Cost Optimization for Hidden Markov Model of Hospital Patient Flow using Heuristic Algorithms and Nurse Staffing Strategies

Vimhala K, Member, IAENG, Lavanya G, Member, IAENG, C.M. Yazhini

Abstract— The most common scenario of admission and discharge in a nurse station which includes patients nurses and HR (Human Resource Personnel) is considered for the model under study. In this study, the Quasi-birth-and-death (QBD) process is applied to optimize nurse, nurse-assistant interactions in hospital settings. By modeling the system as a finite state machine within a hidden Markov model, with states defined as {discharge, admission}, we capture the sequential pattern of events during a single shift. This model is further elucidated through numerical illustrations utilizing the Viterbi algorithm to determine optimal paths for patient navigation. To enhance system efficiency, cost optimization techniques such as genetic algorithms and ant colony optimization are employed. The workflow of a hospital with four stations is modeled as a queueing system and analyzed as a continuous-time Markov chain, revealing steady-state probabilities. Additionally, the study examines patient, nurse-assistant dynamics, emphasizing the diversities within the nursing unit. The interconnectedness of these methodologies demonstrates the comprehensive approach taken: from QBD process modeling to Markov model application and case study validation. Through simulation results and optimal path analysis, heuristic algorithm-based cost optimization, and detailed modeling of multi-station nurse workflows are explained. This thorough analysis culminates in a discussion of system constraints and potential solutions, ultimately contributing to improved healthcare operations management by enhancing efficiency, resource utilization, and patient care outcomes.

Index Terms— Congestion analysis, Markov model, queueing network, hidden Markov model, patient navigation.

NOMENCLATURE

Symbol	Quantity
p_i	Probability of starting in a particular state
A_i	State transition probability matrix which describes the probability of being in state i after state j
B	bed
N	nurse
HR	Human Resource
p_{BB}	Probability of a state transition from bed to Bed
p_{BN}	Probability of a state transition from bed to nurse
p_{BN}	Probability of a state transition from bed to nurse
p_{BHR}	Probability of a state transition from bed to HR
p_{NB}	Probability of a state transition from nurse to Bed
p_{NN}	Probability of a state transition from nurse to nurse
p_{NHR}	Probability of a state transition from nurse to HR
p_{HRB}	Probability of a state transition from HR to bed
p_{HRN}	Probability of a state transition from HR to nurse
p_{HRHR}	Probability of a state transition from HR to HR
π_i	Steady state probability of state i
π_B	Steady state Probability of bed
π_N	Steady state Probability of nurse
π_{HR}	Steady state Probability of HR
a_{ij}	Transition Probability of state i to state j
a_{BB}	Transition Probability of state bed to bed
a_{BN}	Transition Probability of state bed to nurse
a_{BHR}	Transition Probability of state bed to HR
a_{NB}	Transition Probability of state nurse to bed
a_{NN}	Transition Probability of state nurse to nurse
a_{NHR}	Transition Probability of state nurse to HR
a_{HRB}	Transition Probability of state HR to bed
IP	Inpatient
X_i	denotes the order of the requirement
X_n	n^{th} call at the nurse station
a_j	j^{th} Patients in the hospital
A	Admission
D	Discharge

I. INTRODUCTION

Efficient management of patient flows and optimal resource utilization are critical for enhancing healthcare delivery within hospital environments. This study addresses these imperatives through a thorough investigation utilizing advanced modeling techniques and optimization strategies tailored specifically for nurse stations. The foundational framework of the study lies in the Quasi-birth-and-death (QBD) process, which provides a structured approach for understanding the dynamic interactions among patients, nurses, and assistants within hospital units [17].

This theoretical underpinning forms the basis for integrating insights from a robust literature review, which highlights the versatility and applicability of Markov and hidden Markov models in healthcare operations. Previous research has demonstrated the utility of these models in diverse contexts: from assessing quality of care in geriatric wards [4] and classifying medical documents [5], to optimizing resource allocation in intensive care units [7]. [18], In [19] a congestion model of transmission network considering partition. In the proposed model, the distribution companies would relate to HR, nurse and nursing assistant and the distribution of patient workload. Such models offer predictive capabilities essential for anticipating patient flow patterns and optimizing resource allocation in real-time

Manuscript received November 24, 2023; revised October 03, 2024.

Vimhala K is a Research Scholar in the Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur-603203, Tamilnadu, India (e-mail: vk9795@srmist.edu.in)

Lavanya G is an Assistant Professor of the Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur - 603203 Tamilnadu, India (Corresponding author to provide phone: 9941045203; e-mail: lavanyag@srmist.edu.in).

C.M. Yazhini is an undergraduate student in Lakshmiammal College of Education, Chengalpattu, Tamilnadu, India (e-ail: cmyazhini@gmail.com).

scenarios [2, 6]. The literature review underscores the breadth of applications for Markov models, particularly in healthcare operations management. Studies have utilized Markov processes to model patient movements within hospitals, providing insights into capacity planning, staff scheduling, and resource allocation [14, 13]. Hidden Markov models have been employed to analyze complex behaviors and conditions, such as patient readmission risks and treatment outcomes [4, 5]. Simulation studies and optimization techniques, including genetic algorithms and ant colony optimization, have been instrumental in optimizing operational efficiencies in healthcare settings [7, 25, 26]. Additionally, queueing theory has contributed to understanding patient flow dynamics and staffing requirements in hospital environments [15, 16]. In [21], a M/M/1 repairable queueing system with variable input rates and failure rates seem to reflect the hospital scenario with varying crowds of patients and their success rate in treatments offered. [22,23,24,27,28] discuss queueing systems with two types of customers and dynamic change of priority inspiring to incorporate Markov decision process to handle situations. The integration of these methodologies aims to address the multifaceted challenges faced by healthcare facilities in managing patient care and operational workflows. [8,9,10,11] talk about the uncertainty of the availability of bed.

Motivated by the imperative to enhance operational efficiency and patient care outcomes, this study aims to achieve several key objectives: (1) Develop and validate models using Markov and hidden Markov processes to simulate patient movements through nurse stations; (2) Evaluate the efficacy of these models through rigorous numerical simulations, including the application of heuristic algorithms for resource optimization [7]; (3) Explore scalability by analyzing scenarios with multiple nurse stations to assess system performance under varying complexities [12]. In [20] two stage, selection of hyper-heuristic algorithm for solving routing problem is discussed and has been the source of inspiration for incorporating heuristic algorithm for validation.

The motivation behind this research lies in the critical need to improve healthcare operations management. By employing advanced modeling techniques and optimization strategies, healthcare facilities can effectively manage patient flows, reduce waiting times, optimize resource allocation, and ultimately improve patient outcomes. The predictive capabilities offered by Markov models and the optimization potential of heuristic algorithms present promising avenues for addressing these challenges systematically.

The organization of the paper follows a structured approach to systematically address the research objectives: Section II provides a detailed model description of the Quasi-birth-and-death (QBD) process, focusing on its application in healthcare operations and capturing system performance measures, specifically aiming to optimize nurse, nurse assistant interactions in hospital settings. Section III delves into the Markov models and hidden Markov models; Section IV explores their application in modeling patient navigation through nurse stations with case studies and validation discussions. Section V presents the numerical results obtained from simulations and validates model predictions. This section also introduces the Viterbi algorithm, utilized for obtaining optimal paths in modeling patient flows through nurse stations. Section VI examines strategies for cost optimization using heuristic algorithms such as genetic algorithms, ant colony optimization and Markov decision process. In Section VII, the analysis extends to four stations (servers), obtaining steady-state probabilities and illustrating scenarios using state transition diagrams. This section calculates performance measures and graphically illustrates the system's complexity as the number of states increases. Section VIIA briefly discusses system constraints in hospital care and potential solutions using insights gained from the study. The conclusion synthesizes findings from all sections and discusses their implications for healthcare operations management, aiming to foster improved efficiency, resource utilization, and patient care outcomes.

II. MODEL DESCRIPTION

A. Modelling Patient – Nurse - Assistant Dynamics using QBD Process

This section analyses the dynamics between patients, nurses and assistants in a hospital setting, using the Quasi-Birth-Death (QBD) process. The goal is to understand steady-state behaviour, optimize staffing levels and ensure timely patient care. The QBD process accurately models patient arrival and service processes, optimizes staffing levels, and provides insights into system performance, ultimately achieving the goal of minimizing nurses and assistants while meeting the requirement of the patients.

Let X(t) denote the count of assistants at time t, and N(t) represent the count of nurses needed to attend the patients. The model is structured as a two-dimensional continuous-time Markov chain with a state space {(0,1), (0,2), …(0,n)} ∪ {(i,n) : $1 \le n \le N$, i=0,1,2…,n}. This state space accounts for the various combinations of free nurses and assistants, where i denotes the number of occupied nurses and n represents the number of assistants available. The state space now becomes S = {(0,1), (0,2), … (0, n), (1,1), (1,2), …, (1, n), (2,1), (2,2) … (2, n)…}

The transitions signify the movement between states corresponding to the availability of a free nurse or a free assistant. Assuming the arrival rate of a free assistant is denoted by λ_a, when an occupied assistant leaves the system, it is denoted by μ_a, while the arrival rate of a free nurse is denoted by λ_n, and the departure rate of a nurse is denoted by μ_n. The generator matrix is thus formulated as follows:

$$Q = \begin{bmatrix} B_0 & A_0 & \cdot & \cdot & \cdot & \cdot \\ A_2 & A_1 & A_0 & \cdot & \cdot & \cdot \\ \cdot & A_2 & A_1 & A_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & A_1 & A_0 \\ \cdot & \cdot & \cdot & \cdot & A_2 & D_0 \end{bmatrix} \tag{1}$$

Where $B_0 = \begin{bmatrix} -(\lambda_n + \lambda_a) & \lambda_n \\ \mu_n & -(\lambda_a + \mu_n) \end{bmatrix}$,

$A_0 = \begin{bmatrix} \lambda_a & 0 \\ 0 & \lambda_a \end{bmatrix}$, $A_2 = \begin{bmatrix} \mu_a & 0 \\ 0 & \mu_a \end{bmatrix}$

$A_1 = \begin{bmatrix} -(\mu_a + \lambda_a + \lambda_n) & \lambda_n \\ \mu_n & -(\mu_a + \lambda_a + \mu_n) \end{bmatrix}$

$$D_0 = \begin{bmatrix} -(\mu_a + \lambda_n) & \lambda_n \\ \mu_n & -(\mu_a + \mu_n) \end{bmatrix}$$

Stability Analysis And Steady-State Probabilities

The steady state probability is defined as,

$$\pi_{i,n} = \lim_{t \to \infty} P(X(t) = a, N(t) = i), (i, n) \epsilon \Omega. \quad (2)$$

The stability condition ensures that the system remains stable under various operating conditions. It consists of two key components: $\pi Q=0$ and $\pi e=1$, where e is the unit column vector. The vector Π is partitioned as $\Pi = [\Pi_0, \Pi_1, \ldots, \Pi_n]$ where $\Pi_n = [\pi_{0,n}, \pi_{1,n}, \pi_{2,n}]\ for\ n \geq 1$. Using Theorem 3.1.1 in Neuts the necessary and sufficient condition for the system's stability, is expressed as $\pi A_2 e > \pi A_0 e$.

The stationary probability vectors are calculated as follows:

$$\Pi_0 B_0 + \Pi_1 A = 0 \implies \Pi_0 = -\Pi_1 A B_0^{-1} \quad (3)$$

$$\Pi_0 A_0 + \Pi_1 A_1 + \Pi_2 A_2 = 0 \quad (4)$$

$$\Pi_1 A_0 + \Pi_2 D_0 = 0 \quad (5)$$

$$\Pi_n = \Pi_1 R^{n-1}, n \geq 2.$$

and the normalizing condition is

$$\sum_{i=0}^{N} \Pi_0 \mathrm{R}^{\mathrm{i}} \mathrm{e} = 1 \quad (6)$$

This part explains the process of solving for the steady-state probabilities of the system. It involves substituting the equation for R to derive $\pi_0(B_0+RA_2) = 0$. From the last equation (6) we can find the explicit value of rate matrix R. Substituting these values and solving the system of linear equations (3) – (6) we get, the steady state probabilities $\Pi_0, \Pi_1, \ldots, \Pi_n$.

B. System performance measures

Mean number of nursing assistants in the system

$$L_{sys} = \sum_{n=1}^{N} n\pi_n \quad (7)$$

Mean number of non-priority nursing assistants in the system

$$L_{non-prior} = \sum_{n=1}^{N} \sum_{i=1}^{n} n\pi_{i,n} \quad (8)$$

Mean number of priority nursing assistants in the system

$$L_{prior} = L_{sys} - L_{non-prior} \quad (9)$$

The nursing facility comprises of nurses and nursing assistants. However, few nursing assistants may not be trained to assist nurses in all situations. These assistants whose presence does not cause any impact in improving the situation are termed as non-prior assistants. Others fall under prior nursing category. The performance measures are discussed in this context by varying λ_n.

III. OPTIMAL ROUTING TECHNIQUES

A. Markovian model

A Markov model is a random process in which the current value depends only on the previous value and is independent of past values. A Markov process is characterized by the future being independent of the past, given the present. A Markov model, which is a probabilistic model, comprises a sequence of states. The sequence of patient flow in the system is represented by the matrix.

$$\begin{array}{r} \\ \text{Bed} \\ \text{Nurse} \\ \text{HR} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Bed} & \text{Nurse} & \text{HR} \end{array} \\ \begin{bmatrix} p_{BB} & p_{BN} & p_{BHR} \\ p_{NB} & p_{NN} & p_{NHR} \\ - & - & p_{HRHR} \end{bmatrix} \end{array}$$

Transition probability matrix of Queueing Network Diagram

For a three-state model, the initial probability matrix = $[\pi_B\ \pi_N\ \pi_{HR}]$

Here, for every single state S_i the probability is π_i, where

$$\pi_i = \frac{\text{Number of training sequence starting with i}}{\text{Total number of training sequence}} \quad (10)$$

$$\text{State transition matrix } A = a_{ij} = \begin{bmatrix} a_{BB} & a_{BN} & a_{BHR} \\ a_{NB} & a_{NN} & a_{NHR} \\ a_{HRB} & a_{HRN} & a_{HRHR} \end{bmatrix}$$

For every single state S_i the transition probability to state S_j is a_{ij}, where

$$a_{ij} = \frac{\text{Number of transition from i to j}}{\text{Total number of transitions from i to all states}} \quad (11)$$

Steady state probability, $\pi_i A = \pi_i$ (12)

B. Hidden Markov model

Hidden Markov model is a type of Markov chain which uses Markov process that accommodate hidden and unknown parameters to identify hidden parameters. The states cannot be right away observed but can be linked by considering the visible states.

The relationship between the patients and nurses may be modelled as a two dimensional random variable X(N, IP). Let there be i = 1 to n nurses in the station and j= 1 to m patients under admission. Here the hidden states are nurse and bed. The inpatient (IP) are visible. Hidden states S={nurse, bed}. Visible state $V_1=IP_1$, $V_2=IP_2$, $V_3 = IP_3$

Emission matrix	Initial matrix	State transition matrix
$\begin{array}{c} \\ \text{N} \\ \text{B} \end{array} \begin{array}{c} \begin{array}{ccc} V_1 & V_2 & V_3 \end{array} \\ \begin{Bmatrix} p_{NIP_1} & p_{NIP_2} & p_{NIP_3} \\ p_{BIP_1} & p_{BIP_2} & p_{BIP_3} \end{Bmatrix} \end{array}$	$\begin{array}{c} \\ \text{IP} \end{array} \begin{array}{c} \begin{array}{cc} \text{N} & \text{B} \end{array} \\ \{p_N \quad p_B\} \end{array}$	$\begin{array}{c} \\ \text{IP} \\ \text{N} \\ \text{B} \end{array} \begin{array}{c} \begin{array}{cc} \text{N} & \text{B} \end{array} \\ \begin{Bmatrix} p_N & p_B \\ p_{NN} & p_{NB} \\ p_{BN} & p_{BB} \end{Bmatrix} \end{array}$

The conditional probability function of N given $IP=y_j$.

$$p(N = n_i | IP = y_j) = \frac{p(N=n_i, IP=y_j)}{p(Y=y_j)} \quad (13)$$

(i.e.). the probability of the nurse $N= n_i$ attending the $IP=y_j$ is the j^{th} inpatient call for the i^{th} nurse is given by equation (13). the event of discharge of a patient the nurse calls the j^{th} patient after getting assigned to discharge duty.

$$p(IP = y_j | N = n_i) = \frac{p(IP=y_j, N=n_i)}{p(N=n_i)} \quad (14)$$

If k patients call for a service from nurse station assuming 'l' stationed nurses are available then we have

$$p(IP = K | N = l) = \frac{p(IP=K, N=l)}{p(N=l)} \quad (15)$$

If there is a nurse service requirement then

$$p(N) = p(IP)p(N|IP) + p(A)p(N|A) \quad (16)$$

$$p(N) = \sum_{j=1}^{m} \sum_{i=1}^{n} p(IP = y_j) p(N = n_i | IP = y_j) + p(A = a_i) p(N = n_i | A = a_i) \quad (17)$$

An extension of the hidden Markov model (HMM) is represented by the hidden semi-Markov model (HSMM), which allows the underlying process to be a semi-Markov chain with variable durations or sojourn times for each state. Therefore, in addition to the notation defined for the HMM, the duration ‘d’ of a given state is explicitly defined for the HSMM. State duration is a random variable that assumes an integer value in the set D = {1, 2,..., d}. The significant difference between HMM and HSMM is that, in an HMM, one observation per state is assumed, while in an HSMM, each state can emit a sequence of observations. The number of observations produced while in state iii is determined by the length of time spent in state ‘i’; thus, the duration ‘d’ can be considered for the extension of this work.

C. Routing probabilities

If the number of patients in admission attended by a nurse is a_j then consider the sequence $\{X_1 = a_1, X_2 = a_2, X_3 = a_3, X_4 = a_4, \ldots, X_n = a_n\}$.

Proposing an HMM, we consider the hidden states as nurses and beds, while the visible states are determined by the patients.

Case 1: In this section the utility of bed resource and its probability are given.

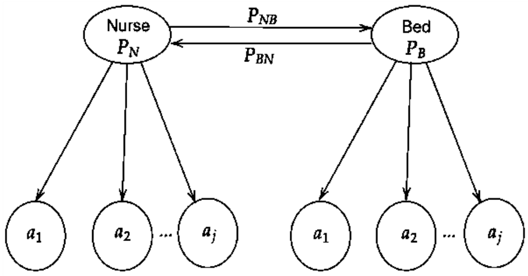

Fig. 1. Flow diagram of ‘a’ patient utilizing the bed

The i^{th} observed output is $X_i = a_j$

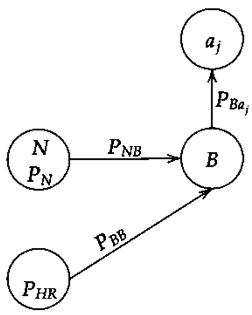

Fig. 2. Flow diagram of j^{th} patient utilizing the bed

$$p(a_j) = [(p_N p_{NB}) p_{Ba_j}] + [\,(p_B p_{BB}) p_{Ba_j}] \quad (18)$$

Here a_j – patients are under admission to the hospital

B to B – denotes that the inpatient continues to be in admission today also.

N to B – denotes the new patient getting admission to bed. Both cases contribute to the admission strength.

Case 2: Here the admission and discharge process are illustrated using a flow diagram. The probability of having an admission or discharge is also obtained.

HR is in-charge of putting admission and sending to nurse for bed assignment. After the completion of treatment nurse sends the patient to HR for discharge procedure.

$$p_{adp}(a_j) = \left[(p_N p_{NNU}) p_{NUa_j}\right] + \left[(p_{HR} p_{HRNU}) p_{NUa_j}\right] \quad (19)$$

Here a_j – denotes j^{th} patient is admission

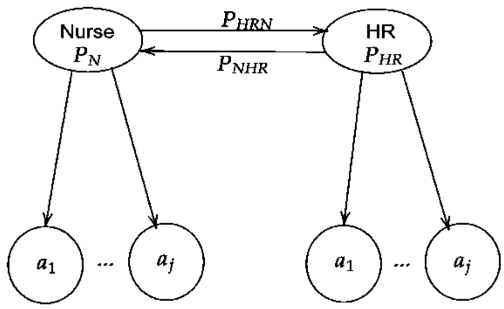

Fig. 3. Flow diagram of ‘a’ patient from nurse and HR

Total number of patients utilizing nursing care unit (NU) at any time instant: S= {N, HR}, V = $\{a_1, a_2, \ldots, a_j\}$.

The i^{th} observed output is $X_i = a_j$

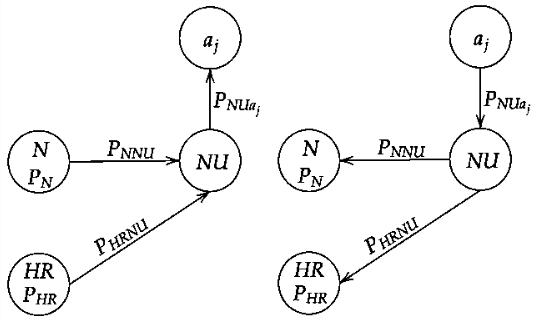

Fig. 4. Flow diagram of j^{th} patient getting admission and discharge

$$p_{dis(a_j)} \left[(p_N p_{NNU}) p_{NUa_j}\right] + \left[(p_{HR} p_{HRNU}) p_{NUa_j}\right] \quad (20)$$

Here a_j – denotes j^{th} patient is discharge

IV. STUDY OF PATIENT NAVIGATION

A. Navigation of two patients through two modules HR and nurse

When two patients are handled by the two modules HR and nurse. By [17], the patient flow from HR to nurse has the state space {(1,1), (0,2), (2,0)}. The two patients may go to HR or nurse to seek admission. The probability of choosing the path that leads to HR is denoted as p and the probability of choosing the path which leads to nurse denoted as q. Two transitions in same path simultaneously is not possible only one transition is possible.

Assume that the system is in the (1,1) state, which means that both HR and nurse are occupied in two ways by both the patients. So, the probability is 2pq.

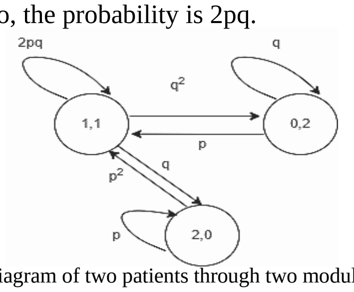

Fig. 5. State diagram of two patients through two modules

$$\begin{array}{c c} & \begin{array}{ccc} (1,1) & (0,2) & (2,0) \end{array} \\ \begin{array}{c} (1,1) \\ (0,2) \\ (2,0) \end{array} & \begin{bmatrix} 2pq & q^2 & p^2 \\ p & q & 0 \\ q & 0 & p \end{bmatrix} \end{array}$$

Transition probability matrix of two patients

Similarly, the probability that a patient requesting admission in bed is directed by HR to nurse is q^2. The probability that a patient being discharged from the bed has to be reported to HR for discharge process by the nurse is p^2. Two independent transition in same path simultaneously is not possible so the probability of (2,0)→(0,2) is 0.

The steady state probability vector is given by

$$\pi_{(1,1)} = 2pq\pi_{(1,1)} + p\pi_{(0,2)} + q\pi_{(2,0)} \quad (21)$$

$$\pi_{(0,2)} = q^2\pi_{(1,1)} + q\pi_{(0,2)} \quad (22)$$

$$\pi_{(2,0)} = p^2\pi_{(1,1)} + p\pi_{(2,0)} \quad (23)$$

From (22), $\pi_{(0,2)} = \frac{q^2}{1-q}\pi_{(1,1)}$ (24)

From (23), $\pi_{(2,0)} = \frac{p^2}{1-p}\pi_{(1,1)}$ (25)

$$\pi_{(1,1)} + \pi_{(0,2)} + \pi_{(2,0)} = 1 \quad (26)$$

Substituting the equation (24) and (25) in equation (26), we get, $\pi_{(1,1)} = \frac{pq}{1-2pq}$ (27)

The steady state probability vector $\pi = [\pi_{(1,1)}\ \pi_{(0,2)}\ \pi_{(2,0)}] = [\pi_1\ \pi_2\ \pi_3]$ (28)

(i.e.) $\pi_1 = \pi_{(1,1)}, \pi_2 = \pi_{(0,2)}, \pi_3 = \pi_{(2,0)}$

$\pi(z)$- Probability generating function of π_i, where π_i denotes the steady state probability of i patients in the system

$$\pi(z) = \sum_{i=0}^{2} \pi_i z^i \quad (29)$$

Substitute (21), (22) and (23) in equation (29), we get

$$\pi(z) = [2pq\pi_{(1,1)} + p\pi_{(0,2)} + q\pi_{(2,0)}]z^0 + [q^2\pi_{(1,1)} + q\pi_{(0,2)}]z^1 + [p^2\pi_{(1,1)} + p\pi_{(2,0)}]z^2 \quad (21)$$

By solving equation (21), we get

$$\pi(z) = \pi_{(1,1)}\left[2pq + \frac{pq^2}{1-q} + \frac{p^2q}{1-p} + \frac{q^2z}{1-q} + \frac{p^2z^2}{1-p}\right] \quad (30)$$

Taking the derivative and substituting z=1 in equation (22), we get

$$E[N] = \pi'(1) = \frac{q^3+2p^3}{1-2pq} \quad (31)$$

B. Navigation of three patients through two modules HR and nurse

The patient flow from HR to nurse has the state space {(3,0), (2,1), (1,2), (0,3)}. Due to the restriction imposed those two transitions in the same path simultaneously is not possible. The probability of choosing the path that leads to HR denoted as p and which leads to nurse denoted as q.

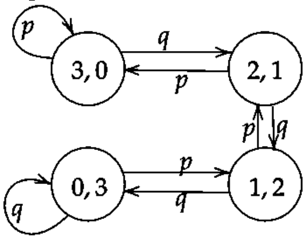

Fig. 6. State Diagram of three patients through two modules

$$\begin{array}{c} \\ (3,0) \\ (2,1) \\ (1,2) \\ (0,3) \end{array} \begin{array}{c} \begin{array}{cccc} (3,0) & (2,1) & (1,2) & (0,3) \end{array} \\ \begin{bmatrix} p & q & - & - \\ p & - & q & - \\ - & p & - & q \\ - & - & p & q \end{bmatrix} \end{array}$$

Transition Probability Matrix of three patients

Following the same procedure as in two patient navigation and solving it ,

we get

$$\pi(z) = \pi_{(3,0)}[1 + \frac{q}{p}z + \frac{q^2}{p^2}z^2 + \frac{q^3}{p^3}z^3] \quad (32)$$

Taking the derivative and substituting z=1 in equation (32), we get

$$\pi'(1) = \frac{qp^2+2q^2p+3q^3}{p^2+q^2} \quad (33)$$

In Section IV patient movement between two modules, HR and nurse is discussed. The system's routing probabilities, steady state probability distribution are discussed and derived.

V. NUMERICAL RESULTS

TABLE I
PERFORMANCE MEASURES AS λ_n VARIES

λ_n	Mean nursing assistants	L_{sys}	$L_{non-prior}$
1.0	3.026636	1.806166	1.551310
1.1	2.966979	1.779459	1.516780
1.2	2.909075	1.753314	1.483363
1.3	2.853094	1.727871	1.451130
1.4	2.799088	1.703194	1.420093
1.5	2.747088	1.679333	1.390254
1.6	2.697077	1.656307	1.361590
1.7	2.649070	1.634153	1.334098
1.8	2.603001	1.612855	1.307732
1.9	2.558814	1.592403	1.282454

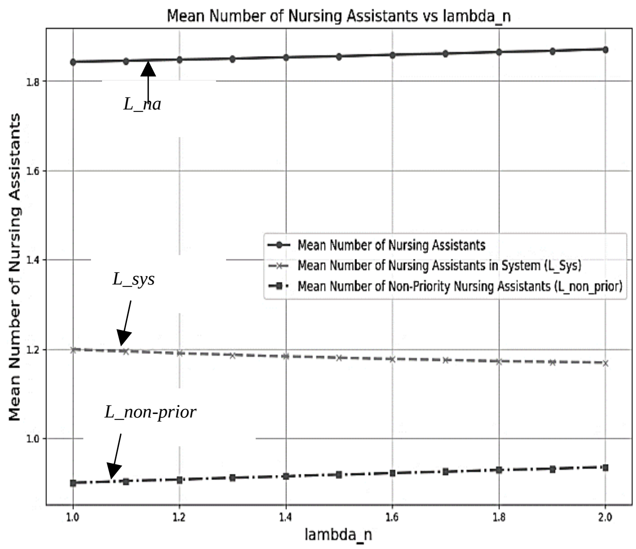

Fig. 7. λ_n versus mean number of nursing assistants

As λ_n is increased, more dynamic state changes with higher nurse arrival rates are indicated. The stationary distributions (π_0 and π_1) are shifted with increasing λ_n, influencing the equilibrium states of the system. The mean number of nursing assistants is decreased as λ_n is increased, indicating higher nurse activity and reduced availability of nursing assistants on average. The decreasing numbers in buffer (L_Sys) and non-prior (L_non_prior) indicate a shift in staffing dynamics towards meeting immediate demands efficiently. A comprehensive understanding of the system's behavior under different scenarios is offered by the data analysis, aiding in optimizing staffing strategies for efficient healthcare delivery. Valuable insights into how changes in nurse activity levels affect the availability and distribution of nursing assistants in different roles are provided by the model.

The average time a nursing assistant spends in the system, from arrival to departure, decreases as λ_n increases. This suggests that the system becomes more efficient at handling arrivals quickly.

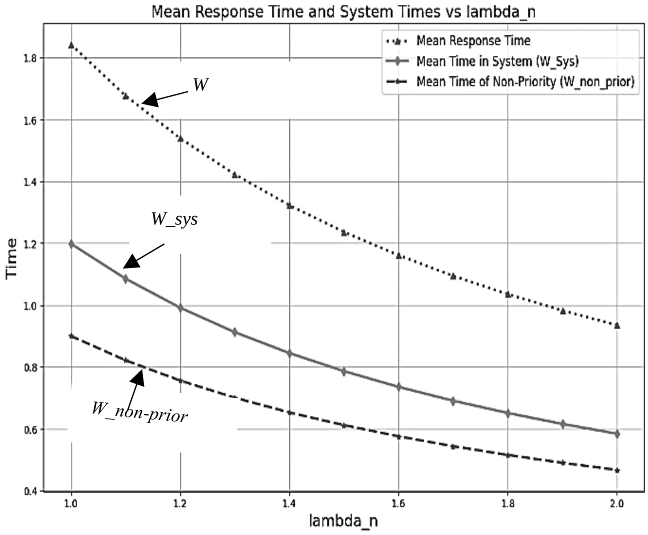

Fig. 8. λ_n versus mean response time

The mean time in the system, which is the average time a nursing assistant waits in the system, also decreases with higher λ_n, indicating better system management and faster processing times as nurse arrival rates rise.

Similarly, the mean time of non-priority, or the average time non-priority nursing assistants spend in the buffer, decreases as λ_n increases. This demonstrates that the system manages non-priority cases more efficiently under higher loads. The CTMC model is valuable for understanding the dynamic behavior and long-term performance of a system with multiple states and transitions.

The integration of CTMC into the Quasi-birth-and-death (QBD) process is beneficial for several reasons. The real-time dynamics of patient, nurse, and assistant interactions are captured by the CTMC model, which is crucial for analyzing system responses to changes and the time taken to reach steady states. The exact rates and transitions between different states are provided by the CTMC model, aiding in the formulation of the generator matrix Q. The steady-state probabilities derived from the CTMC model inform the long-term behavior of the system, critical for resource optimization and efficiency improvement in the QBD framework. Additionally, the integration helps in understanding the evolution of different states over time, assisting in more effective resource allocation and planning. The QBD process can be validated by comparing the predicted state probabilities with actual observations, and the system can be optimized by identifying bottlenecks and areas for improvement.

TABLE II
STEADY STATE PROBABILITIES OVER TIME

Time (hours)	State 0	State 1	State 2	State 3	State 4	State 5	State 6	State 7
0	0.7000	0.3000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.3784	0.2833	0.1428	0.1069	0.0406	0.0304	0.0101	0.0075
16	0.3344	0.2508	0.1516	0.1137	0.0614	0.0460	0.0241	0.0180
24	0.3178	0.2383	0.1523	0.1142	0.0697	0.0523	0.0317	0.0237
32	0.3106	0.2329	0.1524	0.1143	0.0733	0.0550	0.0352	0.0264
40	0.3074	0.2305	0.1524	0.1143	0.0749	0.0562	0.0368	0.0276
48	0.3059	0.2295	0.1524	0.1143	0.0756	0.0567	0.0375	0.0281
56	0.3053	0.2290	0.1524	0.1143	0.0759	0.0569	0.0378	0.0284
64	0.3050	0.2287	0.1524	0.1143	0.0761	0.0571	0.0380	0.0285
72	0.3049	0.2287	0.1524	0.1143	0.0761	0.0571	0.0380	0.0285
80	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
88	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
96	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
104	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
112	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
120	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
128	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
136	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
144	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
152	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
160	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286
168	0.3048	0.2286	0.1524	0.1143	0.0762	0.0571	0.0381	0.0286

Given the parameters and the generator matrix Q, the states can be interpreted as follows: State 0 represents both nurse and assistant being free, State 1 indicates the nurse is occupied while the assistant is free, State 2 shows the nurse is free but the assistant is occupied, State 3 denotes both are occupied, State 4 involves the assistant arriving with the nurse remaining free, State 5 means the assistant arrives and the nurse becomes occupied, State 6 signifies the nurse leaves while the assistant remains, and State 7 reflects both leaving.

Initially, at time 0 hours, the system is most likely in state 0(70%) or state 1 (30%). Overtime, the probabilities of different states evolve. For example, at 8 hours, state 0 has a probability of 37.84%, State 1 has 28.33%, State 2 has 14.28% and so on.

These probabilities gradually stabilize, with the system reaching equilibrium by 168 hours, showing probabilities of 30.48% for State 0, 22.86% for State 1, 15.24% for State 2, among others.

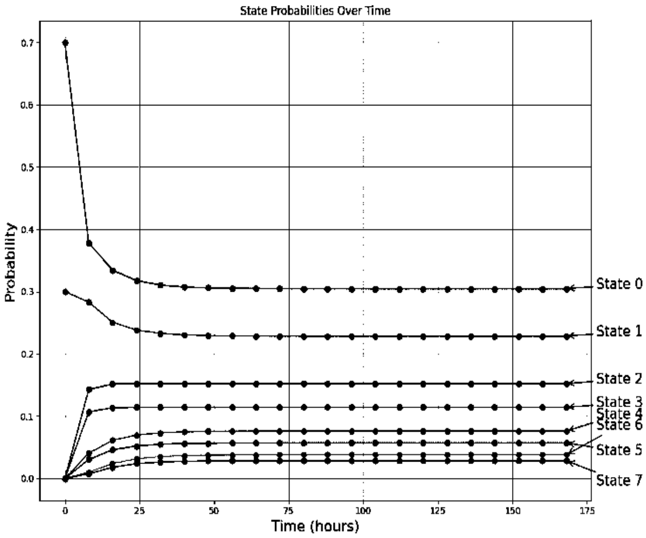

Fig. 9. State probabilities over time

Figure 9 describes the steady-state probabilities indicate the long-term likelihood of each state, with State 0 being the most probable at 30.48%, suggesting both nurse and assistant are often free. This distribution helps in planning and resource allocation by highlighting which states are more likely, thus ensuring adequate staffing and resources.

Overall, the system transitions smoothly and the state probabilities stabilize, reflecting predictable behaviour in the long run. The high initial probabilities of states with fewer transitions spread over time due to ongoing transitions, illustrating the model's dynamic nature. This robust CTMC framework aids in analysing nurse and assistant dynamics, enabling efficient resource management and optimization in healthcare settings, ultimately improving efficiency, resource utilization, and patient care outcomes.

After the nurse and nurse-assistant relationship and the patient's concern are discussed, his wait for service is optimized. An optimal routing path illustrating the sequence of discharge and admission during a specified time interval is now provided. This would enable the most efficient rate of functioning.

A. *Viterbi Algorithm*

The Viterbi algorithm is employed to determine the most probable sequence of events. Operating a hospital is a challenging task as it requires adherence to multiple regulations set by medical councils, strict guidelines, and eligibility criteria. Moreover, satisfying patient needs is of utmost importance. To achieve this, hospitals must ensure they have an adequate and well-trained staff capable of meeting patient requirements. In this context, the two types of staff considered are HR (Human Resource) and nurse. The HR staff's responsibilities encompass collecting patient data, scheduling appointments, and managing billing for patient care. Nurses, on the other hand, are responsible for assisting patients to their beds, monitoring their health and needs, and providing assistance to doctors in emergency situations. The admission of an inpatient signifies their arrival, while the discharge or return of an inpatient to HR represents their departure. The hidden states in this scenario are HR and nurse. While the hospital needs to assess the quality of service provided by the nurses and HR staff, this information remains hidden. Hence, the admission and discharge events discussed earlier are considered as observation states. The parameters required to calculate the optimal path for this analysis are depicted in the context of the situation.

TABLE III
HMM PARAMETERS

	N	HR	A	D
S	0.60	0.40		
N	0.60	0.40	0.50	0.5
HR	0.20	0.80	0.90	0.1

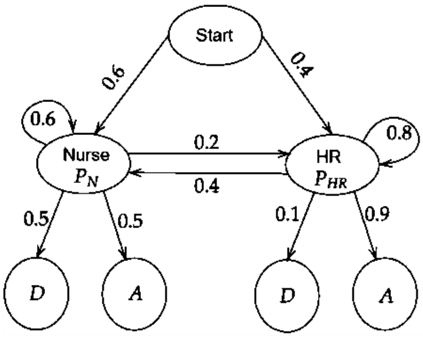

Fig. 10. State diagram of HMM parameters

The following figure pertains to a sequence of order six. The goal is to identify the optimal path for these sequences. For example, consider the occurrence of event A in the first order of the sequence, representing admission.
This is calculated by taking into account the hidden states of HR and nurse as the starting points, multiplying the probability of admission for both. This yields two probability values: one for HR and one for the nurse. In the next step, these initial HR and nurse values are replaced with the calculated probabilities, and the process is repeated. The HR probabilities include transitions from HR to HR and HR to nurse, while the nurse probabilities include transitions from nurse to nurse and nurse to HR. These probabilities are recalculated at each step.

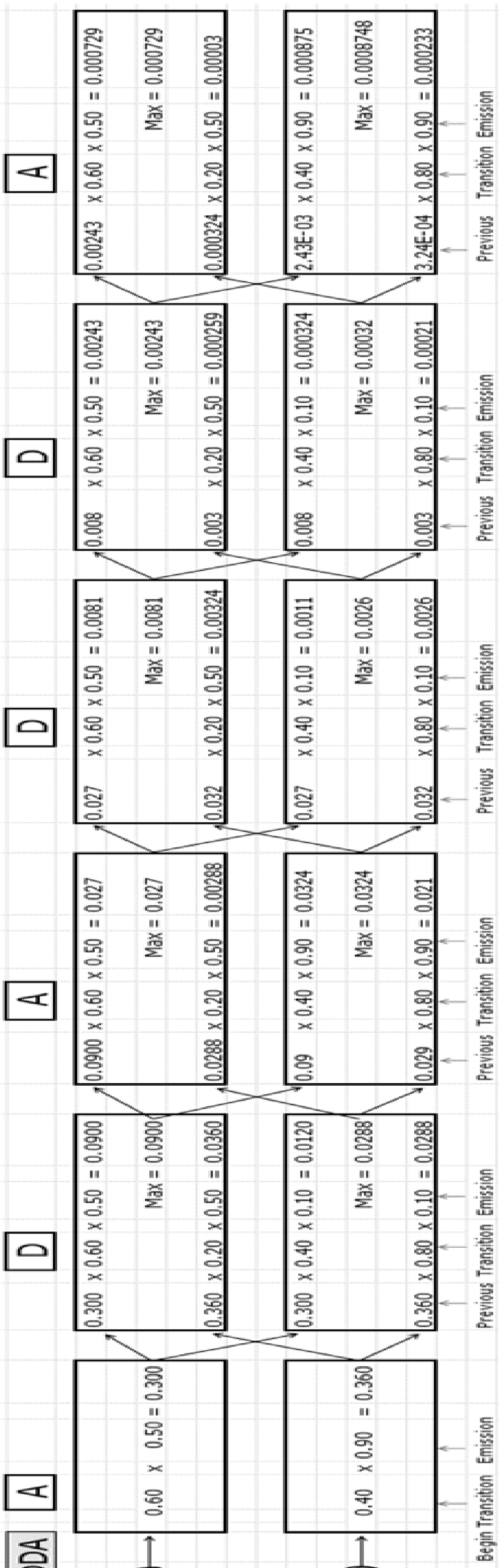

Fig. 11. Viterbi algorithm for sequence of length six

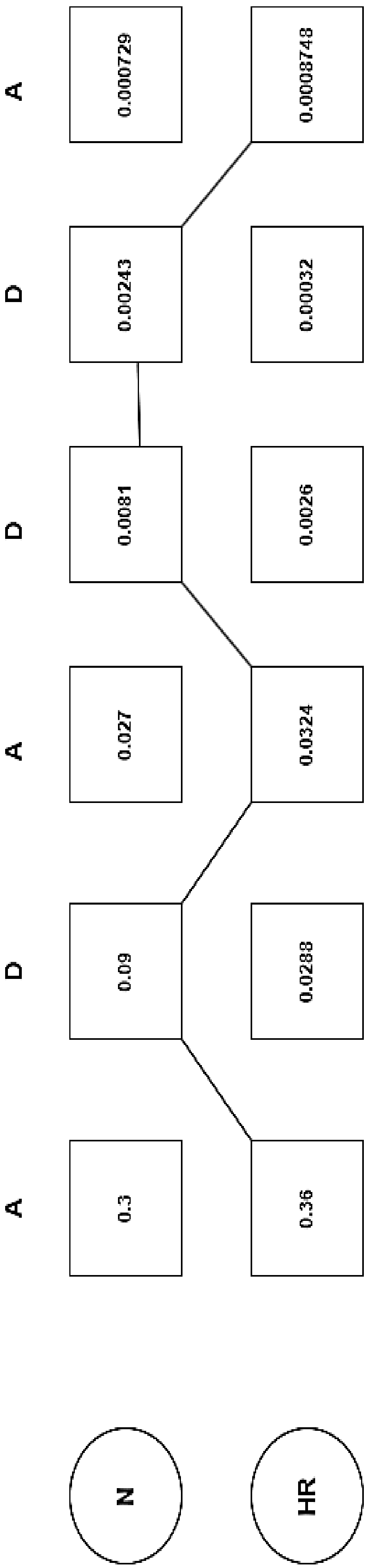

Fig. 12. Optimal path for sequence of length six

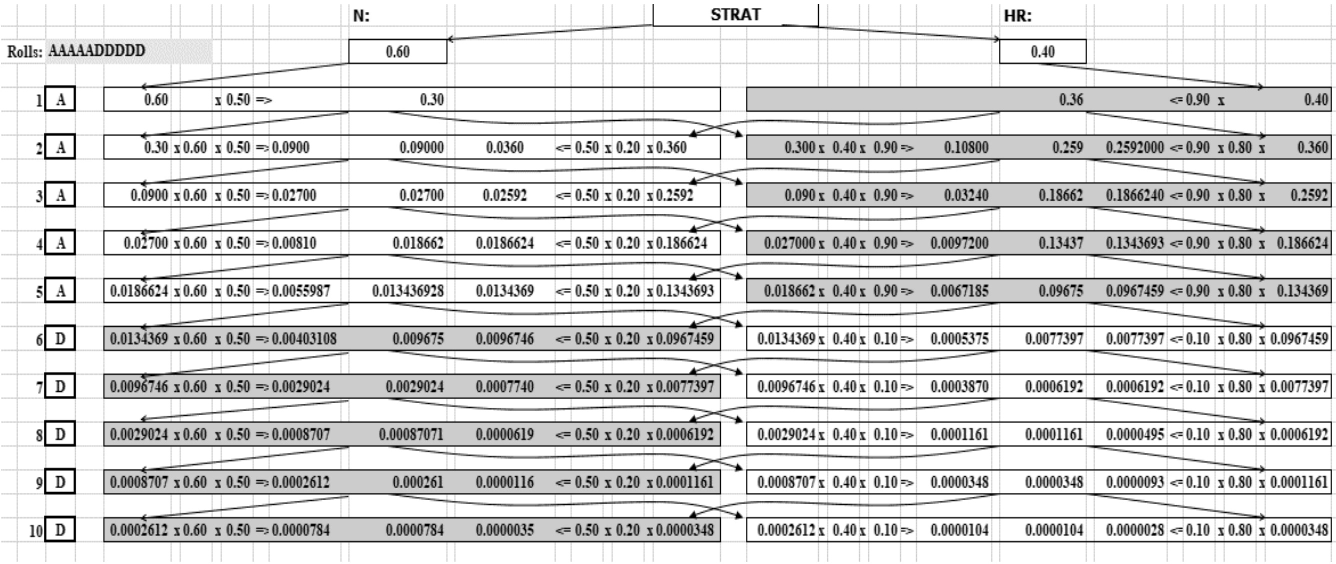

Fig. 13. Viterbi algorithm for sequence of length ten

After each calculation, the maximum value is determined by comparing the two HR possibilities and the two nurse probabilities, using this maximum value in the subsequent step. This procedure continues until the final stage of the sequence is reached. To find the optimal path for the sequence, the calculated maximum values of HR and nurse are compared. By doing so, the most favorable path for the given sequence is determined. Similarly, a sequence of order ten is also evaluated for the optimal path, as shown in Fig. 11.

For the sequence ADADDA, the optimal path is shown in Fig.10.

B. Steady state probability of two patient

In previous section IV A, navigation of two patients through two modules is discussed. From the calculation, we get the following values of equation (27)

(i.e.)$\pi_{(1,1)} = \frac{pq}{1-2pq}$

Substituting this equation in equation (24) and (25), we get the values for

$\pi_{(0,2)} = \frac{q^3}{1-2pq}$ and $\pi_{(2,0)} = \frac{p^3}{1-2pq}$

TABLE IV
STEADY STATE PROBABILITY OF TWO PATIENT

p	q	pi(1,1)	pi(0,2)	pi(2,0)	max value
0.3	0.7	0.3621	0.5914	0.0466	0.5914
0.35	0.65	0.4174	0.5039	0.0787	0.5039
0.4	0.6	0.4615	0.4154	0.1231	0.4615
0.45	0.55	0.4901	0.3295	0.1804	0.4901
0.5	0.5	0.5	0.25	0.25	0.5
0.55	0.45	0.4901	0.1804	0.3295	0.4901
0.6	0.4	0.4615	0.1231	0.4154	0.4615
0.65	0.35	0.4174	0.0787	0.5039	0.5039
0.7	0.3	0.36207	0.0466	0.5914	0.5914

C. Steady state probability of three patient

In Previous chapter Section IV.B, navigation of three patients through two modules is discussed. From the calculation, we get the following values

(i.e.,)$\pi_{(3,0)} = \frac{p^3}{p^2+q^2}$, $\pi_{(2,1)} = \frac{p^2q}{p^2+q^2}$, $\pi_{(1,2)} = \frac{pq^2}{p^2+q^2}$,
$\pi_{(0,3)} = \frac{q^3}{p^2+q^2}$.

TABLE V
STEADY STATE PROBABILITY OF THREE PATIENT

p	q	π(3,0)	π(2,1)	π(1,2)	π(0,3)	max value
0.3	0.7	0.0466	0.1086	0.2535	0.5914	0.5914
0.35	0.65	0.0787	0.1461	0.2713	0.5039	0.5039
0.4	0.6	0.1231	0.1846	0.2769	0.4154	0.4154
0.45	0.55	0.18045	0.2205	0.2695	0.3295	0.3295
0.5	0.5	0.25	0.25	0.25	0.25	0.25
0.55	0.45	0.3295	0.2696	0.2206	0.1805	0.3295
0.6	0.4	0.4154	0.2769	0.1846	0.1231	0.4154
0.65	0.35	0.5039	0.2713	0.1461	0.0787	0.5039
0.7	0.3	0.5914	0.2535	0.1086	0.0466	0.5914

The steady state probabilities are obtained by varying the value of p and is depicted in the table IV and table V. Maximum probability occurs shows the most freequently occuring event.

VI. COST OPTIMIZATION

Now, the cost function is defined as follows:

Cost = C_nurse * N + C_hr * H + C_waiting * Satisfaction (34)

where, C_nurse * N, represents the cost associated with nurses. C_hr * H, represents the cost associated with HR staff. C_waiting * Satisfaction, represents the cost associated with patient satisfaction.

Here

- N: Number of nurses.
- HR: Number of HR staff.
- C_nurse: Cost per nurse per unit time.
- C_hr: Cost per HR staff per unit time.
- C_waiting: Cost associated with patient waiting time.
- Satisfaction: A measure of patient satisfaction, typically ranging from 0 (dissatisfied) to 1 (satisfied).

(34) is a cost that accounts for the degree of patient satisfaction. The higher the satisfaction (closer to 1), the lower the associated cost. The balance between staffing levels and patient satisfaction is achieved by optimizing the cost function.

A. Optimization by genetic algorithm

The main staffing issue is that we need to establish the number of nurses (N) and HR personnel (HR) needed to minimize staffing expenses while maintaining patient satisfaction.

Cost = C_nurse * N + C_hr * HR + C_waiting * Satisfaction is the cost formula.

Genetic algorithm's (GA) are population-based optimization algorithms with natural selection as their primary inspiration. A problem's specific parameters, such as population size, chromosome length (which represents solutions), and the number of generations (or iterations), are first defined by the GA code.

Additionally, it establishes constants like the costs associated with hiring nurses, HR personnel, and patients. The population is started with chromosomes drawn at random. A certain number of generations are spent in the primary GA loop. Based on the cost formula, the fitness function assesses the fitness of each chromosome. Better solutions tend to be more affordable. Crossover parents are chosen according on their physical health. Better solutions are more likely to be chosen. Crossover mixes the offspring of two sets of parents. Random modifications are introduced in offspring solutions via mutation. The old population is replaced by the new population (offspring).

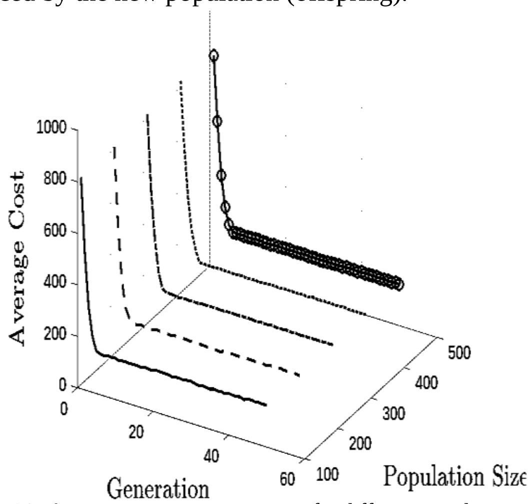

Fig. 14. Average cost over generations for different population size

The optimal staffing solution (N and HR) is displayed.

To illustrate the optimization procedure, a MATLAB code has been developed to generate three plots: one depicting the optimal average cost, and the others showcasing the progression of population size over generations (Fig.12).

B. Optimization by ant colony optimization

A staffing issue akin to GA, with the same cost formula and the aim of reducing staffing expenses while considering patient happiness, is discussed.

Ant colony optimization (ACO), inspired by ant foraging behavior, is particularly suitable for pathfinding or routing challenges. The parameters initially defined in the code include the number of ants, iterations, initial value and decay rate of pheromones, as well as influence parameters (alpha and beta). Additionally, it specifies cost factors for nurses, HR personnel, and patient satisfaction. The main ACO loop runs for a set number of iterations. Paths are created for each ant in each iteration, representing various nurse and HR combinations as staffing options. The cost formula assigns a score to each path, with lower scores indicating better staffing olutions. Pheromone values are updated based on the path

scores, with higher updates for paths with lower costs. The most cost-effective route is regarded as the ideal one. The optimized staffing solution (N and HR) is displayed in Fig.13. A graph plot is created, in order to visualize the best score history over iterations.

The task is to compare the performance of the ant colony optimization (ACO) and genetic algorithm (GA) for varying population sizes or generations (from 100 to 500). Whereas GA employs 2 to 3 HR staff and 2 to 9 nurses, ACO uses 2 to 5 HR staff and assigns 3 to 8 nurses.

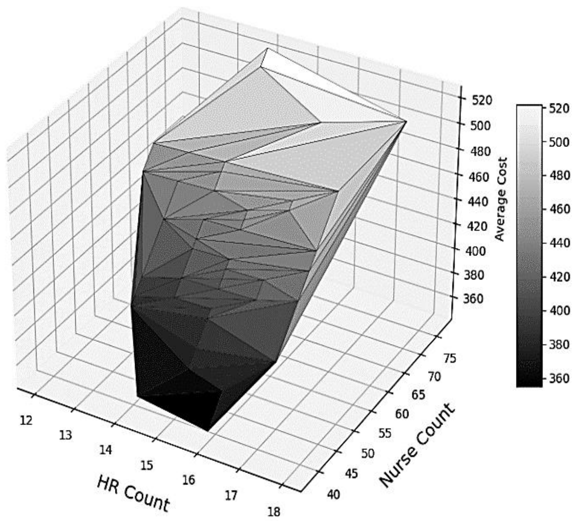

Fig. 15. Surface plot of HR, nurse and Average cost

Based on a patient satisfaction scale of 100, costs are determined by allocating $100 for each nurse, $200 for HR staff, and $50 for patient waiting time. The entire costs of GA are between $5600 and $6300, and the total expenses of ACO are between $5700 and $6800. According to the data, ACO typically has greater expenses than GA, which may indicate variations in resource or efficiency. To discover the most economical solution while keeping patient happiness, more optimization is advised.

TABLE VI
COMPARISON OF ACO AND GA FOR COST FUNCTION

Population /Ants	Generation /trials	ACO	GA	Conclusion
100	100	2(HR)+3(N)	2(HR)+2(N)	GA is better
150	150	3(HR)+9(N)	3(HR)+4(N)	GA is better
200	200	4(HR)+4(N)	3(HR)+6(N)	ACO is better
300	300	5(HR)+3(N)	3(HR)+5(N)	ACO is better
500	500	5(HR)+8(N)	2(HR)+9(N)	GA is better

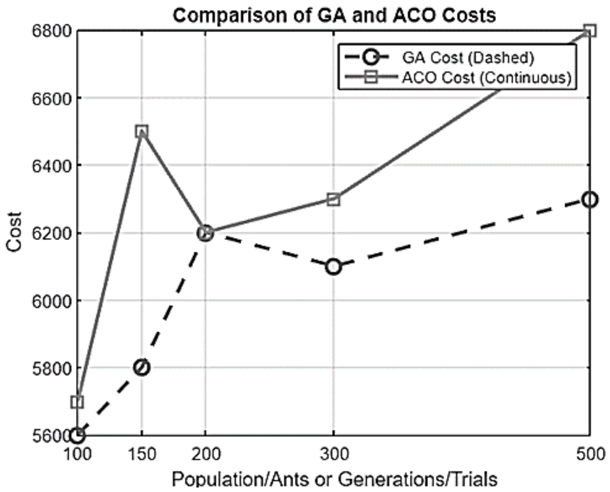

Fig. 16. GA and ACO HR and Nurse comparison

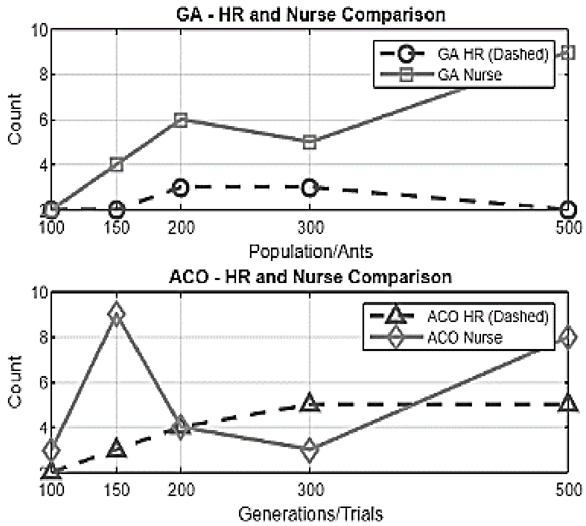

Fig. 17. Cost comparison of GA and ACO

C. Optimization by Markov decision process

A Markov decision process (MDP) is a mathematical framework used for modeling decision-making in scenarios where outcomes are partially random and partially under the control of a decision-maker. It consists of states, actions, transition probabilities, and rewards. MDPs are used to determine the optimal policy, which is a strategy for selecting actions to maximize cumulative rewards or minimize costs over time. In our scenario, MDP is applied to optimize staffing decisions for HR and nurses. The model is set up with states representing various combinations of the number of nurses and HR staff. Actions include increasing or decreasing the number of nurses and HR staff. Transition probabilities are used to describe the likelihood of moving from one state to another based on the chosen action, while rewards are derived from the cost function, which incorporates the costs associated with staffing and patient satisfaction.

The decision-making process is guided by the value iteration algorithm, which uses the Bellman equation to iteratively update the value function and determine the optimal policy. The Bellman equation is expressed as:

$$\mathrm{V}(s) = \max_{a}[\sum_{s'} P(s,a,s')(R(s,a) + \gamma V(s')]$$

where V(s) represents the value of state s, a represents actions, P(s,a,s′) is the transition probability, R(s,a) is the reward, γ is the discount factor.

The value iteration algorithm is used to find the optimal policy. It iteratively updates the value function (V) until it converges to the optimal values. The policy for each state is then determined based on the maximum expected reward. Rewards (R) are the immediate costs incurred when performing a particular action in a given state. They are defined as the negative of the total cost, which includes cost of nurses, cost of HR staff, cost associated with patient waiting time (which is based on satisfaction).

The output is summarized in the policy table, which details the optimal action for each state. Table VII gives the values correspond to action indices where 0: Increase nurse, 1: Decrease nurse, 2: Increase HR, 3: Decrease HR

TABLE VII
POLICY TABLE

Number of HR Staff	1	2	3
Number of nurses	Policy Actions		
10	3	1	2
20	2	0	2
30	2	1	1

The policy table suggests that, generally, increasing HR staff is the optimal action across various staffing scenarios. However, specific conditions indicate that decreasing the number of nurses can also be beneficial. For instance, with 20 nurses and 2 HR staff, the optimal action shifts to decreasing the number of nurses. This table and the accompanying policy insights help guide effective staffing decisions to minimize overall costs and enhance patient satisfaction.

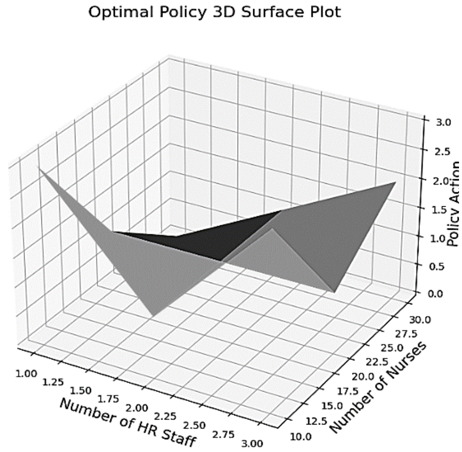

Fig. 18. Cost optimization by MDP

VII. ILLUSTRATION

In a queueing system, that resembles the workflow of a medical clinic or hospital, four stations are involved: HR (Human Resources), Nurse, Doctor, and Bed. Patients move through these stations, receiving services at each one, and then returning to the starting point, which is the HR station. Patients enter the system and follow a predefined pattern as they move through the stations. Patients initially arrive at the HR desk to register and provide necessary information. At the HR Desk, they may complete paperwork, update personal information, or schedule appointments.

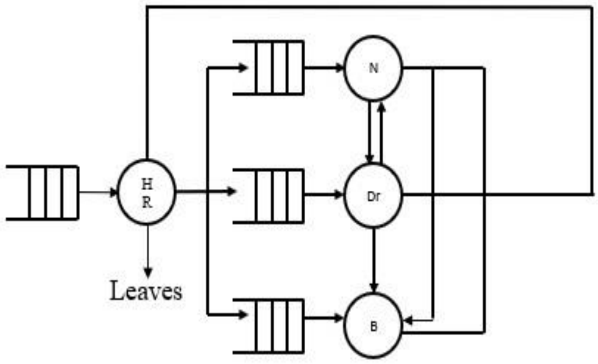

Fig. 19. Network diagram of a 4-station

After registering, patients proceed to the nurse station for an initial check-up and assessment. The nurse may record vital signs, perform basic examinations, and gather additional medical history. Following the nurse's assessment, patients move to the doctor consultation area for a more thorough examination and treatment plan. The doctor evaluates the patient's condition, provides medical advice, prescribes medications if necessary, and recommends further tests or procedures. In some cases, patients may require immediate care or observation. They are then directed to the Bed Area for treatment, monitoring, or rest. After completing the necessary steps, patients return to the HR desk to finalize paperwork, schedule follow-up appointments, or exit the system.

For routine cases, patients follow the usual loop with service rates:

- HR Desk: Mean service duration $\frac{1}{\mu_1}$ for registration and paperwork.
- Nurse station: Mean service duration $\frac{1}{\mu_2}$ for quick assessment.
- Doctor consultation: Mean service duration $\frac{1}{\mu_3}$ for thorough examination and treatment plan.
- Bed area: Mean service duration $\frac{1}{\mu_4}$ for immediate care or observation.

Each station has a mean service duration: μ_1 for HR, μ_2 for nurse, μ_3 for doctor, and μ_4 for Bed. The transition probabilities are represented by p_{12}, p_{13}, and p_{14}.

Figure 19 describes the network diagram of a 4-station queueing system.

This system can be modelled as a continuous-time Markov chain (CTMC), where the state space represents the distribution of patients across the four stations. In this model, each state corresponds to a specific combination of patients in stations one, two, three, and four. Transitions between states occur probabilistically, driven by the service rates at each station and the rules governing patient movement. For instance, when a patient completes their service at a station, they may move to another station or exit the system, triggering a state transition.

These probabilistic transitions are depicted in Figure 20, which illustrates how patients flow through the stations over time.

The number of possible states in this CTMC model can be determined using the combinatorial formula for combinations with repetition, given by $(N+k-1)C_{(N-1)}$, where N is the number of stations and k is the number of patients. This formula accounts for all possible ways to distribute k patients among N stations. In the specific case of our model, there are 4 stations and 3 patients.

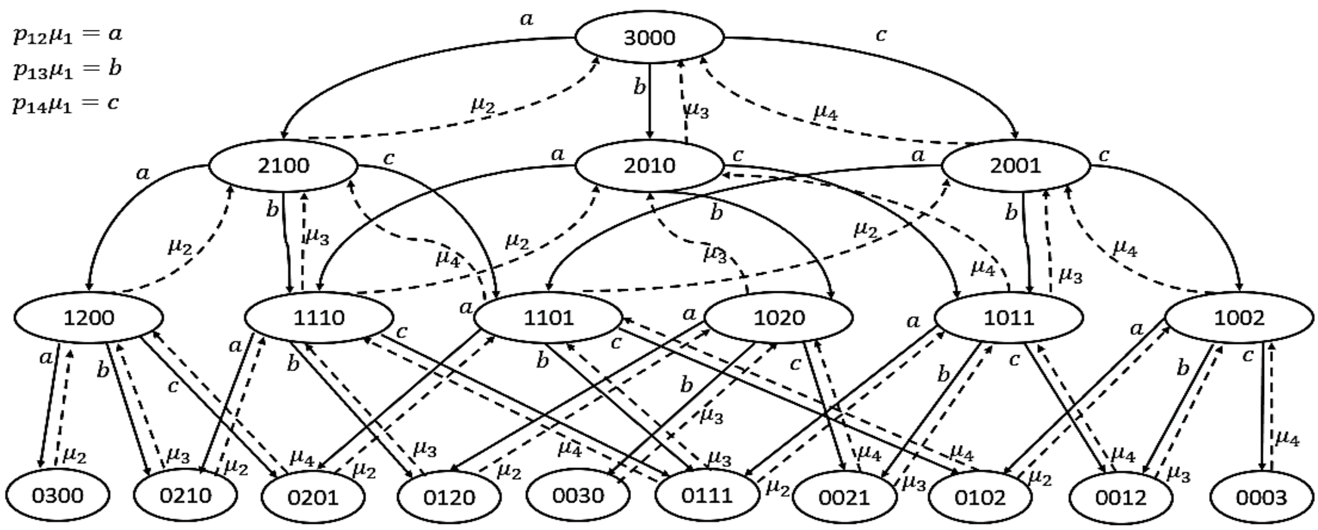

Fig. 20. State transition diagram of 4-station 3-customer queueing network

Plugging these values into the formula, we get (4+3-1)$C_{(3)}$, which simplifies to 6C_3, yielding a total of 20 possible states. This means that there are 20 unique configurations of patients across the four stations that the system can occupy at any given time.
These configurations capture all possible ways that the three patients can be distributed among the four stations, considering every potential combination of patient placements.

Each state in the system can be represented as (k, l, m, n), where k indicates the number of patients in station one, l represents the number of patients in station two, m denotes the number of patients in station three, and n signifies the number of patients in station four, with the condition that k + l + m + n = 3. This means that the total number of patients across the four stations is always three. The 20 possible states that satisfy this condition are as follows: (1, 0, 2, 0), (3, 0, 0, 0), (0, 2, 1, 0), (0, 1, 0, 2), (0, 3, 0, 0), (2, 1, 0, 0), (1, 0, 1, 1), (0, 2, 0, 1), (0, 1, 2, 0), (0, 0, 3, 0), (1, 1, 0, 1), (0, 0, 2, 1), (0, 1, 0, 2), (1, 0, 0, 2), (0, 0, 1, 2), (0, 1, 1, 1), (0, 0, 0, 3), (1, 0, 1, 1), (0, 0, 2, 1), and (0, 1, 0, 2). Each of these states represents a unique distribution of patients among the four stations. A potential sequence of these states, illustrating the transitions between different patient distributions, is depicted in Figure 20.

$$\text{Expected Customers} = \sum_{i=1}^{20} i * \text{Steady State Probability}_i$$

$$\text{Utilization}_i = \sum_{j=i}^{20} \text{Steady State Probability}_j * \min(j - i, 3)$$

Python code is generated to obtain the steady state probabilities of all the 20 states and is listed above in table VIII.

TABLE VIII
STEADY STATE PROBABILITIES

π_1	0.3213	π_{11}	0.0018
π_2	0.1205	π_{12}	0.0026
π_3	0.1208	π_{13}	0.0038
π_4	0.1666	π_{14}	0.0033
π_5	0.0143	π_{15}	0.0053
π_6	0.0186	π_{16}	0.009
π_7	0.0264	π_{17}	0.0048
π_8	0.0241	π_{18}	0.0081
π_9	0.0378	π_{19}	0.0141
π_{10}	0.0691	π_{20}	0.0277

The code is structured to handle arbitrary transition probabilities and service rates. Therefore, we obtain the following performance measures,

Expected number of patients in the system: 3.6024
Expected time a patient spends in the system: 1.2008 time units
Throughput of the system: 3.0000 patients per time unit
Station Utilizations:
Station 1 utilization: 1.3943
Station 2 utilization: 1.0546
Station 3 utilization: 0.8034
Station 4 utilization: 0.6372
Expected Time at Each Station:
Expected time at Station 1: 0.5581 time units
Expected time at Station 2: 0.4221 time units
Expected time at Station 3: 0.3216 time units
Expected time at Station 4: 0.2551 time units

The relationship between service rates and the number of patients in the system is directly proportional. As the service rate increases, the system becomes capable of accommodating more patients, leading to higher expected numbers of patients in the system. In the future, it would be advantageous to employ well-known algorithms to establish bounds for these relationships.

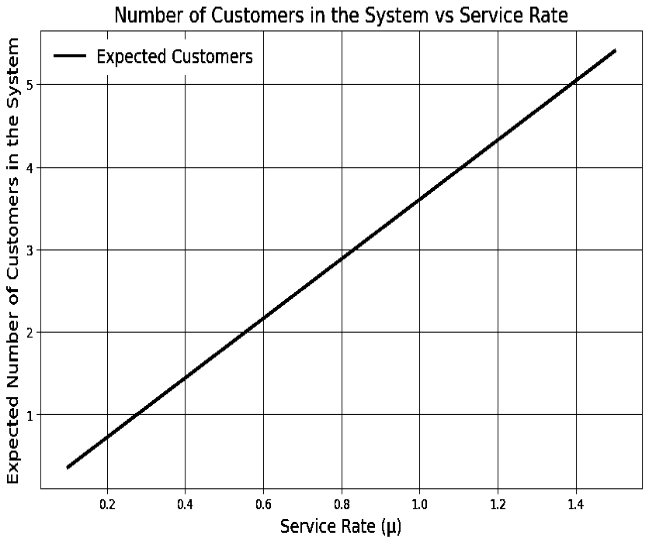

Fig. 21. Number of patients versus service rate

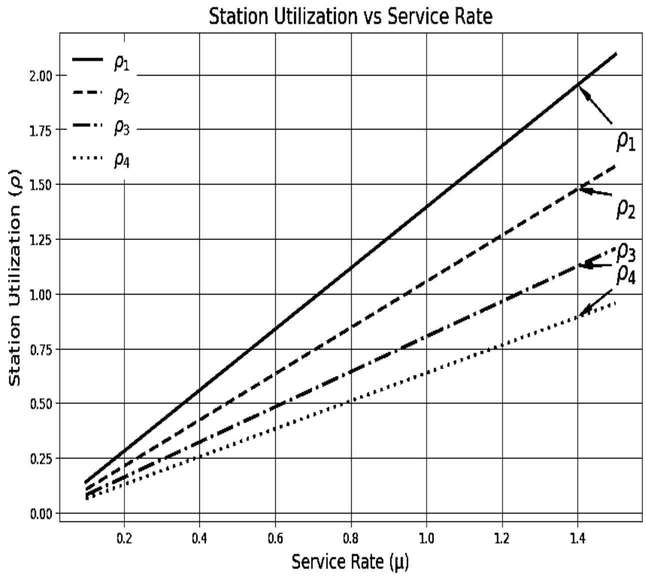

Fig. 22. Server Utilization versus service rate

A. System constraints in hospital care

There are separate queues for each specific patient requirement. Let C be denoted as the number of nurses, K as the number of nursing assistants, and n as the different requirements of patients. The pair corresponds to having C nurses and C+K total availability of manpower for attending patients. Each patient is attended by a single nurse. However, a bed is occupied by the patient even when not being served but under admission. Patient calls that arrive when all C+K staff are occupied are not allowed to enter the system. Nurses are trained to attend to specific tasks such as bedside help, emergency situations, surgery dressings, etc. Assuming they specialize in only one of the skill sets out of k skills, patient calls are attended to only if the requirement matches their skill set. With these restrictions, a routing policy is specified. It is assumed that the hospital nursing facility is well-equipped so that the delay in attending to a patient call is negligible. The goal is to lexicographically minimize (C, K) subject to per class performance constraints.

Our idea is to minimize C i.e., the set of trained well qualified nurses and then for a given C to minimize the nursing assistant K. The skills are treated as part of constraints. Each nurse is assumed to be trained to attend two different situations only. The two main constraints for our study is

i) The speed to answer service level constraint i.e., to minimize the time delay in attending patient call.

ii) Unavailability of resource i.e., turning down admission due to lack of free space.

Q_k be the steady state number of patients in the system, experience by an arrival of type k. The need for attending patients without delay is the atmost requirement. The speed to answer a particular patient call is set to have an upper hound of $\delta_k$time for type k target. For otherwise the patient may be lost. This may be represented as

$$P(W_k \leq \tau | Q_k < C + K) \geq \delta_k, 1 \leq k \leq n \quad (35)$$

τ_k is the type k target where k could be bedside help, emergency care, surgical dressing etc. The probability of turning down a patient admission is

$$P(Q_k = C + K) \leq \epsilon_k, 1 \leq k \leq n \quad (36)$$

Using these two equations we can provide a static priority routing scheme as follows. The nursing staff (trained nurse + assistant) are considered to have 1 to n skills. However, they have their priority based on their level of perfection in addressing issues.

VIII. CONCLUSION

In conclusion, this article has thoroughly explored various aspects of modelling patient flow and optimizing healthcare delivery systems. Section II begins with the Quasi-birth-death (QBD) process, which models patient, nurse-assistant dynamics and system performance measures, detailing how QBD captures staffing levels and patient care requirements. Section III introduces the Markov model, illustrating patient movement within nurse stations through network transition diagrams and steady-state probabilities, and extends this exploration to hidden Markov models with practical case studies.

Section IV studies patient navigation, examining steady-state probabilities under varying conditions and identifying frequent system events. Section V focuses on numerically visualizing performance measures as λ_n varies, introducing different sequence lengths optimized using the Viterbi algorithm in VA determines the most probable event sequences. Section VI discusses cost optimization using heuristic algorithms like genetic algorithms and ant colony optimization, highlighting their effectiveness in cost efficiency and minimizing costs under specific scenarios. The model is further validated using MDP in decision making regarding staff strength.

Section VII analyses scenarios involving four stations, using state transition diagrams to derive and present system evaluation metrics. It addresses hospital care constraints, emphasizing nursing staff allocation based on patient needs and introducing a routing policy to minimize trained nurses and nursing assistants while ensuring efficient patient care. The utilization of network transition diagrams, hidden Markov models, and heuristic algorithms throughout these sections has provided a robust framework for understanding and improving system performance.

Future research into heuristic algorithm applications and model validation will be crucial in enhancing the effectiveness and practicality of these methodologies. By continuing to refine and validate these models, we can strive towards more efficient and effective healthcare delivery systems that benefit both patients and healthcare providers. Additionally, future developments may consider continuous time modelling and discuss challenges arising from time constraints using HSMM and fluid queues.

REFERENCES

[1] Mariette Awad, and Rahul Khanna. "Hidden markov model." *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (2015): 81-104.

[2] Qaisar, Saad & Imtiaz, Sahar & Farooq, Fatima & Jamal, Amna & Iqbal, Wafa & Glazier, Paul & Lee, Sungyoung, "A hidden Markov model for detection and classification of arm action in cricket using wearable sensors," *Journal of Mobile Multimedia*, vol. 9, no. 2, pp. 128-144, 2013.

[3] Liu, Yang, "A Study of Hidden Markov Model," Master's Thesis, University of Tennessee, 2004.

[4] H. Mitchell, A. H. Marshall and M. Zenga, "Using the hidden Markov model to capture quality of care in Lombardy geriatric wards," *Operations Research for Health Care*, vol. 7, pp. 103-110, 2015.

[5] Yi, Kwan & Beheshti, Jamshid, "A hidden Markov model-based text classification of medical documents," *Journal of Information Science*, vol. 35, no. 1, pp. 67-81, 2009.

[6] Bartolomeo, N., Trerotoli, P., Moretti, A. *et al.* "A markov model to evaluate hospital readmission," *BMC Med Res Methodol* **8**, 23 (2008). https://doi.org/10.1186/1471-2288-8-23

[7] Wu, Jingna, et al., "Optimization of Markov Queuing Model in Hospital Bed Resource Allocation," *Journal of Healthcare Engineering*, vol. 2020, pp. 6630885, 2020.

[8] Sato, Renato Cesar, and Désirée Moraes Zouain, "Markov Models in Health Care," *Einstein (Sao Paulo, Brazil)*, vol. 8, no. 3, pp. 376-379, 2010.

[9] Cahill, W., and M. Render, "Dynamic Simulation Modeling of ICU Bed Availability," *WSC'99. 1999 Winter Simulation Conference Proceedings. 'Simulation - A Bridge to the Future' (Cat. No.99CH37038)*, IEEE, vol. 2, pp. 1573-1576, 2003.

[10] Ortiga B, Salazar A, Jovell A, Escarrabill J, Marca G, Corbella X, "Standardizing admission and discharge processes to improve patient flow: a cross sectional study," *BMC Health Services Research*, vol. 12, pp. 180, 2012.

[11] Kim, Seung-Chul, et al., "Analysis of Capacity Management of the Intensive Care Unit in a Hospital," *European Journal of Operational Research*, vol. 115, no. 1, pp. 36-46, 1999.

[12] Kao, Edward P. C., "Modeling the Movement of Coronary Patients within a Hospital by Semi-Markov Processes," *Operations Research*, vol. 22, no. 4, pp. 683-699, 1974. https://doi.org/10.1287/opre.22.4.683

[13] Green, Linda, "Queueing Analysis in Healthcare," *International Series in Operations Research & Management Science*, vol. 100, pp. 281-307, 2006.
[14] Aiken, Linda H., et al., "Hospital Nurse Staffing and Patient Mortality, Nurse Burnout, and Job Dissatisfaction," *JAMA: The Journal of the American Medical Association*, vol. 288, no. 16, pp. 1987-1993, 2002.
[15] Bolch, Gunter, Stefan Greiner, Hermann De Meer, and Kishor S. Trivedi. "Queueing networks and Markov chains: modeling and performance evaluation with computer science applications," John Wiley & Sons, 2006.
[16] Kishore, S., "Probability and Statistics with Reliability, Queuing and Computer Science Applications," John Wiley and Sons Ltd., 2001.
[17] Yankovic, Natalia, and Linda V. Green. "Identifying good nursing levels: A queuing approach." *Operations research* 59, no. 4 (2011): 942-955.
[18] Aarthi, S., and M. Shanmugasundari, "Comparison of Single Server Queuing Performance Using Fuzzy Queuing Model and Intuitionistic Fuzzy Queuing Model with Finite Capacity, "*IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp. 231-241, 2023.
[19] Du, Puliang, Zhong Chen, Ziqi Zhang, and Qi Zhao, "A congestion mitigation model of transmission network considering the participation of distribution companies in power market," *IAENG International Journal of Applied Mathematics*, vol. 51, no. 4, pp. 1017-1026, 2021.
[20] Hou, Yan-E., Wenwen He, Congran Wang, and Xinhui Ren, "A Two-stage Selection Hyper-heuristic Algorithm for the Capacitated Vehicle Routing Problem," *IAENG International Journal of Applied Mathematics*, vol. 52, no. 4, pp. 979-988, 2022.
[21] Lv, Shengli, Yueying Wu, Jingbo Li, and Ruiyu Wang, "The M/M/1 Repairable Queueing System with Variable Input Rates and Failure Rates," *IAENG International Journal of Applied Mathematics*, vol. 52, no. 4, pp. 784-790, 2022.
[22] Klimenok, Valentina, Alexander Dudin, Olga Dudina, and Irina Kochetkova, "Queuing system with two types of customers and dynamic change of a priority," *Mathematics*, vol. 8, no. 5, pp. 824, 2020.
[23] Tarabia, A. M. K., "Two-class priority queueing system with restricted number of priority customers," *AEU-International Journal of Electronics and Communications*, vol. 61, no. 8, pp. 534-539, 2007.
[24] Green, Linda V., Peter J. Kolesar, and Ward Whitt, "Coping with time-varying demand when setting staffing requirements for a service system," *Production and Operations Management*, vol. 16, no. 1, pp. 13-39, 2007.
[25] Wallace, Rodney B., and Ward Whitt, "A staffing algorithm for call centers with skill-based routing," *Manufacturing & Service Operations Management*, vol. 7, no. 4, pp. 276-294, 2005.
[26] Ta, Thuy Anh, Wyean Chan, Fabian Bastin, and Pierre L'Ecuyer, "A simulation-based decomposition approach for two-stage staffing optimization in call centers under arrival rate uncertainty," *European Journal of Operational Research*, vol. 293, no. 3, pp. 966-979, 2021.
[27] Green, Linda, Peter J. Kolesar, and Ward Whitt, "A staffing algorithm for call centers with skill-based routing," *Manufacturing & Service Operations Management*, vol. 7, no. 4, pp. 276-294, 2005.
[28] Broyles, J. R., Cochran, J. K., & Montgomery, D. C. (2011). "A Markov decision process to dynamically match hospital inpatient staffing to demand," *IIE Transactions on Healthcare Systems Engineering*, Vol. 1, no. 2, pp. 116–130, 2011. https://doi.org/10.1080/19488300.2011.609523