

基于鞅论的时延性能分析

于宝珠, 焦子阳*

沈阳理工大学信息科学与工程学院, 辽宁 沈阳

收稿日期: 2024年1月23日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

摘要

为了更加精确地评估网络性能, 本文引入了一种新的理论方法——鞅论, 构建网络时延性能分析框架。通过鞅论建模时延, 能够实现更加准确对网络性能的分析评估。在通信网络中, 考虑单一业务到达通信系统的场景。在常速率服务机制下, 分别建模泊松业务流以及马尔可夫业务流, 基于鞅论进一步推导了时延违反概率。最后, 通过设置不同的到达方式、服务参数, 探索影响时延违反概率的因素, 利用MATLAB仿真进行验证分析。

关键词

鞅论, 排队论, 时延违反概率

Delay Performance Analysis Based on Martingale Theory

Baozhu Yu, Ziyang Jiao*

College of Information Science and Engineering, Shenyang Ligong University, Shenyang Liaoning

Received: Jan. 23rd, 2024; accepted: Feb. 22nd, 2024; published: Feb. 29th, 2024

Abstract

In order to evaluate network performance more accurately, this paper introduces a new theoretical method, martingale theory, to construct a network delay performance analysis framework. Through martingale theory and modeling delay, more accurate analysis and evaluation of network performance can be realized. In the communication network, the scenario in which a single service arrives at the communication system is considered. Under the constant-rate service mechanism, the Poisson service flow and Markov service flow are modeled respectively, and the delay violation probability is further deduced based on the martingale theory. Finally, by setting different

*通讯作者。

文章引用: 于宝珠, 焦子阳. 基于鞅论的时延性能分析[J]. 计算机科学与应用, 2024, 14(2): 341-349.

DOI: 10.12677/csa.2024.142035

arrival modes and service parameters, the factors affecting the delay violation probability were explored, and MATLAB simulation was used for verification and analysis.

Keywords

Martingale Theory, Queuing Theory, Delay Violation Probability

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着通信网络技术的发展, 通信系统在日常生活中扮演着越来越重要的角色。从电话、互联网到移动通信, 通信技术的进步使得信息传输速度和质量得到了极大的提升。然而, 随着各种应用对实时性、可靠性和带宽的需求不断增加, 通信系统的时延问题逐渐成为了一个亟待解决的问题。时延要求对于通信系统的性能有着重要的影响, 尤其是在语音通话、视频会议、在线游戏等实时性要求较高的应用场景中。因此, 对于 QoS (Quality of service) 要求较严格的业务来说, 在系统性 QoS 时延保障的背景下, 对于时延精确的建模分析, 是实现网络性能评估、业务 QoS 保障、合理资源分配以及提高网络资源利用率的关键。在理论上对网络时延性能进行分析的难点可以总结如下: ① 网络中具有时延 QoS 要求的业务生成的数据包具有复杂的随机特性; ② 无线接入网为终端提供的接入机制在数学上也只能用随机过程来描述; ③ 基于排队论建模终端数据包接入, 时延是业务的到达过程和接入节点的服务过程相互纠缠而成的产物, 两种异质随机过程的相互作用在数学上难以分析。现有的分析时延 QoS 的理论方法, 例如排队论, 只能实现对到达、服务均为简单随机过程的排队系统时延性能进行分析。有效容量有效带宽理论提供了一个有效的分析时延超过某一阈值的概率(时延违反概率)的方法, 然而, 这个框架存在小队长估计误差大的问题, 针对具有突发性到达的排队系统, 难以提供一个精确的时延性能分析理论结果。鞅论是 QoS 理论系统中的新兴工具, 基于鞅论得到的时延违反概率理论上界比已有的结论更加精确。因此, 本文基于鞅论研究时延的建模与系统性能分析问题, 在常速率服务机制下, 探讨具有不同随机特性到达过程对网络时延性能的影响。

文献[1]给出了一种基于指数上鞅端到端时延表达式, 分析可实现大规模网络的 QoS 控制, 运用统计网络演算理论中最小加代数的卷积运算规则计算端到端时延界, 该时延界不仅可以线性扩展, 而且数值分析结果表明, 在相同假设条件下, 该时延界比现有的线性时延界具有更好的紧致性。文献[2]基于室内 VLC 系统建立了多业务到达的网络排队模型, 为了更加贴近实际场景中的突发状态, 采用马尔可夫达到过程, 服务采用多接受能力的 ALOHA 随机接入系统, 实验仿真表明基于鞅理论的时延违反概率相对于其他传统方法来说更加贴近真实值。文献[3]研究背景以高速数据网为例, 在推导出单一业务到达网络的有效带宽与系统队列溢出概率的基础上, 进一步地以多业务到达网络场景中, 对网络有效带宽与队列溢出概率进行了分析。文献[4]提出了在保障时延 QoS 要求下的 SFC 逐跳粒度的带宽分配框架。针对异质业务到达的场景将其建模成复杂的马尔可夫过程, 根据有效带宽和有效容量理论的指导下, 保证了异质业务不同的 QoS 要求。

文献[5] Yuming Jiang 等人开始利用随机网络微积分理论对系统端到端时延界进行了分析, 使用统计网络中的最小加代数卷积运算, 得出了一种相对更加紧凑的性能界限。对于随机网络微积分中对系统的

到达和服务建模方法有一定的特殊性, 并不能普遍地使用。文献[6] M Roughan、KAngrishi 等人通过构建较为复杂的排队系统队列流量模型, 通过使用鞅理论的停时定理以及鞅理论的不等式性质, 解决分析系统的性能界限问题。文献[7] F. Ciucu、F. Poloczek 等人提出了到达鞅、服务鞅的概念, 利用全新的视角——鞅论建模并分析时延 QoS 的性能。通过仿真分析得到的时延界比有效带宽和有效容量理论更加精确, 从而有效地得到更加准确的带宽需求, 从而提高了频带利用率。文献[8] Gan Luan 利用鞅理论的方法, 通过分析数据中心网络的队列溢出, 从模型中获得了队列溢出概率与队列溢出的平均时间。文献[9]研究了一种新的分析聚合流量时延违反概率界的方法, 基于鞅理论, 构造出一个超鞅过程, 使得对具有多个到达和单个服务器的排队系统进行时延分析成为可能。用其对应的超鞅的参数来表征聚集流量的延迟。基于鞅的停止时间理论, 推导出了时滞违反概率界, 与仿真结果吻合较好, 且紧。文献[10]研究了在统计型时延的约束条件下, 随机补偿切片方案的性能。提出了一种基于鞅理论的方法来研究随机切片方案并获得高性能的时延界。文献[11]为了满足 5G 通信业务对时延可靠性的严格要求, 网络功能虚拟化和软件定义网络提供了新的解决方案。该论文基础鞅论给出一种精确的时延可靠性分析框架, 并且提出一种可以同时满足业务请求 QoS 约束以及更高系统可靠性要求的系统可靠性设计方案。

2. 系统建模

排队论是对通信网内的流量进行分析的数学工具。在通信时间内, 对用户提供良好的服务, 需要有足够的线路(信道)容量。线路容量不够, 将导致拥塞现象, 在电路域中, 就意味着增加业务量损失在分组域中, 就意味着增加信息延时。

本文将单一业务流的传输过程建模成一个排队系统, 系统建模如下(图 1):



Figure 1. A queuing model for the arrival of a single service
图 1. 单一业务到达的排队模型

在排队系统模型中, 业务的到达用 $a(n)$ 来表示, 这个过程可以用一个概率分布的描述随机过程。本文选择常速率服务机制, 服务过程用来 $s(n)$ 表示。因此在单一业务到达的情况下, 可以用 $A(n)$ 来表示 n 时刻的累计到达, 即 $A(n) = \sum_{i=0}^n a(i)$ 以用 $S(n)$ 来表示 n 时刻的累计服务过程, $S(n) = \sum_{i=0}^n s(i)$ 。

当到达流为马尔科夫过程时, 建模如下(图 2):

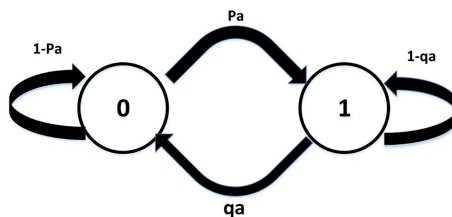


Figure 2. MMOO model
图 2. MMOO 模型

对于马尔可夫到达过程, 0 状态代表的为关闭状态, 此时并没有进行数据包传送。在 1 状态代表的为打开状态, 此时进行数据包传送。0 状态转移到 1 状态的概率为 p_a , 保留在 0 状态的概率为 $1-p_a$ 。1 转移到 0 状态的概率为 q_a , 保留 1 状态的概率为 $1-q_a$ 。

到达过程和服务过程在鞅域进行建模的实质就是对其到达和服务过程进行变换处理, 使之变为特定的鞅形式。因此, 本文将采取指数上鞅结构对到达业务和服务业务进行建模。

到达鞅: 将到达过程进行鞅变换处理, 能够得到到达鞅过程——到达鞅:

$$M_i^a(n; h_{a_i}(a_i(n)), \vartheta, K a_i) = h_{a_i}(a_i(n)) e^{\vartheta(A_i(n) - n K a_i)} \quad (1)$$

若对于每个 $\vartheta > 0$, 都有一个 $K a_i > 0$ 成立, 那么式(1)是一个指数上鞅。其中 $a_i(n)$ 是终端 i 在时隙 n 的到达, $A_i(n)$ 为 n 时隙的累积到达; $h_{a_i}(a_i(n))$ 是终端 i 到达过程的相关特性函数, 体现业务随机变量间的相关性; $K a_i$ 代表的是在保持到达鞅的上鞅性质; ϑ 为到达鞅的参量。不同业务的随机特性决定其 $h_{a_i}(a_i(n))$ 与 $K a_i$ 具体形式。

服务鞅: 将服务过程进行鞅变换处理, 能够得到服务鞅过程——服务鞅:

$$M_i^s(n; h_s(s(n)), \vartheta, K s) = h_s(s(n)) e^{\vartheta(n K s - S(n))} \quad (2)$$

服务鞅与到达鞅类似, 即对于每一个 $\vartheta > 0$, 都有一个 $K s > 0$ 成立时, 此时的(2)式为一个指数上鞅的过程。其中有 $s(n)$ 为时隙 n 的服务, $S(n)$ 来表示 n 时刻的累计服务过程, $h_s(s(n))$ 为服务过程的相关特性函数。 ϑ 为到达鞅的参量。不同业务的随机特性决定其 $h_s(s(n))$ 与 $K s$ 具体形式。

3. 到达鞅和服务鞅推导过程

对于到达过程为马尔可夫调制的随机过程时, 即累计到达为 $A(n) = \sum_{i=0}^n a(n)$, 此时为了保证到达鞅是指数上鞅的结构, 需要保证下式成立:

$$E \left[h_{a_i}(x_{n+1}) e^{\vartheta(A_i(n+1) - (n+1) K a_i)} \middle| x_1, x_2, \dots, x_n \right] = h_{a_i}(x_n) e^{\vartheta(A_i(n) - n K a_i)} \quad (3)$$

在马尔可夫到达过程中, 还需要定义一个矩阵 T 的指数列变换, 可以写为 T^ϑ :

$$T_{i,j}^\vartheta = T_{i,j} e^{\vartheta f(x_j)} \quad (4)$$

即矩阵中每一个元素都需要乘以对应的列进而构成指数函数。在上式中 $f(\cdot)$ 对应其平均到达速率。

在矩阵 T^ϑ 中 $sp(T^\vartheta)$ 表示为最大特征值, 即为谱半径。 h 对应矩阵 T^ϑ 中的特征向量, 即特征值和特征向量都为正值。因此:

$$\begin{aligned} & E \left[h_{a_i}(x_{n+1}) e^{\vartheta(A_i(n+1) - (n+1) K a_i)} \middle| x_1, x_2, \dots, x_n \right] \\ &= e^{\vartheta(A_i(n) - n K a_i)} E \left[h_{a_i}(x_{n+1}) e^{\vartheta(x_{n+1} - K a_i)} \middle| x_n \right] (a) \\ &= e^{\vartheta(A_i(n) - n K a_i)} E \left[h_{a_i}(x_{n+1}) e^{\vartheta(x_{n+1})} \middle| x_n \right] e^{-\vartheta K a_i} (b) \\ &= e^{\vartheta(A_i(n) - n K a_i)} (T^\vartheta h)(x_n) e^{-\vartheta K a_i} (d) \\ &\leq h_{a_i}(x_n) e^{\vartheta(A_i(n) - n K a_i)} (e) \end{aligned} \quad (5)$$

对于常速率服务过程, 即累计服务为 $S(n) = \sum_{i=0}^n s(n)$, 此时服务的鞅构造式为:

$$\begin{aligned}
& M_i^s(n; h_s(s(n)), \mathcal{G}, K_s) \\
&= h_s(s(n)) e^{\mathcal{G}(nK_s - s(n))} \\
&= e^{\mathcal{G}(nC - s(n))}
\end{aligned} \tag{6}$$

根据参数 \mathcal{G} 定义, \mathcal{G}^* 为 $K_{a_1} = K_s$ 的取值解。对于马尔可夫到达过程以及常速率服务机制下有:

$$K_{a_1} = \log sp[T^{\mathcal{G}}] / \mathcal{G} \tag{7}$$

$$K_s = C \tag{8}$$

对于泊松到达过程即累计到达为 $A(n) = \sum_{i=0}^n a(n)$, 此时为了保证到达鞅是指数上鞅的结构, 则到达鞅过程构造如下:

$$\begin{aligned}
& M_i^a(n; h_{a_1}(a_1(n)), \mathcal{G}, Ka_1) \\
&= h_{a_1}(a_1(n)) e^{\mathcal{G}(A_1(n) - n\lambda_p(e^{\mathcal{G}} - 1)) / \mathcal{G}} \\
&= e^{\mathcal{G}(A_1(n) - n\lambda_p(e^{\mathcal{G}} - 1)) / \mathcal{G}}
\end{aligned} \tag{9}$$

根据鞅参量 \mathcal{G}^* 定义, \mathcal{G}^* 为 $K_{a_2} = K_s$ 的取值解。即:

$$K_{a_2} = \frac{\lambda_p(e^{\mathcal{G}} - 1)}{\mathcal{G}} \tag{10}$$

$$\frac{\lambda_p(e^{\mathcal{G}} - 1)}{\mathcal{G}} = C \tag{11}$$

有时延违反概率不等式:

$$P(W \geq k) \leq \frac{E[h_a(0)]E[h_s(s(0))]}{H} e^{-\mathcal{G}^* \lambda k} \tag{12}$$

其中 $E[h_s(0)] = 1$, 对于马尔可夫到达过程, $E[h_a(0)] = h_{a_0}(0)\pi_0 + h_{a_1}(1)\pi_1$ 。泊松到达过程时, $E[h_a(0)] = 1$ 。

4. 仿真分析

为了验证算法的准确性, 本文在建模排队系统的到达过程和服务过程, 利用鞅理论对时延违反概率进行了计算。

在马尔可夫到达过程中, 给定实验参数设置如下: $p_a = 0.2$, $q_a = 0.3$, $R = 2$ packets/slot, $P = 0.7$, 设置不同值 $C = 10$ 、 $C = 100$ 。分析其对时延违反概率的影响, 如图 3 所示。

由系统仿真可以得出, 在给定到达过程中参数一定的条件下, 通过改变不同大小的服务机制。设置服务机制参数 $C = 10$ 、 $C = 20$, 由仿真得出, 在同一时间下, 对于较大的服务机制而言, 目标时延违反概率相对更加严格。对于不同的服务机制参数下, 时延违反概率都会随着时间阈值的增加而减小。

给定实验参数设置如下: $p_a = 0.2$, $q_a = 0.3$, $C = 10$, $P = 0.7$, 设置不同值 $R = 10$ packets/slot、 $R = 20$ packets/slot 分析其对时延违反概率的影响, 如图 4 所示。

对于恒定的速率 $C = 10$ 服务机制下, 设置不同参数的到达强度, 由系统仿真可以明显的得出在同一时刻, 到达强度越大的情况下, 时延违反概率值越大。在同一时刻的条件下, 时延违反概率会随着到达强度的增加而增加, 时延违反概率都会随着时间阈值的增大而减小。

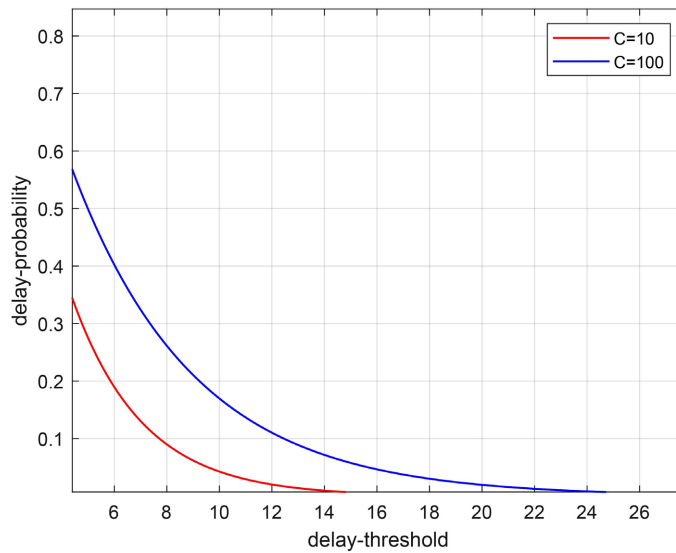


Figure 3. The influence of the non-constant rate service mechanism on the probability of Markov arrival delay violation

图 3. 不同常速率服务机制下对马尔科夫到达时延违反概率的影响

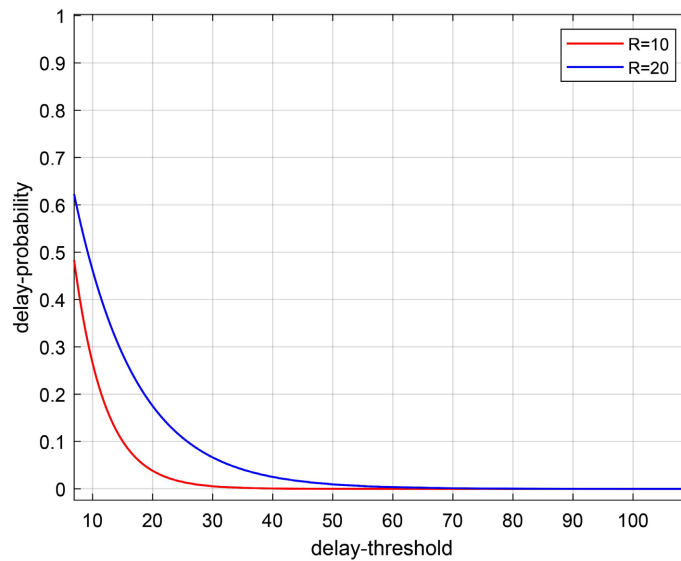


Figure 4. The influence of different Markov arrival intensity mechanisms on the probability of delay violation

图 4. 不同马尔可夫到达强度机制下对时延违反概率的影响

给定实验参数设置如下： $p_a = 0.2$ ， $q_a = 0.3$ ， $C = 10$ ， $P = 0.7$ ，设置不同负载率的值 $\rho = 0.5$ 、 $\rho = 0.9$ 分析其对时延违反概率的影响，如图 5 所示。

对于给定的速率 $C = 10$ 服务机制下，通过分别设置不同的负载率 $\rho = 0.5$ 、 $\rho = 0.9$ 的条件下。由系统仿真可以得出，时延违反概率随着时间阈值的增加而减小。对于负载率较大的系统，时延违反概率要求相对更为宽松，而负载率小的系统时延违反概率更加严格。

在泊松到达过程时，给定参数 $\lambda = 1$ 、 $P = 1$ 、 $h_a = 1$ 、 $h_s = 1$ ，设置不同的常数 $C = 5$ 、 $C = 10$ 分析其对时延违反概率的影响，如图 6 所示。

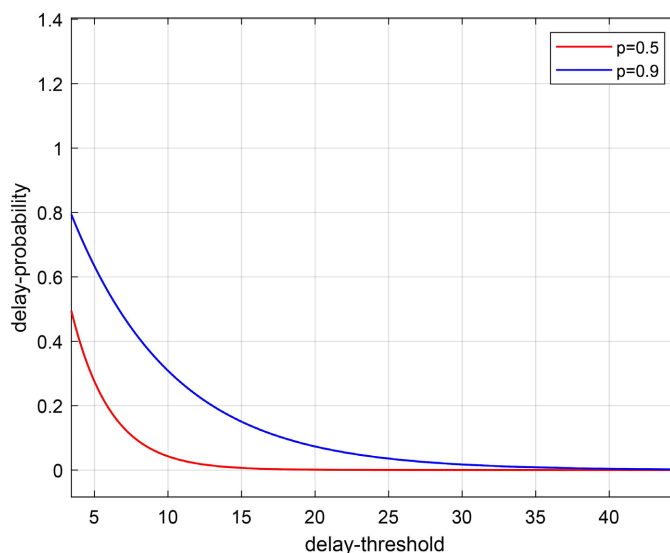


Figure 5. The influence of different Markov arrival load factor mechanisms on the probability of delay violation

图 5. 不同马尔可夫到达负载率机制下对时延违反概率的影响

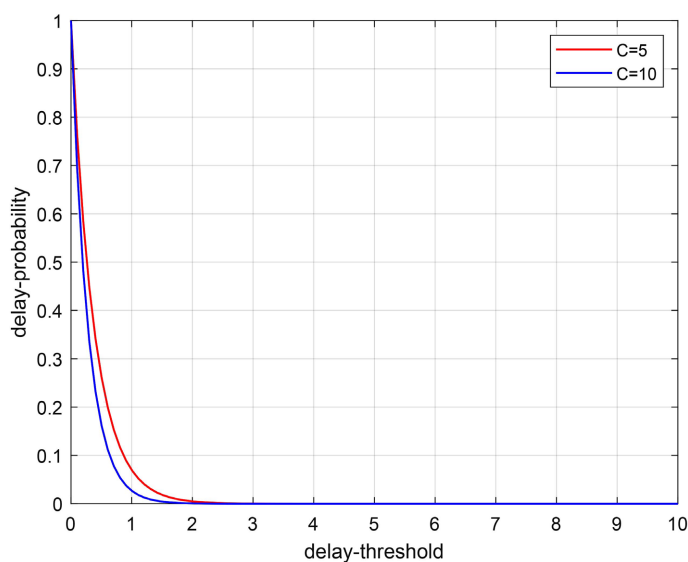


Figure 6. The influence of the non-constant rate service mechanism on the probability of Poisson arrival delay violation

图 6. 不同常速率服务机制下对泊松到达时延违反概率的影响

由系统仿真分析, 在服务机制分别为 $C=5$ 、 $C=10$ 的情况下, 在同一时刻, 时延违反概率随着服务机制常数 C 的增加而更加严格。随着时间阈值的增加, 时延违反概率都在减小, 对于常数较大的服务机制而言, 下降的相对较快。

对于给定 $P=0.6$ 、 $h_a=1$ 、 $h_s=1$ 和 $C=20$ 的情况下, 分别设置 $\lambda=1$ 和 $\lambda=10$ 的情况下。如图 7 所示, 由仿真可以分析出在到达强度较大的情况下, 在同一时间阈值的条件下时延违反概率相对较为宽泛。对于泊松到达强度小的情况下, 时延违反概率相对更加严格。在不同的到达强度下, 时延违反概率都在随着时间阈值的增加而减小。

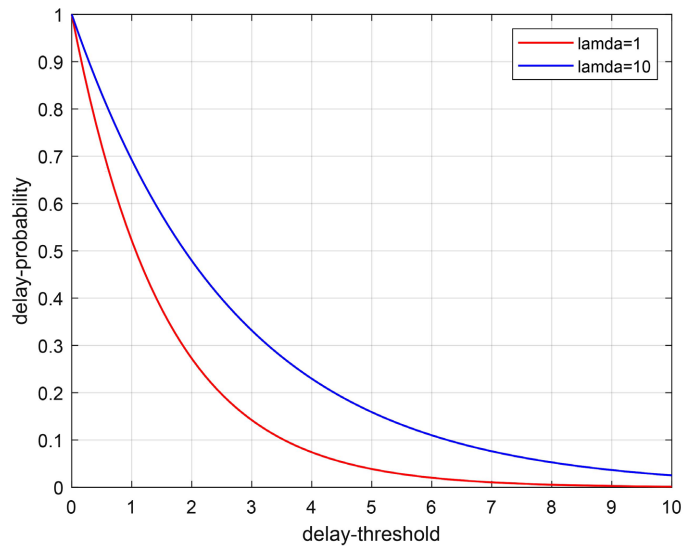


Figure 7. The influence of different Poisson arrival strength mechanisms on the delay violation probability

图 7. 不同泊松到达强度机制下对时延违反概率的影响

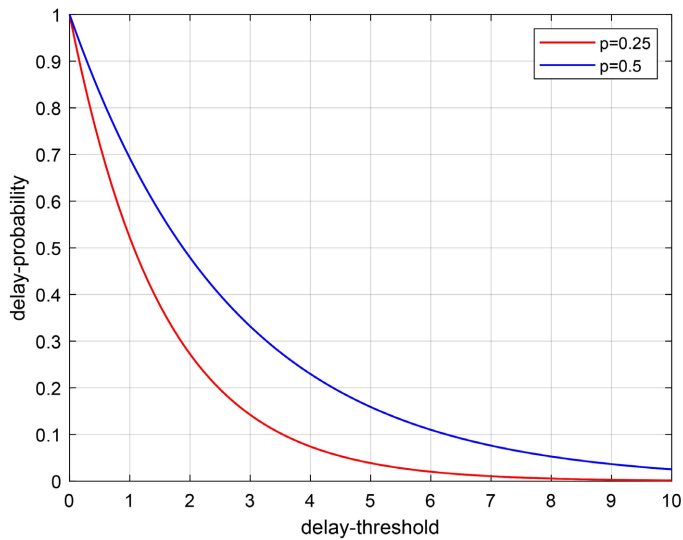


Figure 8. The influence of different Poisson arrival load factor mechanisms on the delay violation probability

图 8. 不同泊松到达负载率机制下对时延违反概率的影响

对于给定 $P=0.6$ 、 $h_a=1$ 、 $h_s=1$ 和 $C=20$ 的情况下, 分别设置 $\rho=0.25$ 和 $\rho=0.5$ 的情况下。在不同泊松到达负载率机制下对时延违反概率的影响如图 8 所示, 经系统仿真可以得出, 在同一时刻随着负载率的增加, 时延违反概率相对于负载率较小的系统而言更为宽松, 对于负载率较小的系统, 时延违反概率更严格。对于不同的负载率, 两者的时延违反概率都随着时间阈值的增加而减小。

5. 结论

本文针对于网络系统时延分析进行了建模, 精确地评估分析了网络时延性能问题。在此文中引入了一种新的理论方法, 鞅理论可以获得更加严格的时延性能。基于鞅理论, 本文通过利用泊松到达和马尔

可夫到达的情况, 研究了在常速率服务机制下的两种时延违反概率, 并利用 MATLAB 软件进行仿真分析。通过设置不同的到达以及服务参数, 利用软件仿真分析其对时延违反概率的影响。

参考文献

- [1] 韩悦, 刘增基, 姚明昨. 基于指数上鞅的统计端到端时延分析[J]. 计算机学报, 2012, 35(10): 2016-2022.
- [2] 胡雪. 基于鞅理论的 VLC 系统时延 QoS 保障算法[D]: [硕士学位论文]. 吉林: 吉林化工学院, 2023.
- [3] Chang, C.S. and Thomas, J.A. (1999) Effective Bandwidth in High-Speed Digital Networks. *IEEE Journal on Selected Areas in Communications*, **13**, 1091-1100. <https://doi.org/10.1109/49.400664>
- [4] 孙玥鑫. 时延 QoS 保障下的 SFC 逐跳带宽分配和部署方法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2023.
- [5] Jiang, Y. (2006) A Basic Stochastic Network Calculus. *ACM SIGCOMM Computer Communication Review*, **36**, 123-134. <https://doi.org/10.1145/1151659.1159929>
- [6] Roughan, M. and Pearce, C. (2002) Martingale Methods for Analysing Single-Server Queues. *Queueing Systems*, **41**, 205-239. <https://doi.org/10.1023/A:1015851021001>
- [7] Ciucu, F. Poloczek, F. and Schmitt, J. (2013) Sharp Bounds in Stochastic Network Calculus. *ACM SIGMETRICS Performance Evaluation Review*, **41**, 367-368. <https://doi.org/10.1145/2494232.2465746>
- [8] Luan, G. (2014) Buffer Stopping Time Analysis in Data Center Networks. *IEEE Communications Letters*, **18**, 1739-1742. <https://doi.org/10.1109/LCOMM.2014.2356478>
- [9] Yu, B., Chi, X. and Sun, H. (2020) Delay Analysis for Aggregate Traffic Based on Martingales Theory. *IET Communications*, **14**, 760-767. <https://doi.org/10.1049/iet-com.2019.0282>
- [10] 荆一航. 时延 QoS 约束下无线资源切片技术研究[D]: [硕士学位论文]. 长春: 吉林大学, 2021.
- [11] 李帅. 通信网络的可靠性冗余分配问题研究及可靠性设计[D]: [硕士学位论文]. 长春: 吉林大学, 2023.