

The Fine Thread between Stability and Randomness of the Non-stationary D/M/1 Queue's GI/M/1 Pointwise of Stationary Fluid Flow Approximation Model with Application Ultra-Low Latency of Autonomous Driving Service

[Ismail A Mageed](#) * and Amina Becheroul

Posted Date: 10 May 2024

doi: 10.20944/preprints202405.0649.v1

Keywords: State variable; mean arrival rate; time; time dependent root parameter; PSFFA; ultra-low latency; autonomous driving service



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

The Fine Thread between Stability and Randomness of the Non-Stationary $D/M/1$ Queue's $GI/M/1$ Pointwise of Stationary Fluid Flow Approximation model with $GI/M/1$ Application Ultra-Low Latency of Autonomous Driving Service

Ismail A. Mageed ^{1,*} and Amina Becheroul ²

¹ School of Computer Science, AI, and Electronics, Faculty of Engineering and Digital Technologies, University of Bradford, United Kingdom

² Department of Mathematics, Faculty of Exact Sciences and Informatic, Hassiba Benbouali University of Chlef, Algeria; a.becheroul@univ-chlef.dz

* Correspondence: drismail664@gmail.com

Abstract: The current work reveals the fine tuning between stability zones and randomness of $GI/M/1$ Pointwise Stationary Fluid Flow Approximation (PSFFA) model of the non-stationary $D/M/1$ queueing system. More specifically, this clearly provides more insights into developing a contemporary PSFFA theory that unifies non-stationary queueing theory with chaos theory and fields in both theoretical physics and chaotic systems. This opens new grounds for stability analysis of non-stationary queueing systems. A notable application of $GI/M/1$ queueing model to achieve ultra-low latency of autonomous driving service is highlighted. Concluding remarks associated with future avenues of research are given.

Keywords: State variable; mean arrival rate; time; time dependent root parameter; PSFFA; ultra-low latency; autonomous driving service

1. Introduction

Day-to-day queues include time varying arrival process of customers, which is interpreted by its variance- based nature on the time of day. This can be caused by factors like failure of network resources or non-stationary input loads. These bursty and non-stationary in character networks' traffic as communication networks become more complicated with fluctuating data speeds and quality of service needs. Queuing theory deals with analyzing and understanding waiting times in various scenarios, such as waiting for service in banks or supermarkets, waiting for a response from computers, waiting for failures to occur, or waiting for public transport.

Simulation techniques in the context of queueing systems involve tracking the behavior of the system through repeated execution of the simulation and averaging relevant quantities over different runs at specific time points. By collecting data at various time instants, the system's Behavior can be evaluated over time[1].

In analytical transient investigations, transform techniques are commonly employed to solve differential/difference equation models that arise from an embedded Markov process/chain. These techniques help in analyzing the behaviour of the system over time by transforming the equations into a more manageable form, facilitating the study of transient phenomena in queueing systems.

This paper's road map is: PSFFA theory is overviewed in section 2. In section 3, the $GI/M/1$ Queueing Model is discussed in more details. In section 4, q - threshold of the non-stationary $D/M/1$ queue's $GI/M/1$ PSFFA model of the is revealed. In section 5, typical numerical experiments to evidence the derived analytic results against the numerical portraits. A notable application of $GI/M/1$ queueing model to achieve ultra-low latency of autonomous driving service is highlighted in section 6 Closing remarks combined with the next phase of research are highlighted in Section 7.

2. PSFFA

The PSFFA is a simulation technique that uses a single non-linear differential equation to estimate the queue's average number of users. An equation's form based on steady-state queueing relationships is obtained by this revolutionary approach to provide advantages in terms of generality, simplicity, and computational efficiency. Moreover, these methods have potential applications in developing dynamic network control mechanisms[1].

Think about a queueing system for a single server that has a non-stationary arrival process. $\mu(t)$ and $\lambda(t)$ serve as the time-dependent average queue service and arrival rates, respectively. The system's ensemble average time-dependent state variable is referred to as $x(t)$, $\dot{x}(t) = \frac{dx(t)}{dt}$. Define $f_{in}(t)$ and $f_{out}(t)$ respectively, to be the system's time-dependent flow into and out. Notably, $\dot{x}(t)$, $f_{in}(t)$ and $f_{out}(t)$ are related by:

$$\dot{x}(t) = -f_{out}(t) + f_{in}(t) \quad (1)$$

Consequently,

$$f_{out}(t) = \mu(t)\rho(t) \quad (2)$$

Here $\rho(t)$ defines the underlying queue's server utilization.

For an infinite queue waiting space is infinite,

$$f_{in}(t) = \lambda(t) \quad (3)$$

Equation (1)'s fluid flow model becomes:

$$\dot{x}(t) = -\mu(t)\rho(t) + \lambda(t), \quad 1 > \rho(t) = \frac{\lambda(t)}{\mu(t)} > 0 \quad (4)$$

Setting $\dot{x}(t) = 0$, implies

$$x = G_1(\rho) \quad (5)$$

Additionally, we assume the numerical invertibility of $G_1(\rho)$, namely

$$\rho = G_1^{-1}(x) \quad (6)$$

Equationally, PSFFA rewrites to:

$$\dot{x}(t) = -\mu(t)(G_1^{-1}(x(t))) + \lambda(t) \quad (7)$$

Notably, (7) is extremely general in nature, since the closed form representation of G_1 can be computed for many queues. However, we can numerically or by data of an existing system's fitting curve calculate G_1 .

3. The GI/M/1 Queueing Model

This section discusses the GI/M/1 queueing model, in which the service time has an exponential distribution, and the inter-arrival process has an identical distribution with successive inter-arrival periods. Let $A(t)$ stand for the distribution of inter-arrival times. The GI/M/1 queue's steady state probability for the number of customers a new arrival finds in the system is a geometric distribution:

$$\pi_n = (1 - \sigma)\sigma^n \quad (8)$$

σ ($1 > \sigma > 0$) uniquely solves:

$$\sigma = f_a^*(s)|_{s=\mu(1-\sigma)} \quad (9)$$

where $f_a^*(s)$ is the Laplace-Stieltjes transform of the inter-arrival time distribution $A(t)$, that is:

$$f_a^*(s) = \mathcal{L}^*(A(t)) = \int_0^\infty e^{-st} dA(t) \quad (10)$$

Notably, $\sigma = 1$ solves (9), and the state variable, x reads as:

$$x = \frac{\lambda}{\mu(1-\sigma)} = \frac{\rho}{(1-\sigma)} \quad (11)$$

In determining the PSFFA model, equation (11) re-writes to

$$\rho(t) = x(t) \mu(1 - \sigma(t)) \quad (12)$$

We believe that the non-stationary load will exhibit sinusoidal mean behaviour, which will describe the cyclic load pattern over a specified time period (for example, day) in accordance with the prior research on non-stationary analysis of communication networks[2–6], namely $\lambda(t) = A + B \sin(\omega t + D)$, for more details see[7–9].

Thus, the required model reads as:

$$x'(t) = \mu x(t)(1 - \sigma(t)) + \lambda(t) \quad (13)$$

We can numerically solve (15) to visualize the queue's time varying behaviour.

Depending on the inter-arrival distribution $A(t)$, the precise process for figuring out will vary, although it usually involves a root-finding approach like Laguerre's method. The time varying $D/M/1$ queue's $GI/M/1$ PSFFA model reads:

$$x'(t) = -\mu x(t)(1 - \sigma(t)) + \lambda(t), \quad \sigma(t) = e^{\frac{(\sigma(t)-1)}{\rho(t)}} \quad (14)$$

$$\rho(t) = \text{time - dependent server utilization} = \frac{\lambda(t)}{\mu}$$

The $D/M/1$ case in equation (16) corresponds to a deterministic arrival process where the inter-arrival distribution $A(t)$ is a delta function (i.e., $dA(t) = f_a(t)dt$ and $f_a(t) = \delta(t - \frac{1}{\lambda})$)

Mastering the increasability(decreasability) for a function, $f(x)$, a shorthand note reads:

$$f' > 0 \Leftrightarrow f \uparrow$$

$$f' < 0 \Leftrightarrow f \downarrow$$

In a more tangible form, we can visualize figure 1(c.f., [10]).

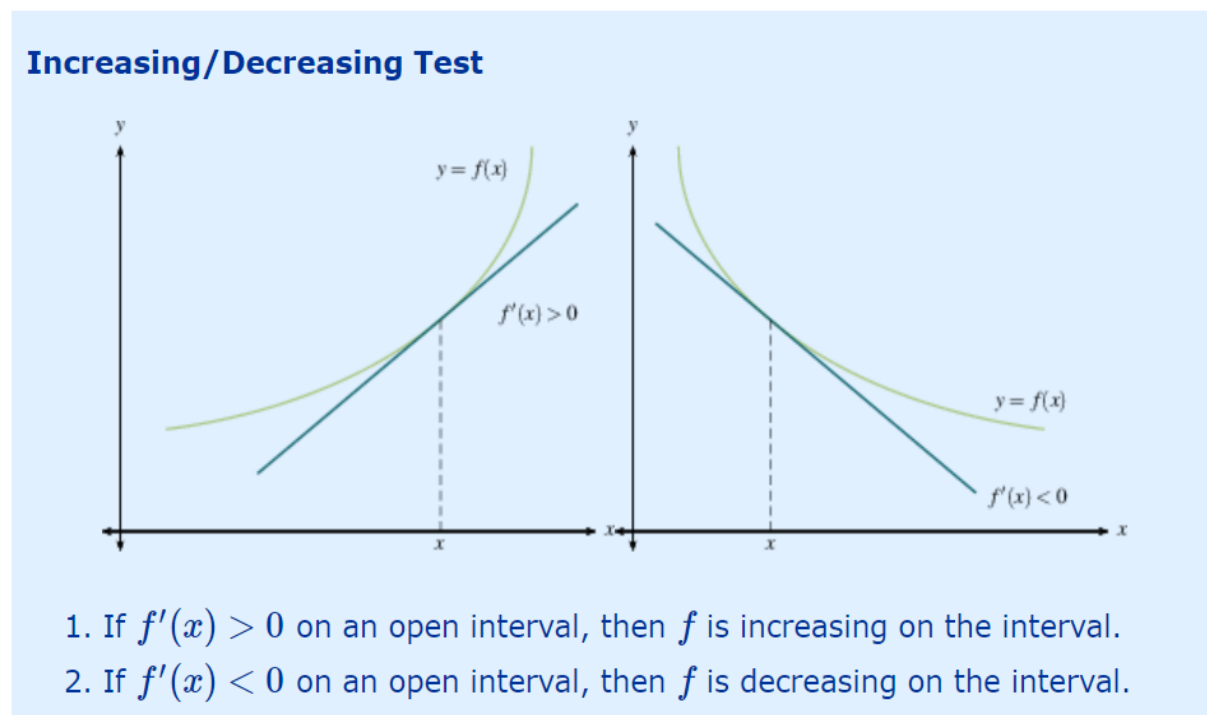


Figure 1. Equational Scenarios of Increase-Decrease Threshold

4. The ρ –Threshold Theory of the Non – Stationary the $D/M/1$ Queue $GI/M/1$ PSFFA Model.

Theorem 1. The time-dependent server utilization, $\rho(t)$ (c.f., (14)) is forever increasing in $\sigma(t)$ ($\sigma(t) \in (0,1)$)

Proof. Let the time-dependent root parameter, $\sigma(t)$ be such that $1 > \sigma(t) > 0$. By (16), it follows that:

$$\rho(t) = \frac{\lambda(t)}{\mu} = \frac{(\sigma(t) - 1)}{\ln(\sigma(t))} \quad (15)$$

We have

$$\frac{\partial \rho}{\partial \sigma} = \frac{\ln \sigma - 1 + \frac{1}{\sigma}}{(\ln \sigma)^2} \quad (16)$$

Following mathematical analysis(c.f., [11]),

$$1 - \frac{1}{\sigma} < \ln \sigma < \sigma - 1 \quad (17)$$

Communicating (16), and (17), the result follows.

We can see that:

$$\lim_{\sigma(t) \rightarrow 1} \rho(t) = \lim_{\sigma(t) \rightarrow 1} \frac{(\sigma(t)-1)}{\ln(\sigma(t))} = \lim_{\sigma(t) \rightarrow 1} \frac{1}{\frac{1}{(\sigma(t))}} = 1 \quad (\text{L'Hopital's rule}) \quad (18)$$

and

$$\lim_{\sigma(t) \rightarrow \infty} \rho(t) = \lim_{\sigma(t) \rightarrow \infty} \frac{(\sigma(t)-1)}{\ln(\sigma(t))} = \lim_{\sigma(t) \rightarrow \infty} \frac{1}{\frac{1}{(\sigma(t))}} = \infty \quad (\text{L'Hopital's rule}) \quad (19)$$

5. Typical Numerical Experiments

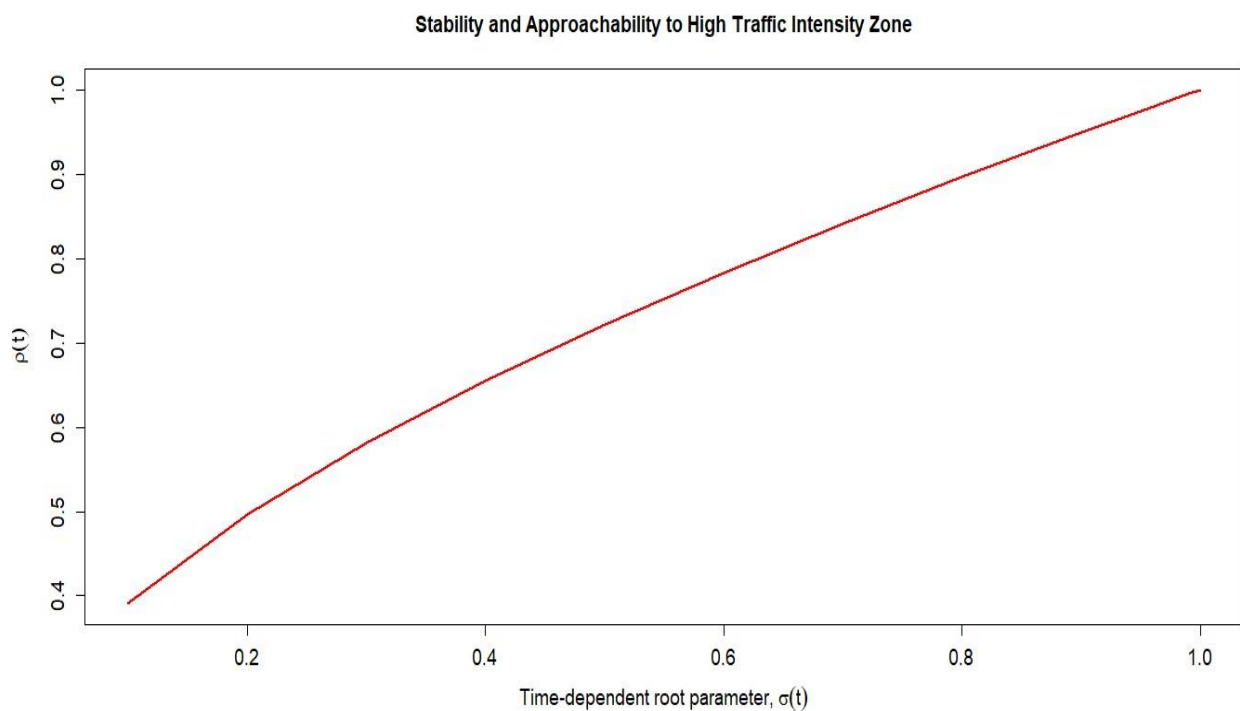


Figure 2. Threshold visibility in stability zone and reaching High Traffic Intensity region($\rho(t) = 1$).

It is observed from Figures 2, and 3 , that the time-dependent root parameter, $\sigma(t)$ has a significant impact on the underlying queue's stability, by directly impacting the time- dependent server utilization, $\rho(t)$. Touching upon stability, it can be seen that $\sigma(t)$ acts as a cutting-edge fine tuning to either approaching a high traffic intensity zone, corresponding to $\rho(t) = 1$.

R code for figure 2

```
sigma <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, 1)
rho <- c(0.3908650337, 0.4970679476, 0.5814084816, 0.6548140008, 0.7213475204,
0.7830460756, 0.8411019756, 0.8962840235, 0.9491221581, 0.9747862873, 0.99499162471, 1)

plot(sigma, rho,
      type="l",
      col="red",
      xlab=expression(paste("Time-dependent root parameter, ", sigma(t))),
      ylab=expression(rho(t)),
      main="Stability and Approachability to High Traffic Intensity Zone",
      lwd=2,
      cex.lab = 1.2,
      cex.axis = 1.2
)
```

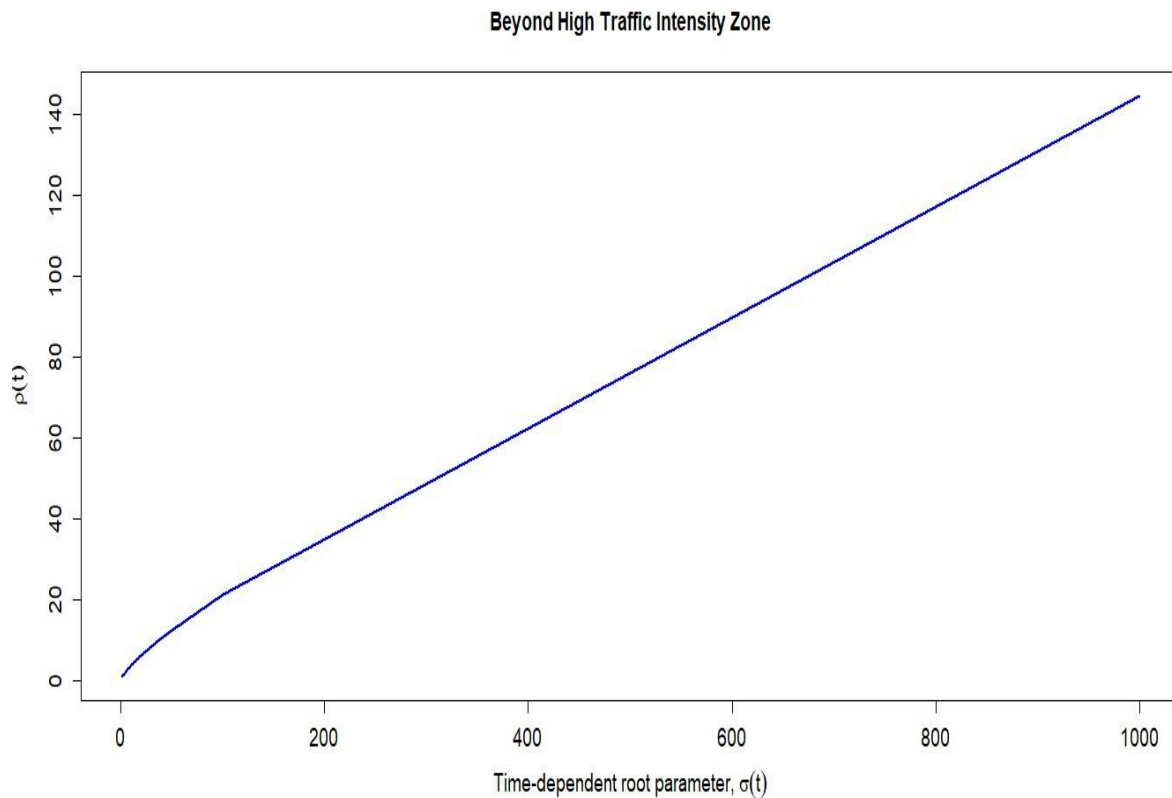


Figure 3. Theshold visibility beyond High Traffic Intensity region($\rho(t) > 1$).

R code for figure 3

```

sigma <- c(1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000)
rho <- c(1.049205869, 1.09696299, 1.143448406, 1.188805365, 1.233151731, 1.276585887, 1.319190975,
1.361038022, 1.402188285, 1.442695041, 1.820478453, 2.164042561, 2.485339738, 3.908650337,
6.342355813,
8.52640901, 10.5723162, 12.52548871, 21.49757685, 144.6200625)

plot(sigma, rho,
      type="l",
      col="blue",
      xlab=expression(paste("Time-dependent root parameter, ", sigma(t))),
      ylab=expression( rho(t)),
      main="Beyond High Traffic Intensity Zone",
      lwd=2,
      cex.lab = 1.2,
      cex.axis = 1.2
)

```

Moreover, the progressive increase of $\sigma(t)$, will steer the whole system into a randomness zone, corresponding to $\rho(t) > 1$.

It can be easily verified that the numerical setup is validating the obtained analytic results of Theorem 1.

6GI/M/1-. Based Ultra-Low Latency of Autonomous Driving Service

The investigation on how 5G networks are anticipated to accommodate different network services with varying performance criteria, such as high-rate traffic, low latency, and high reliability, was conducted by the authors of [12]. New technologies including network functions virtualization (NFV), network slicing, and software-defined networking (SDN) are being implemented to satisfy these expectations. By combining fog, edge, and cloud computing with network slicing to map autonomous driving functionalities into service slices, these technologies will help build a distributed and scalable SDN core network architecture that will ultimately improve transmission efficiency and meet low latency constraints. The goal is to improve the quality of service for autonomous driving applications.

The global system manager in an autonomous driving system [12] plays a crucial role by overseeing three main functions: driving mode management, which includes manual, run, and pause modes; fault management system, which monitors module statuses for safe driving; and emergency response to system faults or operator interventions to ensure the safety and efficiency of the autonomous vehicle operations. These functions are essential for maintaining the operational integrity and safety of the autonomous driving system by managing driving modes, monitoring module statuses, and responding to emergencies effectively. Moreover, this describes a 5G-enabled scalable SDN core network architecture for autonomous driving systems[12], emphasizing the importance of ultra-low latency and high reliability. It introduces the concept of a global system management service implemented as a service slice to meet the stringent requirements of autonomous driving modes, ensuring efficient operation and safety through features like fail-safe mechanisms and human-machine interfaces displaying critical vehicle information, as shown by Figure 4 (c.f., [12]).

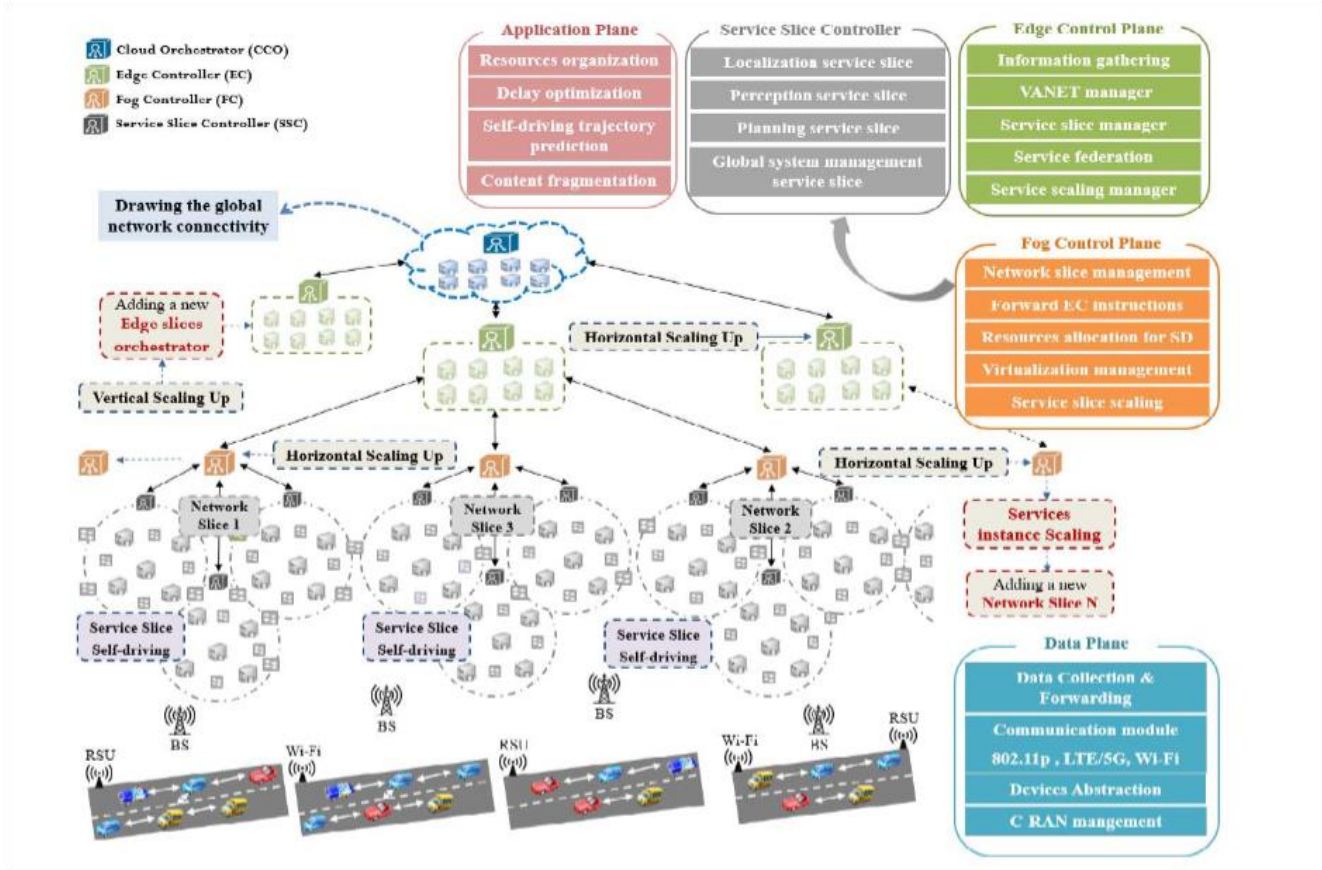


Figure 4. Scalable SDDN core network architecture allowed by 5G slicing (i.e., SliceScal).

Accordingly, the authors[12] provided a simulation-based assessment of their system with a particular emphasis on SliceScal and the 5G slicing model for driverless vehicles. They evaluate the system’s performance in a realistic urban setting using the network simulator NS-3, the cars traffic simulator Veins-SUMO, and the OpenFlow SDN controller, considering elements like vehicle dispersion, radio propagation, and latency management in autonomous driving resource slicing. The evaluation platform and algorithm developed aim to optimize resource allocation and latency handling for autonomous driving services in a dynamic urban environment.

Figure 5 (c.f., [12]) offers a visual description for handling latency of autonomous vehicle service requests in different scenarios involving service slicing and a slice management algorithm. It shows that as the density of autonomous vehicles(Avs) increases, the handling latency also increases. The use of service slicing reduces the handling latency by 60%, and the SliceMan algorithm further reduces it by 90% compared to scenarios without service slicing, demonstrating the effectiveness of managing resources for service slices in improving latency.

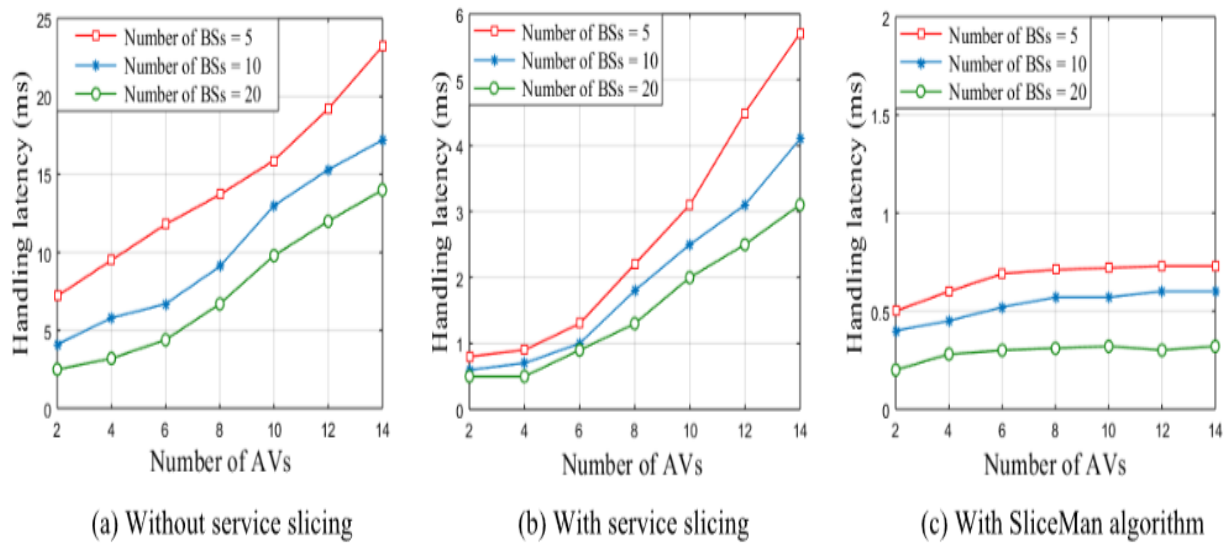


Figure 5. Comparative analysis of how autonomous car service requests are handled when they are late.

7. Conclusion and Future Work.

An exposition is undertaken to reveal the threshold theory of the time-dependent server utilization of $D/M/1$ queueing system's $GI/M/1$ PSFFA closed form expression. Moreover, some numerical experiments are provided to validate the analytic results. The influential impact of $GI/M/1$ queueing model on ultra-low latency of autonomous driving service is investigated. Future work involves further investigation of similar threshold theorems of $G/M/1$ PSFFA model of the non-stationary $E_k/M/1$ and $IPP/M/1$ queueing systems.

References

- [1] Wang, W-P., Tipper, D., and Banerjee, S. 1996: A simple Approximation for Modelling Non-Stationary Queues, IEEE.
- [2] Poryazov S, Andonov V, Saranova E. Methods for Modelling of Overall Telecommunication Systems. Research in Computer Science in the Bulgarian Academy of Sciences. 2021:325-61.
- [3] Zhu M, Gao S, Tu G, Chen D. Multi-Access Edge Computing (MEC) Based on MIMO: A Survey. Sensors. 2023 Apr 11;23(8):3883.
- [4] Mageed IA, Zhang Q, Modu B. The Linearity Theorem of Rényian and Tsallisian Maximum Entropy Solutions of The Heavy-Tailed Stable M/G/1 Queueing System entailed with Potential Queueing-Theoretic Applications to Cloud Computing and IoT. electronic Journal of Computer Science and Information Technology. 2023 Apr 23;9(1):15-23.
- [5] Grassmann, W.K. 1990: Computational Methods in Probability Theory, Handbook in Operations Research and Management Science, Vol.2, North Holland.
- [6] Mageed IA, Zhang,Q. Solving the open problem for $GI/M/1$ pointwise stationary fluid flow approximation model (PSFFA) of the non-stationary $D/M/1$ queueing system. electronic Journal of Computer Science and Information Technology. 2023 Apr 23;9(1):1-6.
- [7] Chang J, Chan HK, Lin J, Chan W. Non-homogeneous continuous-time Markov chain with covariates: Applications to ambulatory hypertension monitoring. Statistics in Medicine. 2023 May 30;42(12):1965-80.
- [8] GRASSMANN WK. Finding Transient Solutions in Markovian Event Systems Through Numerical Solution of Markov Chains. 2021 Jun 30:357.
- [9] A Mageed, D.I. Effect of the root parameter on the stability of the Non-stationary $D/M/1$ queue's $GI/M/1$ model with PSFFA applications to the Internet of Things (IoT). Preprints 2024, 2024011835. <https://doi.org/10.20944/preprints202401.1835.v1>

- [10] Mageed, I. A., & Zhang, Q. (2023). Threshold Theorems for the Tsallisian and Rényian (TR) Cumulative Distribution Functions (CDFs) of the Heavy-Tailed Stable M/G/1 Queue with Tsallisian and Rényian Entropic Applications to Satellite Images (SIs). *electronic Journal of Computer Science and Information Technology*, 9(1), 41-47.
- [11] Kozma, L. 2020: Useful inequalities, online source, available at http://www.Lkozma.net/inequalities_cheat_sheet
- [12] Chekired, D. A., Togou, M. A., Khoukhi, L., & Ksentini, A. (2019). 5G-slicing-enabled scalable SDN core network: Toward an ultra-low latency of autonomous driving service. *IEEE Journal on Selected Areas in Communications*, 37(8), 1769-1782.