mathcal{N}} = \{\bar{N}_{i_n}\}_{n=1}^{\infty}$, where $\bar{N}_{i_n} = (\bar{V}_{i_n}, \bar{E}_{i_n})$ is a connected subnetwork of N_{i_n} with $\bar{V}_n \subseteq V_{i_n}$ and $\bar{E}_n \subseteq E_{i_n}$;

3) $\Delta(\bar{N}_{i_n}) \to \infty$ as $n \to \infty$.

The scaling dimension of the sequence $\mathcal{N}$ is defined as

$$dim_S(\mathcal{N}) = \sup_{(\bar{\mathcal{I}},\bar{\mathcal{N}})\in\Omega(\bar{\mathcal{I}},\bar{\mathcal{N}})} \left\{\limsup_{n\to\infty} \frac{\log|\bar{V}_{i_n}|}{\log\Delta(\bar{N}_{i_n})}\right\}. \tag{3}$$

Briefly speaking, the scaling dimension is given by the ratio of log network size over log diameter as the network expands. One can interpret the scaling dimension as a metric to measure how fast network grows as a function of network size and diameter. In particular, if $\mathcal{N}$ converges to a connected infinite graph that is locally-finite, then the scaling dimension is in analog with the growth degree in geometric group theory or the upper internal scaling dimension in Physics. If $\mathcal{N}$ converges to a fractal, then the scaling dimension is in analog with the box counting dimension or the Hausdorff dimension.

4.2 Extended Metric Dimension

Let $\mathcal{W} = \{W_1, W_2, \ldots, W_k\}$ be an ordered set of subsets of nodes in a graph $G = (V, E)$ with $W_t \subseteq V$, $t = 1, 2, \ldots, k$. The extended metric representation of a node v with respect to $\mathcal{W}$ is given by

$$\bar{\boldsymbol{r}}(v|\mathcal{W}) = \big(dis(v, W_1), dis(v, W_2), \ldots, dis(v, W_k)\big), \tag{4}$$

where $dis(v, W_t)$ is the shortest distance between v and any node in W_t, $t = 1, 2, \ldots, k$. A set $\mathcal{W} = \{W_1, W_2, \ldots, W_k\}$ of subsets of V is a Λ-extended resolving set for G, if for all $v \in V$, the number of nodes $u \in V$ with $\bar{\boldsymbol{r}}(u|\mathcal{W}) = \bar{\boldsymbol{r}}(v|\mathcal{W})$ is bounded above by a constant $\Lambda > 0$.

Consider an infinite sequence of FJQN/Bs $\mathcal{N} = \{N_i\}_{i=1}^{\infty}$. The extended metric dimension of $\mathcal{N}$, denoted as $dim_{EM}(\mathcal{N})$, is the minimum integer k such that, $\forall i \in \mathbb{Z}^+$, the undirected counterpart of N_i has a Λ-extended resolving set $\mathcal{W}_i$ with cardinality $\leq k$, where $\Lambda > 0$ is a constant independent of i.

The concept derives from a graph's metric dimension: the minimum cardinality of a basis that uniquely identifies every node by its distance to the basis. Briefly speaking, the extended metric dimension is given by the minimum cardinality of a basis that identifies nodes up to a constant level as network expands. One can interpret the extended metric dimension as the least number of coordinates needed to describe the network viewed far away as it expands.

4.3 Relationship Between Dimensions

With respect to the relationship between the two dimensions, we show that the scaling dimension is bounded above by the extended metric dimension. In most common networks with integer scaling dimensions, the two dimensions coincide.

5 SCALABILITY CONDITIONS

We show that, under regularly varying service times, the throughput scalability of FJQN/Bs is determined by the relationship among the network dimensions and the service time tails. Our main result is given as follows, which includes a necessary condition and a sufficient condition on throughput scalability of FJQN/Bs under regularly varying service times. See detailed proofs in [8].

Theorem 1. *Consider an infinite sequence of FJQN/Bs $\mathcal{N} = \{N_i\}_{i=1}^{\infty}$, where $N_i = (V_i, E_i)$ is a finite-sized FJQN/B with $|V_i| < \infty$, $\forall i \in \mathbb{Z}^+$, and $\limsup_{i\to\infty} |V_i| = \infty$. The service times are i.i.d. regularly varying with index $\alpha > 1$. Under condition (2), the sequence $\mathcal{N}$ is throughput scalable if the extended metric dimension $dim_{EM}(\mathcal{N})$ satisfies*

$$dim_{EM}(\mathcal{N}) < \alpha - 1 \tag{5}$$

and only if the scaling dimension $dim_S(\mathcal{N})$ satisfies

$$dim_S(\mathcal{N}) \leq \alpha - 1. \tag{6}$$

Theorem 1 reveals that Condition (2) is not enough to address throughput scalability in heavy-tailed cases. We need additional conditions on network dimension to ensure that the growth degree of the networks is bounded by the heavy tail index of the service time distribution. This result provides new insights on the scalability of a rich class of FJQN/Bs under various structures, including tandem, lattice, hexagon, pyramid, tree, and fractals. Table 1 provides a list of network examples with scalability conditions in addition to Condition (2).

6 CONCLUSION

This paper investigates throughput scalability of fork-join queueing networks with blocking under heavy-tailed service times. In particular, we focus on cases where service times are regularly varying with index α. We introduce two novel geometrical concepts for generally structured FJQN/Bs: scaling dimension and extended metric dimension. We show that a sequence of FJQN/Bs is throughput scalable if its extended metric dimension $< \alpha - 1$ and only if its scaling dimension $\leq \alpha - 1$. The results apply to a list of FJQN/Bs including tandem, lattice, hexagon, tetrahedron pyramid networks, and even fractals. The results can be useful for designing parallel and distributed computing systems in heavy-tailed environments as well as for analysis of other scaling networks or fractals such as social networks, electrical grid, Internet of Things, etc.

Table 1: Examples with Scalability Conditions

Name	Structure	Scalability Conditions	
		Necessary	Sufficient
Tandem	(a)	$\alpha \geq 2$	$\alpha > 2$
Tandem-alike	(b) (c) (d)	$\alpha \geq 2$	$\alpha > 2$
d-D Lattice	(e) 2-D (f) 3-D	$\alpha \geq d+1$	$\alpha > d+1$
Hexagon	(g)	$\alpha \geq 3$	$\alpha > 3$
Tetrahedron Pyramid	(h)	$\alpha \geq 4$	$\alpha > 4$
Sierpinski Triangle	(i)	$\alpha \geq 1+\log_2 3$	$\alpha > 3$
Binary Tree	(j)	light-tailed	light-tailed

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants CNS-1717060, IIS-0916440, ECCS-1232118, SES-1409214.

REFERENCES

[1] F. Baccelli, A. Chaintreau, Z. Liu, and A. Riabov. 2005. The one-to-many TCP overlay: A scalable and reliable multicast architecture. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, Vol. 3. IEEE, 1629–1640.

[2] A. Chaintreau. 2006. *Processes of Interaction in Data Networks.* PhD thesis, INRIA-ENS.

[3] A. Chaintreau. 2008. Sharpness: a tight condition for scalability. *Structural Information and Communication Complexity* (2008), 74–88.

[4] J.B. Martin. 2002. Large tandem queueing networks with blocking. *Queueing Syst.* 141, 1-2 (2002), 45–72.

[5] J.B. Martin. 2002. Linear growth for greedy lattice animals. *Stoch. Proceedings Appl.* 98, 1 (2002), 43–66.

[6] C.H. Xia, Zhen Liu, Don Towsley, and Marc Lelarge. 2007. Scalability of fork/join queueing networks with blocking. In *Proceedings of ACM Sigmetrics.*

[7] Y. Zeng, A. Chaintreau, D. Towsley, and C.H. Xia. 2016. A Necessary and Sufficient Condition for Throughput Scalability of Fork and Join Networks with Blocking. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science.* ACM, 25–36.

[8] Y. Zeng, J. Tan, and C.H. Xia. 2018. Fork and Join Queueing Networks with Heavy Tails: Scaling Dimension and Throughput Limit. *arXiv preprint arXiv:1805.05197* (2018).