Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

A Portable Server-Based Analysis of One or More Queuing Models

Dr. Sudhakar Jyothula[1], Mr. Aritra Bandyopadhyay[2], Dr. R. Venkata Aravinda Raju[3], Dr. Shobana A[4], Dr. P. Hema[5], Dr. Nellore Manoj Kumar[6], Dr. V. K. Somasekhar Srinivas[7]

[1]Professor, Department of ECE, Vignan's Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India, 530049.Email id: sudhakar.jyo@gmail.com

[2]Assistant Professor, Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu, West Bengal, India, Pincode: 712139. Email: aritra.d90@gmail.com

[3]Assistant Professor, Department of Basic Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India, Pincode: 521180. Email: aravindaraju.1@gmail.com

[4]Professor, Department of Science and Humanities, Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India, Pincode: 641008. Email id shobanaa@skcet.ac.in

[5]Assistant Professor, Department of Mathematics, R.M.K. College of Engineering and Technology, R.S.M. Nagar, Puduvoyal, Thiruvallur District, Tamilnadu, India, Pincode: 601206. Email: hema.p@rmkcet.ac.in

[6]Department of Mathematics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Thandalam, Chennai, Tamilnadu, India, Pincode: 602 105. Email id: nelloremk@gmail.com.
ORCID: 0000-0002-1349-800X

[7]Associate Professor, Department of Mathematics, School of Liberal Arts and Sciences, Mohan Babu University, (Erstwhile Sree Vidyanikethan Engineering College), Sree Sainath Nagar, Rangampeta, Tirupati, Andhra Pradesh, India, Pincode: 517102. Email id : somasekharsrinivas@gmail.com

Article History:

Received: *27-09-2024*

Revised: *28-11-2024*

Accepted: *09-12-2024*

Abstract:
The paper presents a rational approach to analyse the administrative characteristics of Programming Characterised Systems administration. The perspective of "programming characterised systems administration" is intended to improve the efficiency and display of organisational figures. The complexity of organisations nowadays that are built around this concept necessitates characterising the adequate methods for the computational work qualities. The model proposed to be used for QoS analysis is represented by a Poisson appearance process with boundary λ and administration time given by hyper-outstanding dissemination with r phases. We analyse how different bundle size distributions impact Web hub displays. Specifically, we used deterministic, Pareto, dramatic, and typical dispersion to illustrate how the bundle sizes are changeable. The parcels' seasons of manifestation in each of the four information stream scenarios followed a similar dramatic dispersion. Then, in every case, the typical parcel pausing and stay season was broken. It should be noted that regardless of the applied bundle size circulation, the mean bundle size, parcel burden, and server consumption remained essentially the same; however, the article demonstrates that the completed exhibits varied greatly. The displays were constructed with the numerical information generated for the M/G/1/∞/∞/FCFS queuing mechanism.

Keywords: Portable Server-Based, Analysis, Queuing Models

1. INTRODUCTION

Due to their countless uses in complex contemporary correspondence, transportation, administration, and creation systems, queuing models with administration interference have been widely concentrated.

As the aforementioned reasons show, up until this point, the writing on the help interference can be broadly divided into two categories, depending on whether the interference is caused by external factors (like unanticipated setbacks) or, conversely, by internal factors (like the gradual breakdown of the servers (offices)).

The need for flexible and efficient processing arrangements in today's advanced scene has driven advancements in portable server-based analysis apparatuses. These devices serve as vital tools for comprehending and developing intricate systems, particularly in fields like media communications, healthcare, transportation, and assembly. Of the multitude of perceptive theories, queuing models are unique as essential tools for assessing the presentation of systems defined by the appearance of substances, such as customers or tasks, and the management that follows. In this study, we began an exploration into the meaning and uses of portable server-based analysis within the queuing model system. By means of a definitive evaluation of the standards, methodologies, and actual consequences, we hope to elucidate the significant impact of these analyses on improving system performance, asset allocation, and dynamic cycles.

Queuing models provide a systematic framework for deconstructing and optimising various queuing systems. They are based on the standards of stochastic cycles and probability hypothesis. These systems cover a broad spectrum of real-world scenarios, from information parcel lines in PC organisations to customer service lines in retail settings. Queuing models operate with a quantitative understanding of important execution metrics such as holding up times, queue lengths, and administration rates by abstracting complex systems into numerical structures. As a result, partners can modify functional boundaries, remove obstructions, and increase overall system throughput by utilising the knowledge gained from queuing analysis.

The popularity of portable server-based analysis devices has changed which queuing models are available and suitable in practical contexts. Unlike traditional methods that rely on particular hardware or cumbersome software setups, portable server-based arrangements provide flexibility and ease of transmitting across various registration situations. These devices employ improvements in virtualization, containerisation, and distributed computing to empower professionals to oversee remote queuing tests using software stacks that are stage-skeptic, lightweight, and flexible. As a result, associations can send advanced solutions for emerging functional issues more rapidly, repeat plan cycles, and accelerate the displaying system.

Additionally, the range and impact of queuing models are expanded beyond typical office situations by the synchronisation of server-based analysis with portable processing stages. These mobile solutions, which can be sent on handheld devices, edge figuring hubs, or Internet of Things devices, provide continuous verification, analysis, and decision-making at the organisational edge. As a result, partners obtain valuable insights into the components of system execution, enabling proactive mediations and adaptable asset assignment strategies. Portable server-based analysis facilitates advancement, flexibility, and effectiveness across various domains and functional contexts by mitigating any challenges that may arise between hypothetical showing and certifiable execution.

The systems, applications, and contextual analyses that demonstrate the practicality of portable server-based analysis in queuing model contexts are covered in more detail in the next sections of this work.

Through elucidating fundamental concepts, presenting empirical findings, and examining practical implications, our aim is to provide readers with a comprehensive understanding of this innovative scientific paradigm. In the end, we argue that portable server-based analysis solves an important advancement in the field of queuing hypothesis, allowing collaborators to study complexity, improve asset utilisation, and foster functional greatness in a world that is unquestionably linked and dynamic.

2. LITERATURE REVIEW

Bruneel et al. (2016) provide a basic discrete-time line model with a changeable help limit. Their analysis provides insights into a key component of queueing systems that aid with long-term change limitation. The authors delve into the examination of a basic model for these kinds of systems. Through the use of scientific methodologies and the establishment of critical execution metrics, they provide a comprehensive understanding of these queues' behaviour.The significance of this work resides in its dedication to extending traditional queueing models to require dynamic help restrictions, which are common in several certified scenarios. The experiences helped to improve asset utilisation and system performance in the circumstances indicated by the varying support requests.

Chaudhry et al. (2019) focus on the GI(X)/Geo/c queueing model, in which the help times are widely distributed with a finite limit c and the appearance times between them follow an overall dispersion. Their review focuses on finding computationally efficient, yet scientifically simple, solutions for this specific queueing problem.Through the use of probabilistic analysis and numerical methodologies, the authors provide intelligent solutions that integrate with the board of GI(X)/Geo/c lines and comprehension. Their commitments cover both speculative developments and practical implications, providing solutions that balance computational possibilities with perceptive meticulousness.

De Muynck et al. (2017) Examine discrete-time lines with limited help administration limits and requests for general support. Their research covers a broad spectrum of administrative attributes and boundary imperatives, providing a comprehensive framework for comprehending the behaviour of these kinds of systems.The authors elucidate the components of lines under variable support requests and constrained help administration constraints by means of a detailed numerical analysis and illustrative strategies. Their findings contribute to the advancement of queueing theory in several application domains by providing valuable insights into asset portioning strategies and system execution enhancement.

Dudin et al. (2022) suggested a creative paradigm for self-management systems in which client assessments affect appearances. The authors of this review investigate the implications of previous clients' assessments for the appearance pace of subsequent clients. The model aims to capture the essential components of customer behaviour under self-management settings, where future appearances are heavily influenced by the loyalty of the customer. The evaluation provides insights into the plan and the executives of self-administration systems by combining rating-subordinate appearances. It highlights the importance of client feedback in the provision of support.

Grosof et al. (2022) introduced WCFS, a different framework for analysing multiserver systems. "Responsibility Conveying Liquid Help" (WCFS) provides a comprehensive method for decomposing and illustrating complex multiserver scenarios. Through the integration of responsibility analysis and liquid-based showing techniques, WCFS provides an adaptable framework for evaluating execution metrics such as throughput, reaction time, and asset utilisation in multiserver systems. The review provides insights into asset assignment and system streamlining techniques, which advances the development of logical tools for analysing the behaviour of large scope management systems.

He and Alfa (2015) focused on creating Markov chains within a limited aid administration limit (K) for discrete-time lines with general help requests handled by Markovian Arrival Processes (MAP) and Phase-Type (PH) conveyances. The review offers a methodical approach to creating the queueing system-related change likelihood lattice of the Markov chain. The review provides computationally productive calculations to analyse the consistent state behaviour of discrete-time lines with variable assistance limitations by leveraging the features of MAP and PH disseminations. This work advances the development of intelligent protocols for evaluating and visualising queueing systems with various help requirements.

3. QUEUING MODEL

In this section, we provide an accurate description of the queuing model that is the focus of this paper along with the numerical data that was used for the analysis.

The queuing model under examination is discrete-time, meaning that time is divided into time allotments. Only those users who arrive inside the allocated time are allowed to access the system until the end of that session. The number of clients who arrive at schedule opening k is denoted by Ak. With pgf (z)), these arrival quantities $\{Ak|k=1,2,...\}$ are believed to be i.i.d. irregular factors with an overall normal circulation.

$$\boldsymbol{A(Z) \triangleq \sum_{n=0}^{\infty} p(arbitrary\ slot\ contains\ \ n\ \ arrivals)\, z^{n}.} \qquad \mathbf{(1)}$$

The mean number of arrivals per slot is denoted as $\lambda \triangleq A'(1) \triangleq '(1)$.

The amount of labour that each customer expects from the system, or the "administration demands," are considered to be positive numerical quantities of "work units." The intention behind client K's help interest is Sk. With pgf (z), the aid needs $\{Sk|k=1,2,...\}$ are believed to be i.i.d. irregular components with an overall normal circulation. The notation for the mean help request is $\tau \triangleq S'(1) \triangleq '(1)$.

In our queuing paradigm, there are c servers for a given number $c \geq 1 \geq 1$. Every server in the schedule has a "administration limit" that indicates the maximum amount of work that the server can perform for customers within that time slot. A number of work units, denoted as Ri,, is the help limit of server I during space k. It is believed that the assistance limitations $\{Ri,k|i=1,2,...,c,k=1,2,...\}$ are i.i.d. irregular factors with a usual conveyance with pgf R(). The notation $\mu \triangleq R'(1) \triangleq '(1)$ denotes the mean help limit of a server (per space). It is believed that the pgf (z) is an intelligent capability. This is how we introduce the often-occurring prime polynomials $P(z)$ and $QR(z)$, hoping that

$$\boldsymbol{R(1/Z) = \frac{P_R(Z)}{Q_R(Z)},} \qquad \mathbf{(2)}$$

and we let m denote the degree of $QR(z)$.

The Ri, work units of administration limit are consistently carried out in a work-saving manner. That is to say, if a server's accessible help limit is greater than the client's (remaining) administration interest in assistance from that server, then a subsequent client can begin administration right away (during a similar schedule opening). This rehash continues until either the server's accessible aid limit is reached or there are no more customers in the queue. As such, it is possible for a single server to serve multiple customers in a single opening. Alternatively, if the client's excess interest in assistance from server I exceeds Ri,k, then this request for assistance effectively decreases by Ri,k and the client's assistance continues in the subsequent opening k+1+1. As a result, a single client care may have multiple opportunities.

The following is the planning process that servers are assigned to customers. A single line that is shared by every server exists. A single server can provide administration to all of its clients. The client in front of queue goes to an erratic server at the beginning of each opening, which has at least one work unit of administrative limit available. This process is repeated until either all servers are servicing clients with an aggregate (residual) administration request that is more than or equal to their administration limit during that opening, or until there are no more clients in the queue.

The number of clients that were in the system when the kth schedule opened is denoted by Bk. This is often referred to as the system content at the beginning of the kth space. The number of openings between the end of this client's arrival space and the end of the final space in which the customer receives administration is known as the postponement of the kth client, and it is denoted by the symbol Dk. The two essential characteristics of an exhibit that we are interested in are the confining transports of Bk and Dk as k approaches infinity, that is, in the so-called "consistent state". The system is meant to be "steady" at the point where these two irregular components have a limiting conveyance. In other words, when the heap $\rho=\lambda\tau/\mu c<1$, or when the mean all-out assistance interest of all clients coming up per opening is not exactly the mean help limit per space. When the system reaches equilibrium, we denote the pgf of the limiting distribution of $\mathbf{B}k$ and Dk as, respectively, (z) and $D(z)$. Conversely, pgf(z) represents the pgf of the system content at the beginning of an erratic space in a consistent state, and pgf(z) represents the pgf of an erratic client's deferral in a consistent state.

The concept of "incomplete work," which is essentially the total of all unfinished work units of the assistance interest of every client in the system, will also be used in the study of this system. The incomplete work that a client C views is defined as the amount of three free terms that are present in the system immediately following the client C's appearance:

1. The (excellent) management interest of all clients presently enrolled in the system before to client C's arrival space and enrolled in the system after client C's arrival opening;

2. Different clients' work units appearing in the same opening as client C but appearing earlier than client C;

3. The actual help-seeking interest of client C.

4. AN QUEUING MODEL SELECTION

The main justification for the numerical model is its ability to accurately represent the operational principle of an SDN-based network. It should be flexible enough to handle any number of traffic inquiries directed towards the regulator. In addition, the model should consider the anticipated number of hubs and be effectively extendable. As a result, the selection of numerical devices for the SDN-based network's display and analysis should take into account the associated characteristics, such as the kind of traffic and its volume, the number of servers, and the protocols for managing incoming requests. The following highlights should be taken into account:

- In an SDN-based organisation, the persuasiveness of client solicitations can vary greatly. Solicitations may be sent out singly or in erratic groups;
- The characteristics, structure, and management guidelines of the different forms of administrations (video, discourse, or registration gaming administrations) are distinct;
- The customs' requests for assistance, which are gathered in case the number of participants exceeds the capacity to accommodate;
- The cooperation between the regulator and the OpenFlow switch requires input from the regulator.

SDN-based network analysis must take into account the fact that real streams are heterogeneous, meaning that demands originate from different sources and have varying levels of power. The amount of data that has been gathered with the source (video, audio, or information) and the percentage of the total number of sources determine the value of the next stream rate at each second. A few features of Open Flow switches and regulators allow several devices to be serviced simultaneously with different levels of aid power for intricate support offering significance. An SDN-based organization's usefulness can be calculated using the formula M H m r. A Poisson arrival process with boundary λ represents this model; the measure of Open Flow switches increases to m, and the administration time exhibits hyper-outstanding propagation with r phases (H,) with mean1 μr.

There are some limitations to the suggested queueing model. The model facilitates the examination of a two-layered state-space that addresses the number of clients and the number of phases that still require service for the client receiving assistance. The phase circulation r in the suggested model equates to two for disentanglement. The model that is being used is 2 M H m//. In contrast to the suggested queueing model, which acknowledges reasonable methodology, consider several streams between the inputs and the information plane (correspondence hardware). Figure 1 shows the proposed queueing model's layout. Capability can be used to characterise the thickness of the information demand stream (M) from clients to switch t b t e–λ=λ , the thickness of result demands stream from change to regulator (H2) can be characterized by capability$\boldsymbol{a(t) = p\mu_c e^{-\mu_c t} + (1-p)\mu_{sw} e^{-\mu_{sw} t}}$.

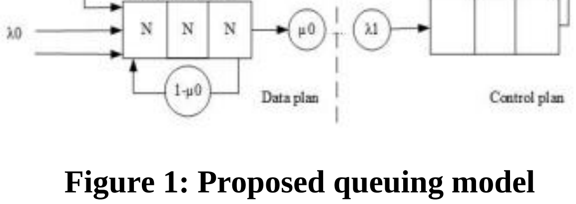

Figure 1: Proposed queuing model

Requests from customers arrive at the switch ports (1 ... m) with different intensity λi. λ0 is common requests arrival intensity from different sources $\boldsymbol{\lambda = \sum_i^k \lambda_i}$. We consider a several possible scenarios:

- The information requests from clients processed on the Open Stream switch and forwarded to the regulator with administrative force and probability. There is a relationship between layers in this scenario.

- Client requests for information are handled by an OpenFlow switch and forwarded to another OpenFlow switch with administrative authority and probability. Demands for this scenario do not depart from the information strategy. Is the method of addressing control plan solicitations (from regulator to regulator or from change to regulator) effective in managing demand from regulators?

For models type 2 H M m / / Laplace transform for H2 has the next form: $A^*(s) = p\frac{\mu_0}{s+\mu_0} + (1-p)\frac{\mu_1}{s+\mu_1}$,

for M has the next form: $B^*(s) = \frac{\lambda_0}{s+\lambda_0}$.

The significant nature of administration boundaries of SDN-based network like an organization with high need and continuous applications are

- Deferral and jitter;

- Regulator and switches handling speed;

- Hindering likelihood and parcels misfortune.

An analysis and assessment of these QoS boundaries by proposed queueing model are given in following segments.

A. An Evaluation of Processing Speed and Delay

The OpenFlow switch on the server is the target of the clients' requests. A few requests are quickly fulfilled by sending someone else to the holding area, cradle and administration queue. Administration discipline is necessary for administration holding up time. A major component of the deferral is the average help holding up time (W). W is defined as the amount of time an employee spends in the company from the moment a package arrives at administration hardware until it departs service gear. One can obtain W in queuing system 2 M H m by varying the Little's low. One way to describe the ability that determines the possibility of administration holding up time in the queuing system is as:

$$\boldsymbol{P_{w(t)} = \sum_{i=0}^{m} P_{w_i}(t) e^{-(m\mu_i - p)t}}, \qquad (3)$$

Where τ is amount of incoming requests, ρ is the rate arrival intensity (λi) of request to serve intensity (μi), wi p is i -th request awaiting probability.

Average duration of waiting in the system W (t) can be calculated as follows:

$$\boldsymbol{W(t) = \frac{P_{w(t)}}{n\mu - p} = \frac{1}{\eta - \mu}}, \qquad (4)$$

Where, η is variation coefficient $\eta = \sqrt{\frac{c_2^2}{4} + c_1} - \frac{c_2}{2}$, $c_1 = \mu(\lambda_i(1-p) + \lambda_{i+1}p) - \lambda_i\lambda_{i+1}$, $c_1 = \lambda_i + \lambda_{i+1} + \mu$

The absolute gain from postponing the SDN-based network's season where hub 1 and hub 2 cooperate via regulator, as shown in Figure 1, may be calculated as:

$$\boldsymbol{W(t)_{ab}} = \begin{cases} \boldsymbol{W(t)_{sw-sw}}; \\ \boldsymbol{W(t)_{sw} + W(t)_{c} + W(t)_{sw}} \end{cases} \qquad (5)$$

Waiting time in this instance can be expressed as:

$$\boldsymbol{W(t)_{ab}} = \begin{cases} \boldsymbol{\dfrac{p^2}{\eta - p}}; \\ \boldsymbol{\dfrac{(1-\eta)^2\mu_1^2 p^2}{\eta\mu_0^2(\eta\mu_0 - (1-\eta)\mu_1 p)}} \end{cases} \qquad (6)$$

The following formula can be used to calculate the average value of the interval between progressive solicitations:

:

$$\boldsymbol{\bar{\tau} = \frac{1}{N}\sum_{i=0}^{N}(t_{i+1} - t_i)} \qquad (7)$$

where N is the number of time intervals between the analysed request1,i, and the timestamp of the i-th request coming.

B. An Evaluation of Packets Loss Probability

Anticipate that solicitations may arrive at the system's entry in a comparable amount of time. "Disavowal of administration" of the state system is possible at any time for any solicitation. Arrival requests likelihood can be used to calculate the likelihood that serve gear will not be accessible:

$$\boldsymbol{p_i = \frac{p(H_i)}{\sum_{i=0}^{k} p(H_k)}} \qquad (8)$$

5. UTILIZATION OF THE M/G/1/∞/∞/FCFS MODEL IN THE PERFORMANCE ANALYSIS

Internet hubs, such as switches, doors, extensions, and switches, are essentially queuing systems that may be quantitatively represented using the models of queuing systems that have been developed. The application of a specific model is restrained by:

- The distribution of the inter-arrival times (t_a),
- The service time (t_s) distribution,
- Number of servers,
- System capacity,

- Source capacity and
- Service policy, i.e. the queue’s discipline.

For M/G/1/∞/∞/FCFS It is anticipated that the model will be widely distributed (i.e., the client arrival cycle can be represented as a Poisson or arbitrary arrival process); it will follow any wide distribution (in this study, we look at remarkable, typical, Pareto, and deterministic conveyance); the number of servers will increase to one (the server's limit in our case is 1 Gbps); the system and source limits will be infinite; and the help strategy will be First-Come-First-Served (FCFS), which can be used to represent the best effort organisations, such as the Internet.
We will introduce basic limits using the articulations from the remainder of this section in an effort to deconstruct the M/G/1/∞/∞/FCFS queuing system exhibits, such as the Internet hub. For instance, the traffic force (a) and the number of servers (c) determine the server utilisation (p).:

$$\boldsymbol{p = \frac{a}{c}.} \qquad (\boldsymbol{9})$$

In this study C=1, hence we can write:

$$\boldsymbol{p = a = \lambda . T_s = \lambda . \frac{\overline{p}}{C}} \qquad (\boldsymbol{10})$$

where λ indicates packet load, and Ts represents the average packet service time, ̅is the server capacity as well as the typical packet length. Furthermore, the service time of a particular packet (t_s) is equal to:

$$\boldsymbol{t_s = \frac{p}{c}} \quad (\boldsymbol{11})$$

For p to be the packet size. Keep in mind that the total of all $t_a$equals:

$$\boldsymbol{T = \sum_{i=1}^{N} t_{a_i}} \quad (\boldsymbol{12})$$

where T is deciphered as all out span of an estimation or reenactment and N is the quantity of clients, for example information parcels. Consequently, it tends to be composed:

$$\boldsymbol{\lambda = \frac{N}{T}.} \qquad (\boldsymbol{13})$$

According to the average waiting time $T_w$for this queuing system can be expressed by:

$$\boldsymbol{T_w = \frac{a . T_s}{2.(1 - a)} . \left[1 + \left(\frac{\sigma_{t_s}}{T_s}\right)^2\right]} \qquad (\boldsymbol{14})$$

where σ_{ts}denotes to standard deviation of t_s. Lastly, the average sojourn $T_q$time is:

$$\boldsymbol{T_q = T_w + T_s.} \qquad (\boldsymbol{15})$$

6. CHARACTERISTICS OF THE INPUT DATA FLOWS

A. Packet inter-arrival times

As previously discussed, the use of the M/G/1/∞/∞/FCFS queuing system recognises that client arrival times (ta)—that is, information bundles—are widely distributed amongst them. Stretching that in this study only happens once is crucial. For instance, each parcel was given a unique arrival time that remained constant throughout the four info stream scenarios.

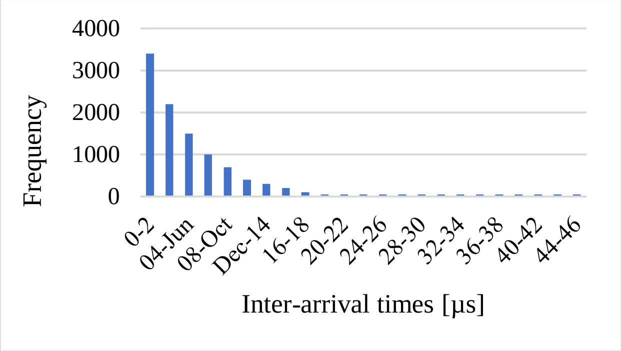

Figure 2: Frequency distribution of t_a

Fig. 2 shows the recurrence dispersion of the mimicked between arrival times. It should be noted that in the span ⟨0,2] µs, 34.46% of all reenacted interarrival times differed. As can be shown from Fig. 3, 56.25% was < 4 µs and 98.39% was ≤ 18 µs.

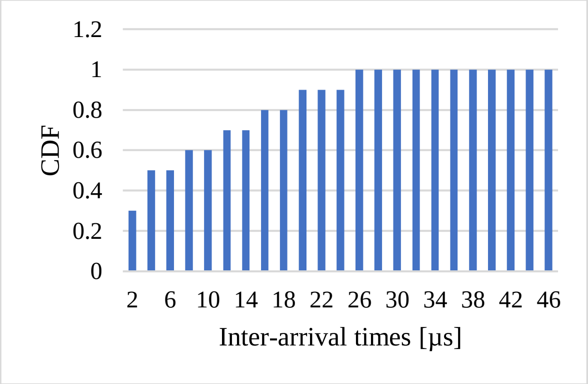

Figure 3: Cumulative distribution function (CDF) of t_a

Using (4) and N = 10,000 parcels, we can determine that the total recreation time is close to 0.0489 seconds given the known arrival timings. Using (5), for instance, we can determine that the bundle load is 204,302 parcels per second.

B. *Simulated variability of the packet size*

In this review, we tested how the bundle size fluctuation affected the hub exhibitions using four different conveyances. Specifically, we replicated four information streams with parcel sizes that corresponded to deterministic, ordinary, Pareto, and dramatic diffusion. The characteristics of these information streams are discussed in the remaining portion of this section, along with an introduction to each stream's important boundaries (administrative time appropriation, $\overline{P}$, Ts, and P).

a) *Input flow 1: Exponential distribution*

The bundle administration times (ts) that were calculated for the initial information stream are displayed in Fig. 4 as the recurrence dispersion. Because there were more produced bundles with sizes ranging from 1000 to 1200 bytes, the help times are communicated considerably with small specific circumstances for (8 <= ts< 8.5) µs. For the connection limit of 1 Gbps, which remained constant in each of the four information stream scenarios, the help times were calculated using (3). In this case, the average bundle size was 423.8344 bytes, with a Ts of 3.3907 µs. (2) indicates that the server utilisation is close to 0.6927, or 69.27%.

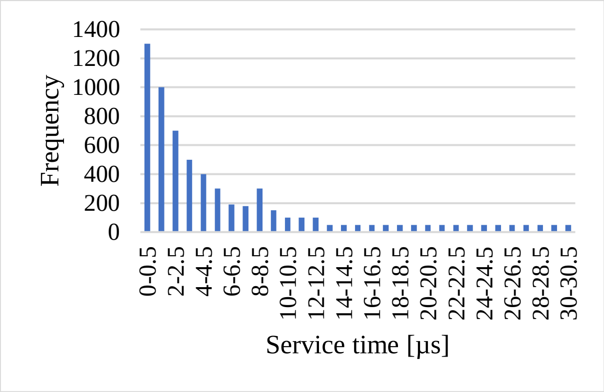

Figure 4: Frequency distribution of $t_s$for the first input flow

Keep in mind that the M/M/1/∞/∞/FCFS model can also be used to illustrate the queuing system for this information stream when it is presented dramatically.

b) Input flow 2: Normal distribution

The average parcel size in this case was 420.509 bytes, which is 3.3254 bytes, not precisely like in the previous case, and Ts was 3.3641 µs. The server's utilisation was almost 0.6873. In this case, the smallest bundle size ever recorded was 22 bytes, while the largest was 809 bytes. Fig. 5 depicts the recurrence frequency of ts in this scenario.

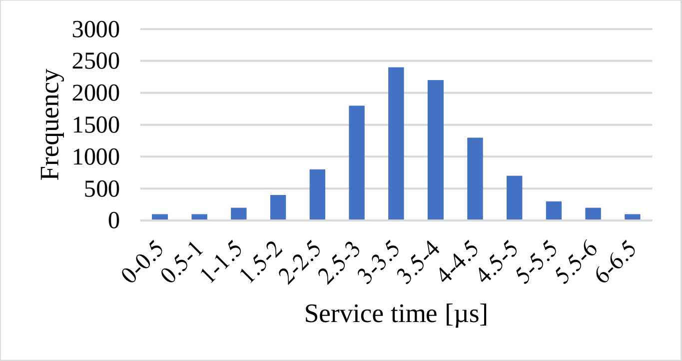

Figure 5: For the second input flow, the frequency distribution of ts

c) Input flow 3: Pareto distribution

Fig. 6 shows the recurrence conveyance for this scenario. In stream no. 3, the average parcel size is equal to 422.6222 bytes, and the average assistance time is 3.3810 µs. Similarly, the determined server use for information streams 1 and 2 is 0.6907.

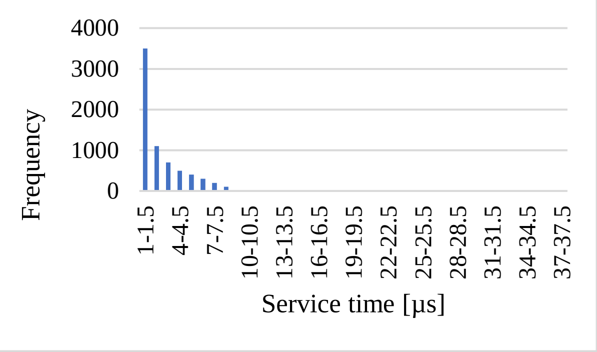

Figure 6: Frequency distribution of $t_s$for the third input flow

d) Input flow 4: Deterministic distribution

Each package in this case had a comparable size of 422 bytes. Therefore, in this case, it is not necessary to visually represent the aid time conveyance. Pressure can be applied for this information flow=t_s=3.376 µs and server utilization equals 0.6897.

It is important to emphasise that the M/D/1/∞/∞/FCFS queuing system model can also be used for this input flow, given $t_s$is deterministic.

7. PERFORMANCE ANALYSIS AND RESULTS

We calculated the upsides of crucial hub execution boundaries, such as the typical pause and stay season of the parcels for each information stream, based on the numerical articulations presented in section 2 and involving the qualities for Ts and p introduced in part 3. We used (6) and (7) separately because of this. Tab. 1 introduces the results.

Table 1: Determined Parameter Values for Various Input Flows

	Input Flow 1 (Exponential)	**Input Flow 2 (Normal)**	**Input Flow 3 (Pareto)**	**Input Flow 4 (Deterministic)**
T_w[µs]	8.3519	4.9304	10.9626	4.7525
T_q[µs]	11.7426	8.2945	14.3436	8.1285

In addition to computing the parameters using the mathematical equations (created for the M/G/1/∞/∞/FCFS queuing system), we also wished to compare the computed values with the simulation's results. Tab. 2 displays the values of the parameters that the simulation produced. As anticipated, the table displays a comparatively minor deviation from the calculated values.

Table 2: Simulated Parameter Values for Various Input Flows

	Input Flow 1 (Exponential)	**Input Flow 2 (Normal)**	**Input Flow 3 (Pareto)**	**Input Flow 4 (Deterministic)**
T_w[µs]	8.5869	4.90	11.565	4.6786
T_q[µs]	11.9776	8.256	14.9440	8.0546

Sequential deliveries of parcel pause and stay durations, separately, for specific information streams are shown in Figures 7 and 8. As can be observed from Fig. 7, when the parcel sizes adhered to the deterministic circulation, tw = 0 for bundles north of 3000. With respect to the next three deliveries, we observe that a larger percentage of the items whose distribution is normal

$t_w \in [1,4\rangle\ \mu s$, tw, compared to the exponential and Pareto distribution.

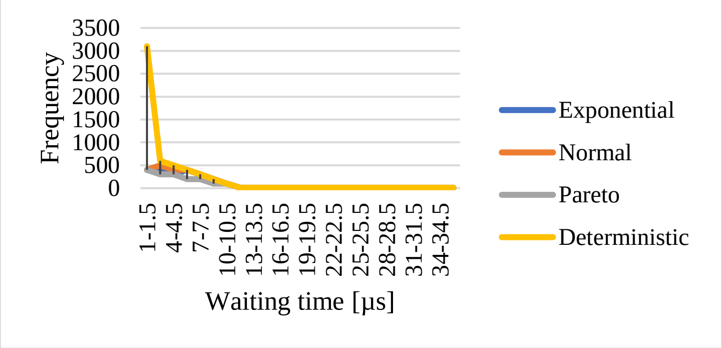

Figure 7: Frequency distribution of $t_w$for different input flows

If we look at the visit times (Fig. 8), we can see that the bundle distribution follows the Pareto and noteworthy conveyance, which records the shortest time. This is due to the fact that, in comparison to the usual or deterministic dispersion, these two conveyances have larger proportions of smaller size bundles. As a result, their administration windows are smaller. The recurrent delivery of tqhas a nearby highest value for the exceptional circulation of the bundle measures at$\boldsymbol{t_q \epsilon [0.5, 1] \mu s}$(430 packets); for the ordinary circulation the most extreme is at$\boldsymbol{t_q \epsilon [3.5, 4] \mu s}$and is equivalent to 914 packets; in contrast, the maximums for the Pareto and deterministic distributions are at $\boldsymbol{t_q \epsilon [1, 1.5] \mu s}$ (1148 packets) and$\boldsymbol{t_q \epsilon [3, 3.5] \mu s}$(3191 packets), respectively.

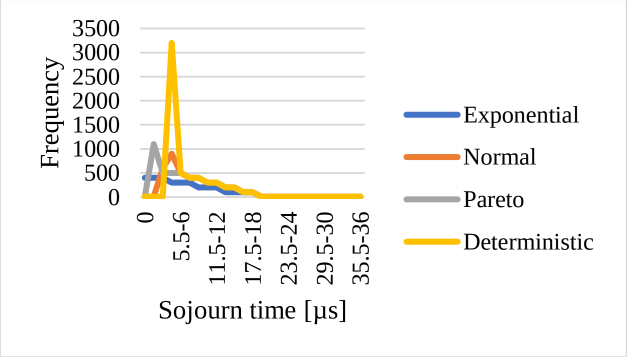

Figure 8: Frequency distribution of $t_q$for all four input flows

Additionally, Figs. 7 and 8 provide a reasonable indication of the importance of the bundle size circulation, as bundle interarrival times were the same in each of the four replications, although $\bar{p}$, ts, and p varied insignificantly.

8. CONCLUSION

Using a portable server-based analysis for at least one of the queuing models provides a robust and efficient method for handling system complexity. Associations can conduct complex investigations with greater flexibility and variety by employing portable server technology, taking ongoing guidance and experiences into account. To understand the effect of bundle size delivery on parcel pauses and visit durations within a speculative Internet hub, we conducted an analysis. In order to determine execution pointers, we developed numerical definitions using the M/G/1/∞/∞/FCFS queuing system model. We were able to observe the impact of bundle size dispersion by simulating four distinct information streams, which ensured relatively small variations in wait times, stay periods, and the probability of queuing. Our findings highlighted the crucial influence of info stream features on hub performance by focusing only on parcel size circulation and excluding between arrival time fluctuation. Significantly, the coefficient of variety emerged as a fundamental execution metric, even when normal help times revealed little differences across various parcel size dispersions. This study provides insights on optimising system execution and asset allocation, highlighting the importance of taking bundle size conveyance into account in network analysis.

REFERENCES

[1] AN Bruno Astuto, M Marc, et al. A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. IEEE Communication surveys and tutorials 2014; 16: 1617-1634.

[2] Brandwajn, A.; Begin, T. Reduced complexity in M/Ph/c/N queues. Perform. Eval. 2014, 78, 42–54.

[3] Bruneel, H.; Wittevrongel, S.; Claeys, D.; Walraevens, J. Discrete-time queues with variable service capacity: A basic model and its analysis. Ann. Oper. Res. 2016, 239, 359–380.

[4] Chaudhry, M.L.; Kim, J.J.; Banik, A.D. Analytically simple and computationally efficient results for the GI(X)/Geo/c queues. J. Probab. Stat. 2019, 2019, 6480139.

[5] De Muynck, M.; Wittevrongel, S.; Bruneel, H. Analysis of discrete-time queues with general service demands and finite-support service capacities. Ann. Oper. Res. 2017, 252, 3–28.

[6] Dudin, A.; Dudina, O.; Dudin, S.; Gaidamaka, Y. Self-service system with rating dependent arrivals. Mathematics 2022, 10, 297.

[7] Grosof, I.; Harchol-Balter, M.; Scheller-Wolf, A. WCFS: A new framework for analyzing multiserver systems. Queueing Syst. 2022, 102, 143–174.

[8] He, Q.; Alfa, A. Construction of Markov chains for discrete time MAP/PH/K queues. Perform. Eval. 2015, 93, 17–26.

[9] K Asif, D Nirav, Enabling Hardware Exploration in Software-Defined Networking: A Flexible, Portable OpenFlow Switch. International Symposium on Field-Programmable Custom Computing Machines (FCCM) 2013; 145-148.

[10] Kim, C.; Dudin, A.; Dudin, S.; Dudina, O. Hysteresis control by the number of active servers in queueing system MMAP/PH/N with priority service. Perform. Eval. 2016, 101, 20–33.

[11] M. Saleh, L. Dong, “Comparing FCFS & EDF Scheduling Algorithms for Real-Time Packet Switching Networks,” In Proceedings of the International Conference on Networking, Sensing and Control (ICNSC), pp. 698-703, Chicago, 2010.

[12] Morozov, E.; Pagano, M.; Peshkova, I.; Rumyantsev, A. Sensitivity analysis and simulation of a multiserver queueing system with mixed service time distribution. Mathematics 2020, 8, 1277.

[13] O Tkachova, S Issam, et al. Mathematical models for analysis Software-defined networks. Information Technologies and Knowledge 2015; 9: 111-123.

[14] R. Srivastava, “Evaluation of Response Time Using Gang Scheduling Algorithm for B2C Electronic Commerce Architecture Implemented in Cloud Computing Environment by Queuing Models,” International Journal of Future Computer and Communication, Vol. 2, No. 2, pp. 71- 75, 2013.

[15] Yao, Y.C.; Miao, D.W.C. Sample-path analysis of general arrival queueing systems with constant amount of work for all customers. Queueing Syst. 2014, 76, 283–308.