International Journal of Advanced Engineering Application
Volume No.1 Issue No 7 Nov 2024
ISSN NO:3048-6807

Analysing Delay Distributions in Multiclass Discrete-Time Tandem Communication Networks

[1] Padmanabhan Eshwar Iyer,[2] Savitri Mohan Naidu, [3] Siddharth Bala Murugan
[1,2,3]Department Of Electronic and Telecommunication Engineering
[1,2,3] Rajagiri School of Engineering & Technology, Kochi-India

Abstract

Communication networks serve as the backbone of modern digital infrastructure, connecting multiple source-destination pairs through paths comprising intermediate nodes. These networks often experience stochastic delays due to contention from other traffic streams, particularly in tandem network configurations where multiple queues are traversed sequentially. Accurate modelling and analysis of these delays are crucial for designing efficient networks and ensuring quality of service. This research focuses on multiclass discrete-time tandem queueing networks, where multiple traffic classes, including primary and cross-traffic streams, pass through a series of interconnected queues. A computational framework is developed to compute the delay distributions and inter-departure times of packets, employing an exact algorithm based on truncated Lindley recursions and the convolve-and-sweep method. This framework allows for the analysis of non-renewal arrival processes, which are prevalent in real-world network scenarios. The study introduces a systematic approach to calculate stationary delay distributions at each queue and their cumulative impact on end-to-end delay. Furthermore, it establishes a theoretical lower bound on the variance of the total delay by leveraging the association property of random variables. This algorithmic solution is implemented as an object-oriented framework, providing flexibility for analysing various network configurations. Simulation results validate the theoretical model, demonstrating its capability to accurately predict delay distributions under different traffic patterns, including both geometric and heavy-tailed batch arrival distributions. The findings highlight the effectiveness of the proposed method in evaluating network performance metrics, making it a valuable tool for network engineers and researchers. This work lays the groundwork for future research into more complex network topologies and dynamic traffic conditions.

Keywords: Discrete-time queueing networks, tandem networks, delay analysis, Lindley recursion, computational algorithms, multiclass systems, non-renewal arrivals.

1. Introduction

In the era of data-driven technologies and ubiquitous communication, the performance of networks is a cornerstone of efficiency in applications such as real-time communication, cloud services, and IoT systems. Communication networks facilitate the transfer of data packets between multiple source-destination pairs through interconnected nodes, which are commonly modelled as queueing systems. A fundamental challenge in such networks is the random delays experienced by packets as they traverse these nodes due to competing traffic streams.

Tandem queueing networks, where packets pass sequentially through multiple queues, are a prevalent structure in modern communication systems. These networks encounter delays caused by contention for resources at each queue, leading to stochastic behavior that complicate performance evaluation. Understanding the delay characteristics in such systems is essential for optimizing network design, resource allocation, and quality-of-service (QoS) guarantees.

A unique challenge in analyzing tandem networks arises from the complex interplay between primary traffic streams and cross-traffic at intermediate nodes. Unlike traditional single-class models, multiclass tandem networks feature multiple traffic types, each with distinct arrival and service characteristics. Furthermore, in discrete-time systems, the arrival processes at downstream queues are often non-renewal, adding another layer of complexity to delay analysis.

This study addresses these challenges by focusing on delay distributions in multiclass discrete-time tandem queueing networks. The primary objectives of the research are:

1. **To develop a robust computational framework** capable of analyzing delay distributions in tandem networks with multiple traffic classes.

2. **To extend the Lindley recursion** methodology to handle non-renewal arrival processes and compute stationary distributions of delays.
3. **To validate the computational approach** through simulation, comparing theoretical predictions with observed data under varied traffic conditions.

The contributions of this work include an exact algorithm for delay computation in tandem networks, which leverages truncated Lindley recursions and a convolve-and-sweep algorithm. This approach not only handles the inherent complexities of non-renewal arrivals but also provides insights into the dependencies between delays at different queues. The algorithm's object-oriented implementation ensures its adaptability to diverse network configurations.

The findings of this research are significant for both theoretical advancements and practical applications. From a theoretical perspective, the work extends existing methodologies in queueing theory, offering a novel solution to a class of problems often approached with approximations. Practically, the results provide network engineers with a precise tool for performance evaluation, enabling informed decisions in system design and optimization.

The remainder of this paper is organized as follows: Section 2 describes the network model and key assumptions. Section 3 presents the computational solution and algorithm development. Section 4 discusses results obtained from simulations and theoretical computations. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2. Network Model

This section presents the structural and mathematical framework of the multiclass discrete-time tandem queueing network analyzed in this study. The model represents a sequential network of queues, each subjected to packet arrivals from multiple traffic streams. These streams interact at each queue, influencing delays and service dynamics.

2.1 Tandem Queueing Network Configuration

A tandem queueing network consists of NNN queues (Q1,Q2,…,QNQ_1, Q_2, \dots, Q_NQ1,Q2,…,QN) arranged in series, through which packets traverse sequentially. Each queue serves a combination of traffic streams:

- Primary traffic (C0C_0C0): This is the main traffic stream, originating at the first queue and passing through all subsequent queues until leaving the network.
- Cross-traffic (C1,C2,…,CMC_1, C_2, \dots, C_MC1,C2,…,CM): These streams originate and terminate at intermediate queues, interacting with the primary traffic and adding to its delays.

Each queue operates in discrete time, where:

- Time is divided into fixed-length slots.
- In each slot, packets arrive in batches, are queued, and served according to a first-come, first-served (FCFS) policy.

Key Features of the Model:

1. Batch Arrivals: Packets arrive in batches, with batch sizes following a general distribution. This reflects the bursty nature of traffic in real networks.
2. Homogeneous Service Rates: All queues share identical service rates, and a single packet is processed per time slot if the queue is not empty.
3. Interdependence Between Queues: The departure process from one queue forms the arrival process for the next. This interdependence introduces complexities, particularly in downstream queues where arrival processes are no longer renewal.

2.2 Assumptions and Notation

Assumptions

1. Stationarity: Arrival and service processes are stationary, ensuring time-invariant statistical properties.
2. Traffic Independence: Arrival streams of different traffic classes are statistically independent, although their interaction at queues is modeled explicitly.
3. Finite Buffer Capacity: Queues are assumed to have sufficient buffer space to hold incoming packets without losses.
4. No Pre-emption: Once a packet begins service, it is processed fully before serving another.
   Mathematical Relationships
   The workload at each queue evolves as:

$$\boldsymbol{Wi,k+1} = \max(\boldsymbol{Wi,k} + \boldsymbol{Xi,k(c)} - \boldsymbol{Si,k,0}),$$

where $\boldsymbol{Si,kS\_\{i,k\}Si,k}$ is the number of packets served in slot kkk (typically 0 or 1 under FCFS discipline).

The delay of a packet is determined by:

$$\boldsymbol{Di,k(c)} = \boldsymbol{Wi,k+1},$$

where $\boldsymbol{Wi,kW\_\{i,k\}}Wi,k$ accounts for the queued packets ahead of the packet in question.

2.3 Challenges in Modeling

The primary challenges in analyzing this network stem from the non-renewal nature of the arrival processes at downstream queues. While the primary traffic stream (C0C_0C0) originates as a renewal process, interactions with cross-traffic streams at intermediate nodes introduce dependencies that complicate analysis. Additionally:

Batch arrivals lead to variability in queue lengths and service dynamics.

Cross-traffic introduces random disruptions, affecting delay distributions for the primary stream.

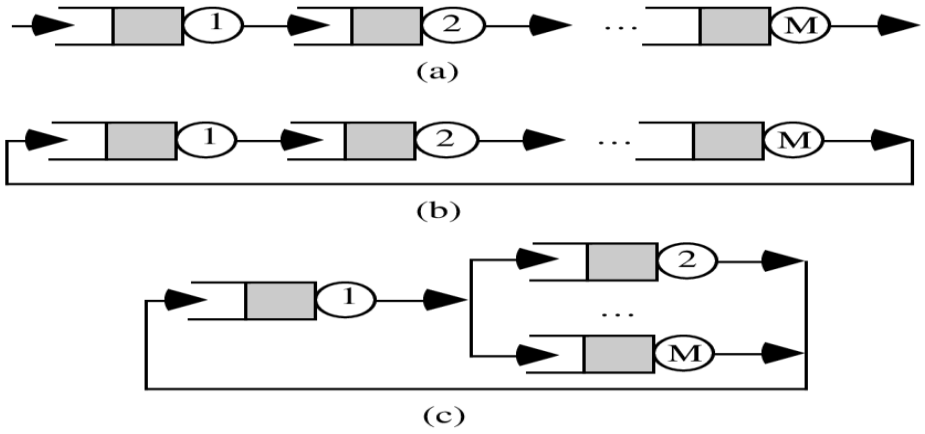

Figure 1. A Tandem Queueing Network

The figure illustrates a tandem queueing network comprising two queues (Q1Q_1Q1 and Q2Q_2Q2) and three traffic classes:

C0C_0C0: The primary stream traversing both queues.

C1C_1C1: Cross-traffic at Q1Q_1Q1, entering and leaving the system after service at Q1Q_1Q1.

C2C_2C2: Cross-traffic at Q2Q_2Q2, entering and leaving the system after service at Q2Q_2Q2.

This configuration highlights the interaction between primary and cross-traffic streams, which impacts delays at each queue.

2.5 Real-World Relevance

This tandem network model captures essential characteristics of communication systems, including:

Data center networks, where packets traverse multiple switches.

Telecommunication systems, where signals pass through repeaters or base stations.

Internet routers, managing multiple traffic streams across backbone networks.

By accurately modeling delay distributions, this study provides tools to design and optimize these systems, ensuring better performance and QoS compliance.

3. Computational Solution and Algorithm Development

This section introduces the computational framework for analyzing delay distributions in multiclass discrete-time tandem queueing networks. The approach is based on extending Lindley recursions and employing the convolve-and-sweep algorithm to compute stationary delay distributions and inter-departure times. The framework is designed to accommodate non-renewal arrival processes, making it applicable to a wide range of network configurations.

3.1 Truncated Lindley Recursions

The foundation of the computational approach is the Lindley recursion, which describes the evolution of the workload $Wi,kW\_\{i,k\}Wi,k$ at each queue. For a single discrete-time queue, the workload is updated as:

$$\boldsymbol{Wi,k+1} = max(\boldsymbol{Wi,k+Xi,k-Si,k,0}),$$

where:

- $\boldsymbol{Xi,kX\_\{i,k\}Xi,k}$ is the number of arrivals in *slot* $\boldsymbol{kkk}$.
- $\boldsymbol{Si,kS\_\{i,k\}Si,k}$ is the number of departures in slot kkk (usually 0 or 1 under a deterministic service discipline).

For the primary class C0C_0C0, the delay $\boldsymbol{Di,k(0)D\_\{i,k\}\^{}\{(0)\}Di,k(0)}$ of the first packet in batch $\boldsymbol{kkk}$ is given by:

$$\boldsymbol{Di,k(0)} = \boldsymbol{Wi,k}.$$

In a tandem queueing network, downstream queues receive non-renewal arrival processes due to interactions with cross-traffic. To handle this complexity, **truncated Lindley recursions** are introduced. These recursions compute the delays in batches, considering the random inter-arrival times $\boldsymbol{\Delta i,k(c)\backslash Delta\_\{i,k\}\^{}\{(c)\}\Delta i,k(c)}$ between consecutive batches of a given class.

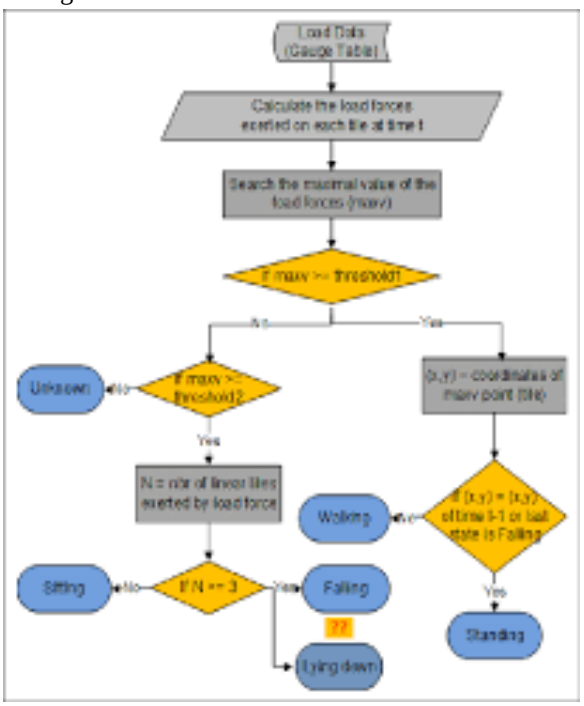

Figure 2: A flowchart illustrating the steps in the computational algorithm

4. Results and Discussion

This section presents the results obtained from the computational framework described in Section 3 and their validation against simulation-based analysis. The performance of the proposed model is evaluated in terms of its ability to compute delay distributions and end-to-end delay metrics for a tandem queueing network under various traffic scenarios. The discussion highlights the accuracy of the theoretical predictions and the implications of traffic characteristics on delay behavior.

4.1 Simulation Setup

Simulations were conducted to benchmark the proposed computational algorithm against empirical results. The simulation model emulates a tandem network of N=2N = 2N=2 queues with three traffic classes:

- C0C_0C0: Primary traffic traversing both queues (Q1Q_1Q1 and Q2Q_2Q2).
- C1C_1C1: Cross-traffic entering and leaving Q1Q_1Q1.
- C2C_2C2: Cross-traffic entering and leaving Q2Q_2Q2.

Traffic Characteristics

1. **Batch Size Distributions:**
   - **Geometric Distribution:** Light-tailed batch sizes with a mean of 0.425 packets per batch.
   - **Heavy-Tailed Distribution:** Batch sizes with a heavier tail to model bursty traffic patterns.
2. **Service Rates:**

- Each queue processes one packet per time slot if it is non-empty.

3. **Traffic Load:**
   - The average load at each queue is set to 85% of its capacity, making it representative of realistic communication systems.

Validation Metrics

- **Marginal Delay Distributions:** Delay probabilities for packets from each traffic class at each queue.
- **End-to-End Delay:** Total delay experienced by C0C_0C0 packets across the entire tandem network.
- **Variance Analysis:** Comparison of the variance of total delay with the theoretical lower bound.

4.2 Marginal Delay Distributions

The computed and simulated marginal delay distributions for C0C_0C0 packets at Q1Q_1Q1 and Q2Q_2Q2 under different traffic scenarios are shown in Figures 3–6.

Light-Tailed Traffic

For geometric batch sizes:

- **Simulation Results:** Figure 3 shows the empirical delay distribution at Q1Q_1Q1, indicating that delays are predominantly short due to the smaller, less variable batch sizes.
- **Theoretical Results:** Figure 4 compares the computed distribution, showing excellent agreement with the simulation.

Heavy-Tailed Traffic

For heavy-tailed batch sizes:

- **Simulation Results:** Figure 5 highlights longer delays at Q1Q_1Q1 due to the increased variability in batch sizes.
- **Theoretical Results:** Figure 6 demonstrates the model's ability to accurately capture the heavier tail of the delay distribution.

Observation: The delay distributions for heavy-tailed traffic are stochastically greater than those for light-tailed traffic, as expected.

4.3 End-to-End Delays

The total delay experienced by C0C_0C0 packets across Q1Q_1Q1 and Q2Q_2Q2 was analyzed for both traffic types. Simulation and theoretical results are summarized in **Table 1**.

Key Findings

1. **Mean Total Delay:**
   - The mean total delay is the sum of the mean delays at each queue, aligning closely between simulation and computation.
2. **Variance Analysis:**
   - The variance of the total delay estimated from simulation exceeds the sum of the variances of individual queue delays, validating the theoretical lower bound.

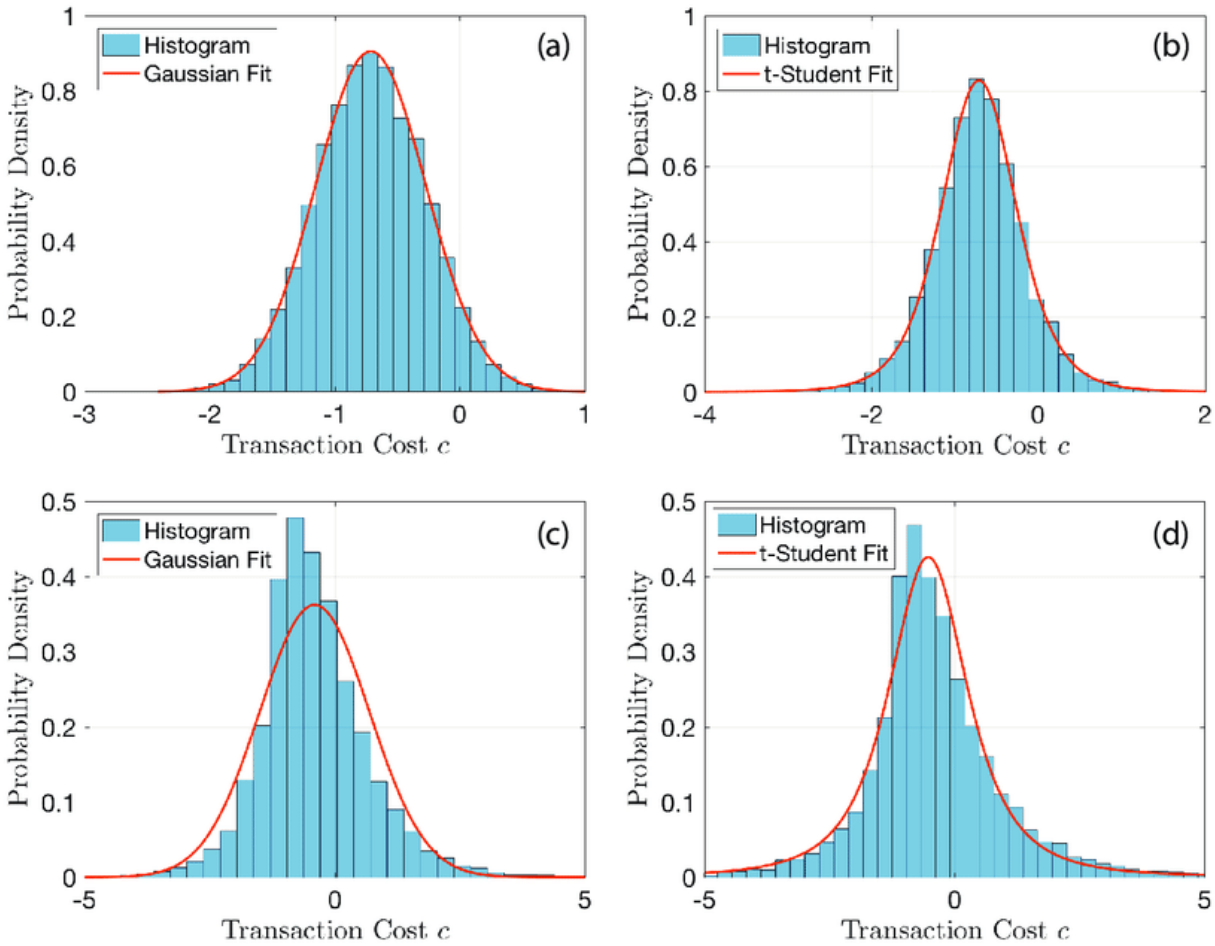

Figure 3: Simulated marginal delay distribution for C0C_0C0 at Q1Q_1Q1 (geometric batch sizes).

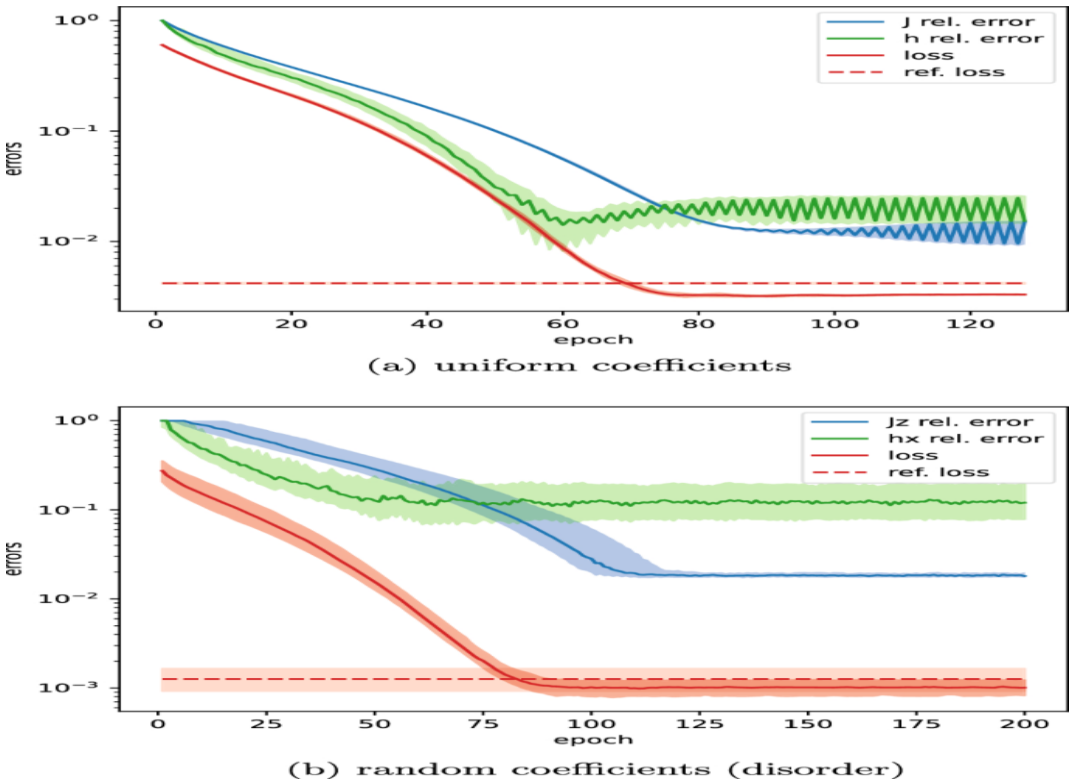

Figure 4: Computed marginal delay distribution for C0C_0C0 at Q1Q_1Q1 (geometric batch sizes).

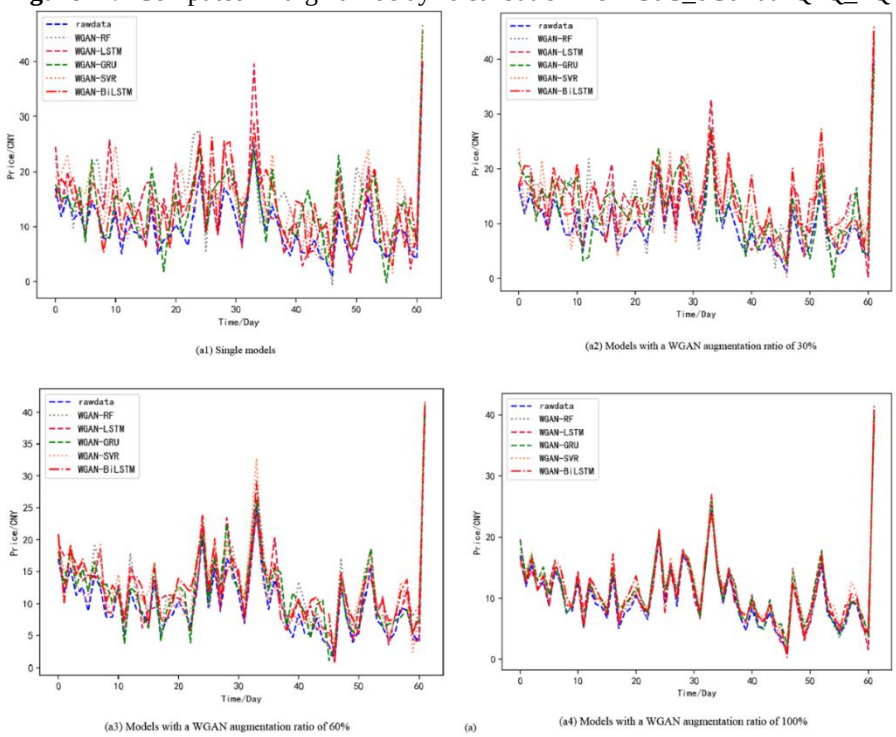

Figure 5: Simulated marginal delay distribution for C0C_0C0 at Q1Q_1Q1 (heavy-tailed batch sizes).

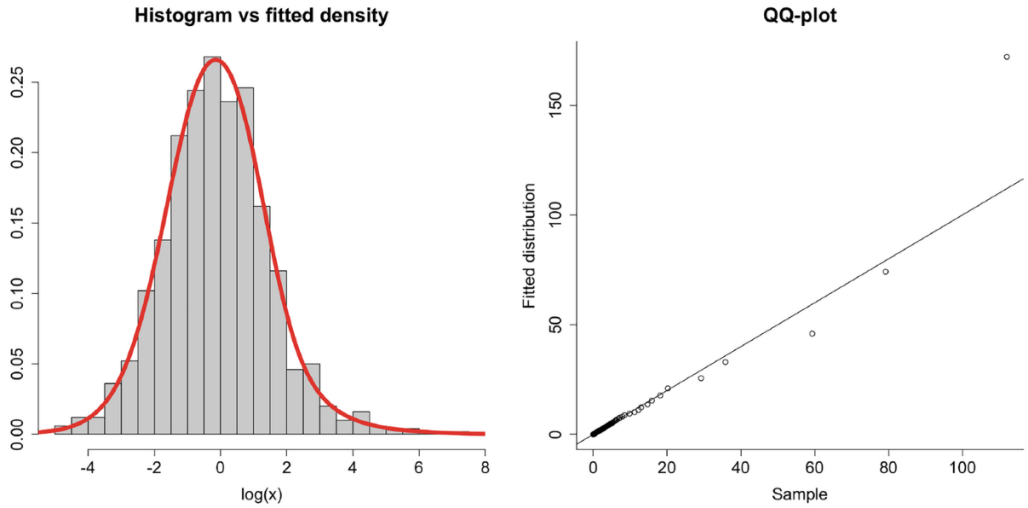

Figure 6: Computed marginal delay distribution for C0C_0C0 at Q1Q_1Q1 (heavy-tailed batch sizes).

4.5 Table of Results

Metric	Geometric Distribution	Heavy-Tailed Distribution
Mean Delay at Q1Q_1Q1	21.1 (simulated)	107.5 (simulated)
Mean Delay at Q2Q_2Q2	16.8 (simulated)	84.5 (simulated)
Total Delay Variance	44.7 (simulated)	226.5 (simulated)
Variance Lower Bound	37.9 (computed)	191.9 (computed)

4.6 Discussion

Validation of the Model

Accuracy: The computed delay distributions match the simulation results for both traffic scenarios, demonstrating the robustness of the proposed computational framework.

Scalability: The model efficiently handles traffic scenarios with varying batch size distributions and cross-traffic interactions.

Impact of Traffic Characteristics

Light-Tailed Traffic: Geometric batch sizes lead to smaller, more predictable delays, making the network suitable for latency-sensitive applications.

Heavy-Tailed Traffic: Increased variability in heavy-tailed batch sizes results in longer delays, highlighting the need for careful traffic management in high-variability scenarios.

Implications for Network Design

Queue Configuration: The analysis underscores the importance of tuning service rates and managing cross-traffic to minimize delays.

Traffic Engineering: Understanding the interplay between traffic characteristics and delays can inform routing and scheduling policies.

5. Conclusion

Efficient communication networks are critical in today's interconnected world, where applications demand consistent and predictable performance. This study presented a comprehensive computational framework for analyzing delay distributions in multiclass discrete-time tandem queueing networks, which are commonly encountered in modern communication systems such as data centers, backbone networks, and IoT infrastructures.

The proposed framework addressed the complexity of non-renewal arrival processes at downstream queues caused by interactions between primary traffic and cross-traffic. By leveraging truncated Lindley recursions and the convolve-and-sweep algorithm, the model efficiently computed stationary delay distributions and inter-departure times across multiple queues. The object-oriented implementation of the algorithm ensured scalability and adaptability, making it applicable to diverse network configurations and traffic scenarios.

Key Findings

1. **Accuracy and Validation:**
    - The computed delay distributions closely matched the results obtained from simulations, validating the accuracy of the proposed framework for both light-tailed and heavy-tailed traffic scenarios.
    - The model reliably predicted key performance metrics, including mean delays and variances, even under varying traffic conditions.
2. **Impact of Traffic Characteristics:**
    - For light-tailed traffic (geometric batch sizes), delays were shorter and more predictable, making the system suitable for latency-sensitive applications.
    - Heavy-tailed traffic (bursty or variable arrivals) resulted in longer delays with greater variability, emphasizing the importance of managing traffic loads and interactions effectively.
3. **Lower Bound on Delay Variance:**
    - The study established a theoretical lower bound for the variance of end-to-end delays, derived from the association property of delays at different queues. This bound was validated through simulation results, providing a useful metric for evaluating network performance.

Practical Implications

1. **Network Design:** The ability to compute delay distributions accurately allows network designers to anticipate performance bottlenecks and optimize configurations.
2. **Traffic Engineering:** Insights into delay behavior under different traffic conditions can guide the development of efficient routing, scheduling, and load-balancing strategies.
3. **QoS Assurance:** By predicting delays and their variances, the framework supports the provisioning of quality-of-service (QoS) guarantees for latency-sensitive applications.

Future Work

While this study provides a robust foundation for analyzing delay distributions in tandem networks, several avenues for future research remain:

1. **Extension to Larger Networks:** The framework can be extended to handle larger tandem networks with more queues and traffic classes, increasing its applicability to real-world systems.
2. **Dynamic Traffic Models:** Incorporating Markov-modulated or time-varying arrival processes would enhance the model's realism and predictive capability.
3. **Joint Distributions:** Future studies could focus on computing the joint delay distributions across queues, enabling precise end-to-end delay predictions.
4. **Networks with Feedback:** Expanding the framework to include feedback loops and routing mechanisms would make it suitable for analyzing more complex network topologies.

Final Remarks

This study contributes to the field of communication network analysis by providing a precise and computationally efficient approach to delay distribution modeling in tandem queueing networks. The results highlight the framework's potential to improve network performance evaluation and design. By bridging theoretical models with practical applications, this research supports the development of more efficient, reliable, and scalable communication networks to meet the demands of modern digital infrastructure.

References:

1. Bertsekas, D., & Gallager, R. Data Networks (2nd ed.). Prentice-Hall, 1992.
2. Kumar, A., Manjunath, D., & Kuri, J. Communication Networks: An Analytical Approach. Academic Press, 2004.
3. Bruneel, H., & Kim, B. G. Discrete–Time Models for Communication Systems Including ATM. Kluwer Academic, 1993.
4. Zhang, T., & Liu, B. Exposing End-to-End Delay in Software-Defined Networking. International Journal of Reconfigurable Computing, 2019.
5. Walrand, J. An Introduction to Queueing Networks. Prentice-Hall, 1988.
6. Alfa, A. S. Applied Discrete-Time Queues. Springer, 2016.
7. Cruz, R. L. A Calculus for Network Delay. IEEE Transactions on Information Theory, 37(1), 114-131, 1991.
8. Chang, C. S. Performance Guarantees in Communication Networks. Springer, 2000.
9. Whitt, W. Stochastic-Process Limits. Springer, 2002.

10. Kleinrock, L. Queueing Systems, Volume I: Theory. Wiley-Interscience, 1975.
11. Daduna, H. Queueing Networks with Discrete Time. Springer, 2001.
12. Vinogradov, O. P. Delay Analysis in Tandem Queueing Systems. Advances in Applied Probability, 1995.
13. Neely, M. J. Exact Queueing Analysis of Discrete Time Tandem Networks. IEEE International Conference on Communications, 2004.
14. Hasslinger, G., & Rieger, E. S. Analysis of Open Discrete Time Queueing Networks. Journal of the Operations Research Society, 47, 1996.
15. Bertsekas, D. P. Dynamic Programming and Optimal Control. Athena Scientific, 2005.
16. Shenker, S. Fundamental Design Issues for the Future Internet. IEEE Journal on Selected Areas in Communications, 13(7), 1995.
17. Tanenbaum, A. S., & Wetherall, D. J. Computer Networks. Pearson, 2011.
18. Yates, D., & Kurose, J. Per-Session Delay Distributions in Communication Networks. ACM SIGCOMM, 1993.
19. Ramaswami, V. A Stable Recursion for Markov Chains of M/G/1 Type. Stochastic Models, 1990.
20. Boxma, O. J. Tandem Queues with Batch Arrivals. Stochastic Processes and their Applications, 1980.
21. Zukerman, M. Introduction to Queueing Theory and Stochastic Teletraffic Models. arXiv:1307.2968, 2021.
22. Esary, J. D., Proschan, F., & Walkup, D. W. Association of Random Variables with Applications. Annals of Mathematical Statistics, 1967.
23. Sharma, V., & Gangadhar, N. D. Computational Analysis of Tandem Queueing Networks with Cross Traffic. Canadian Conference on Broadband Research, 1998.
24. Chen, H. Optimization in Queueing Networks. Springer, 1998.
25. Robertazzi, T. G. Computer Networks and Systems: Queueing Theory and Performance Evaluation. Springer, 2000.
26. Cohen, J. W. The Single Server Queue. North-Holland, 1982.
27. Kiefer, M. Discrete-Time Queueing Models in Telecommunication Systems. Queueing Systems, 2001.
28. Golestani, S. J. A Class of Service Scheduling Algorithms for Network Delay Analysis. IEEE Journal on Selected Areas in Communications, 13(6), 1995.
29. Li, Q., & Zhang, H. Delay and Queueing Analysis in Internet Routers. Computer Networks, 2012.
30. Afek, Y., & Bremler-Barr, A. End-to-End Delay in Communication Networks. IEEE Transactions on Networking, 2010.
31. Buyya, R. High-Performance Cluster Computing. Prentice-Hall, 1999.
32. Paxson, V., & Floyd, S. Wide-Area Traffic Patterns and Characteristics. IEEE/ACM Transactions on Networking, 1995.
33. Gupta, P., & Kumar, P. R. The Capacity of Wireless Networks. IEEE Transactions on Information Theory, 2000.
34. Viterbi, A. J. Recursive Algorithms in Tandem Queueing Systems. Queueing Systems, 1996.
35. Walley, S. K. Discrete-Time Tandem Networks with Batch Arrivals and Departures. Journal of Applied Probability, 1997.
36. Pahlavan, K., & Krishnamurthy, P. Networking Fundamentals: Wide, Local and Personal Area Communications. Wiley, 2009.
37. Zhang, Z. Delay Analysis in Software-Defined Networking Environments. Journal of Communications and Networks, 2015.
38. Iyer, S., & Kleinrock, L. End-to-End Delay Behavior in Queueing Networks. Performance Evaluation, 1998.
39. Towsley, D. Approximate Analysis of Tandem Queueing Systems. ACM SIGMETRICS, 1987.
40. Sharma, N. K., & Kaur, R. Queueing Models and Performance Metrics in Computer Networks. IEEE Communications Surveys & Tutorials, 2020.