Simulation of Patient Flow in Emergency Department

Saasha Joshi
(V00976693)

Abstract

This graduate report discusses the methodology and implementation of a simulation program to model the flow of patients in an emergency department (ED) over a period of one month. The aim of this project is to perform a simulation to recognize various bottlenecks in the ED and optimize the use of resources, such as doctors, nurses, administrative staff, beds, etc., while reducing patient wait times, and improving patients' length of stay in the ED. A lot of focus in this project was given to increasing the technical complexity of the hospital/ED simulation by introducing the concept of the Canadian Triage and Acuity Scale (CTAS) and increasing the number of co-existing departments (namely Triage, ED, and inpatient departments) in a hospital, diagnostic tests (including, MRI, CT, blood test, and X-rays), and waiting rooms.

The report is accompanied by a 1000 LOC program written in Python that uses the Simpy library to create a simulation environment. The GitHub repository for easy viewing of the code is at this link: https://github.com/SaashaJoshi/patient-ER-flow-simulation, however, the repository is private at the moment and can be requested to be made public.

I. Project Description

Emergency Department (ED) management requires careful coordination and allocation of resources. On a daily basis, EDs all over the country continue to experience an increasing amount of patient volume. This increase is expected to continue and therefore, managing the resources and processes effectively becomes an important task. Any inefficiencies in this process can result in extended waiting periods, overcrowding, and delayed care, ultimately leading to unfavorable outcomes.

A. Project Goals

The primary objective of this project is to develop a simulation model of patient flow in the emergency room to optimize the use of resources, reduce patients' length of stay in the ED, and improve patient outcomes. In order to achieve this, a number of smaller goals are undertaken. These include,

- Developing a Discrete Event Simulation (DES) model of higher complexity to model the patient flow in a realistic hospital scenario.
- Using the developed simulation model to identify bottlenecks and inefficiencies in the patient flow process.
- Testing different scenarios (by selecting different parameters) to evaluate the impact of patient flow and suggest potential improvements.
- Process patients according to the Canadian Triage and Acuity Scale (CTAS) to model real-life statistics.
- Provide recommendations for improving patient flow and reducing patients' length of stay in the ED.

II. Simulation Model

EDs are complex systems that require an in-depth understanding before they can be modeled. Taking reference from a previously published paper by De Freitas et al. [1] in 2020, a highly complex ED model was developed for this project. This model replicates a real-life ED scenario where patients are processed according to the Canadian Triage and Acuity Scale (CTAS). This scale was developed by Canadian hospital Emergency Departments (CAD) to classify patients on different levels according to the acuity of their complaints and the severity of their signs and symptoms. Table I shows the different CTAS levels that are modeled in this project.

Level	Label
1	Resuscitation
2	Emergent
3	Urgent
4	Less Urgent
5	Non Urgent

TABLE I
CANADIAN TRIAGE AND ACUITY SCALE (CTAS)

A. Simulation Parameters

In order to model the flow of patients in an ED, various parameters such as patients, doctors, nurses, administrative staff, lab equipment, and machinery are required. Since these parameters provide services in the simulation they are treated as servers.

B. Input parameters and Decision-Making Process

Patients, or the entities that flow through the simulation, are modeled as a resource that enters and exits the simulation according to a series of specific functions. Since patients are modeled in a DES, each event results in some significant change in the state of the patient.

The decision-making in this simulation is done with the help of functions and other process resources. That is, here, patients do not have the skills for the decision-making process. If any decision is to be made, the decisions are generated randomly using some function. Hence, this project does not implement an agent-based simulation.

C. Process parameters (Human Resources)

Process parameters or resources such as doctors, nurses, administrative staff, and consultants are assumed to be variable over different simulation runs. These resources are also assumed to be available all the time, i.e., a time limit, such as a shift period, is not imposed. For doctors and nurses, each of these resources can serve any department in the hospital. They are assumed to be equal in terms of their respective skill levels.

All these human resources provide one or the other kind of service, and hence come with their respective service time distributions. Every service time is sampled from a general independent distribution. For this project implementation, most of the service times are modeled using Triangular distribution.

A Triangular distribution is most commonly used to model situations where the sample data is limited. Since this data is difficult to collect or generalize over a large number of institutions, Triangular distribution seems like the best fit for this type of simulation [3].

D. Process parameters (Non-Human)

Various lab equipment such as blood tubes, medications, and machinery such as ECG, CT, and X-Rays are considered to be of fixed capacity. The only difference is that lab equipment can be replenished easily, since in real life they are minor tools and objects that can be reused, replaced, or reordered, whereas, machinery remains fixed unless specified.

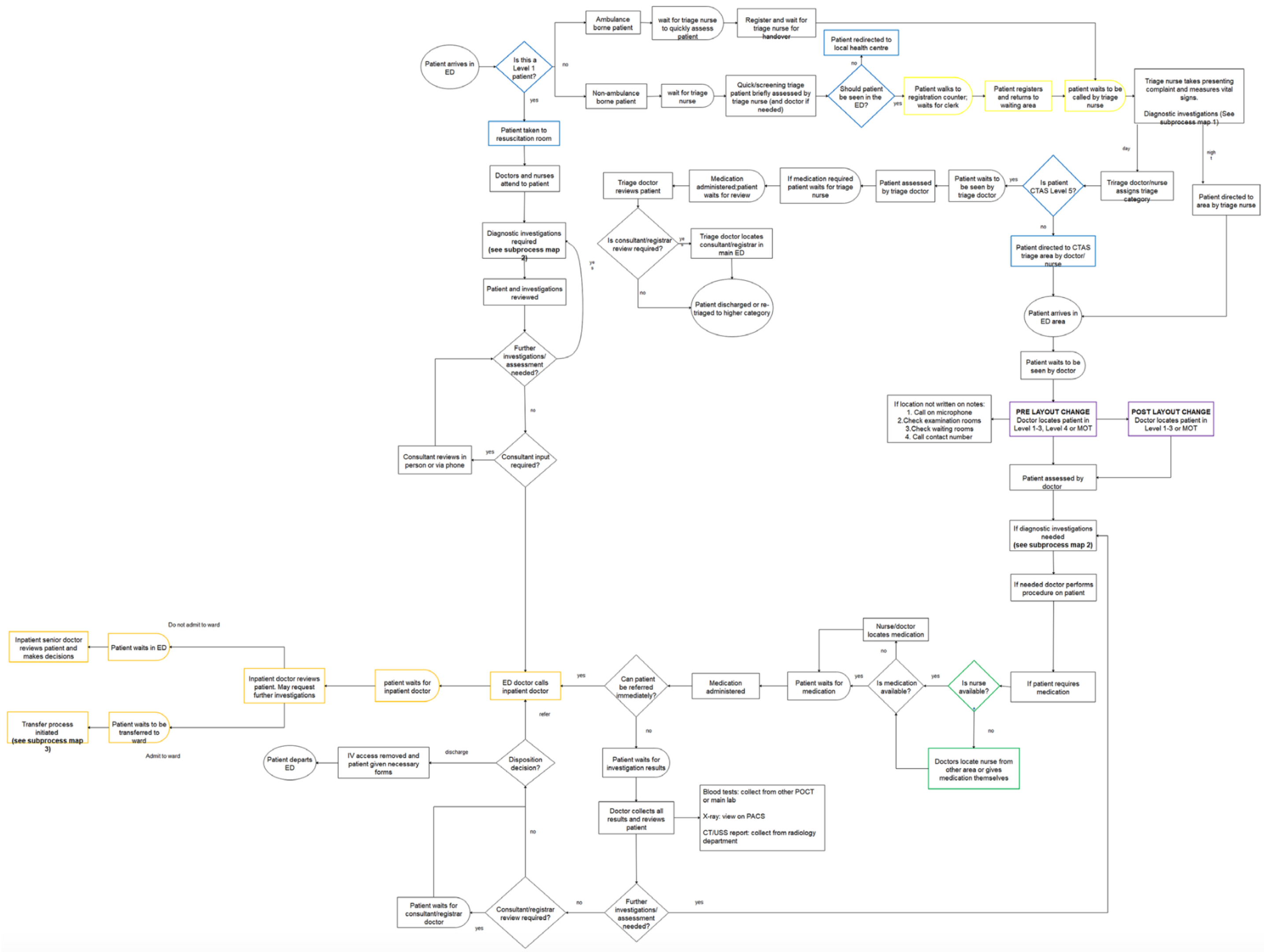

Fig. 1. Main Process

E. Static Model

This simulation model is implemented as a static model to avoid any complexity. In a static system, none of the processes affect or influence each other. For example, in case of a backlog in the ECG waiting room, the decision to increase the number of machines is not given to any resource or function in the simulation. Rather, this decision is made while analyzing the simulation systems in the Analysis section of this report.

F. Output Parameters

The parameters tested using this simulation are the patient's length of stay (LOS) data. The primary aim of this simulation is to model a system that reduces the average LOS.

III. Methodology and Implementation

The ED model implemented in this project, as referred from [1], consists of a larger main process map, as shown in Figure 1, accompanied by three smaller sub-processes depicted in Figures 2, 3, and 4.

A. *Main Process Map*

The main process map of the ED model depicts the flow of patients in an ED from start to stop. The map can be divided into different subsections for easy understanding.

Section 1: Enter, Triage Stage, Registration and Subprocess 1

This subsection of the main process map models the patient flow as soon they enter the ED until they are triaged to be sent to an ED doctor. The first step in this subsection is to perform an on-arrival CTAS to differentiate between CTAS Level 1 patient with the rest. While the CTAS-I patients go to a resuscitation room (Section 5), the rest of the patients are sent to the Triage department for a detailed evaluation and triage.

In this Triage department, the patient waits in the triage waiting room until a nurse is assigned. A quick screening at the beginning of the triage process separates CTAS-V patients that may be referred to a local health clinic from the rest.

At the end of the triage stage, a nurse may assign the need for diagnostic evaluations. These evaluations are a list of certain tests, including ECG, urine test, or X-Ray, that are modeled according to Subprocess 1 shown in Figure 2.

The queuing model for Section 1 is an M/GI/c model with exponential arrival rates of patients and Triangular service times for patient assessment by nurses.

It is also important to note that the implementation of this model in this project doesn't distinguish between the night and day processes as depicted in the figure. In addition to this, a waiting room and a registration process are created in this the project implementation of this section.

Section 2: Triage Doctor, Consultation and Re-Triage

Once the CTAS level of the patient is triaged, the patient is sent either to a Triage doctor or to the ED doctor. These doctors may be distinguished on the basis of their skill levels. If the patient is triaged as CTAS-V, without the need to be referred to a local health clinic, they are sent to the triage doctor for further assessment and medication. A consultation with in-hospital consultants may also be scheduled.

Once the procedure is complete, the patient is discharged. In a few cases, a patient may also be re-triaged to a higher level. In such a case, the patient is made to enter Section 3.

In the project implementation of this section, a waiting room is created until a doctor is assigned to the patient. The queuing model for this section is an M/GI/c model with exponential arrival rates of patients and Triangular service times for patient assessment by nurses. The consultation process also follows the M/GI/c model.

Section 3: ED Doctor, Subprocess 2 and Medication

If after Section 1 the patient is triaged at a level higher than CTAS-V, the patient is directly sent to an ED doctor. This ED doctor accompanies the patient through every process that will be performed in the main ED.

In this section after preliminary assessment, the doctor may assign some diagnostics investigations. These diagnostic tests are different from the ones performed in Section 1. The map depicting the process flow of these tests is given in Figure 3 as Subprocess 2.

A particularly iterative process created in this project is the medication process in this section. This process makes sure that either a doctor or nurse gives the required medication to the patient. However, in case of the unavailability of a nurse, the doctor initially assigned to the patient is responsible to collect and administer the medication. The medication, in the implementation, has a fixed capacity. If that capacity is met before the doctor or nurse can administer it, the medication is replenished. This replenishment takes some time which is again modeled by a triangular distribution.

The queuing model for this section also follows an M/GI/c model with Triangular service times for patient assessment by doctors.

Section 4: Further Investigations and Consultation

Once the medication has been administered in the main ED, a further assessment may be performed by the doctor on the patient. This assessment may include retesting or external consultation.

Once the diagnostic test results are collected from Section 3, if the doctor flags the patient for retesting, the patient is redirected to Subprocess 2. This redirection, particularly, introduces a huge delay in the exit of the patient from the ED.

After the retesting, a consultation with an in-hospital consultant may be requested by the doctor. This consultation process is similar to the one that can be performed for CTAS-V patients in Section 2.

Section 5: CTAS-I Resuscitation Room, Further Investigations and Consultation

Continuing from Section 1, once the CTAS-I patients are recognized, they are transferred immediately to the resuscitation room, skipping the main ED entirely. This separate flow is created in order to treat CTAS-I patients as soon as possible. Once the patients enter this section, doctors and nurses assess the patient and decide upon different diagnostic investigations and consultations.

The diagnostic investigations presented in Subprocess 2 are completed here. The consultation process remains the same as in Sections 2 and 4.

The queuing model for this section also follows an M/GI/c model with Triangular service times for patient assessment by doctors or nurses, however, the service times are a little better than those in other sections. The resuscitation time is also modeled as Triangular distribution.

Section 6: Inpatient process, Transfer to the ward, Subprocess 3, and Exit

The final section of the main process deals with patients of all CTAS levels. Patients from Sections 4 and 5 are directed to an inpatient doctor if there seems a need for assessment by a senior doctor or admission to a ward. Therefore, this section has two exit strategies. If the patient can be assessed within a day without the need to stay, a senior doctor is requested. The patient is discharged once the assessment is complete.

If, however, there arises a need for the patient to stay, a transfer is initiated. This transfer process is described in the Subprocess 3 map.

The queuing model for this section also follows an M/GI/c model with Triangular service times for patient assessment by doctors or nurses. An inpatient waiting room is also created in this section.

B. Subprocess 1: Diagnostic evaluations at triage

In Subprocess 1, Figure 2 diagnostic evaluation at the triage level are performed. These investigations include 3 main tests: ECG, Urine, and X-Ray, each following a different subprocess of its own.

For the diagnostic evaluations and tests in this process, each test is modeled with Triangular service times. The machines required for these tests, however, have fixed availability but are changed strategically to present different analytical statistics.

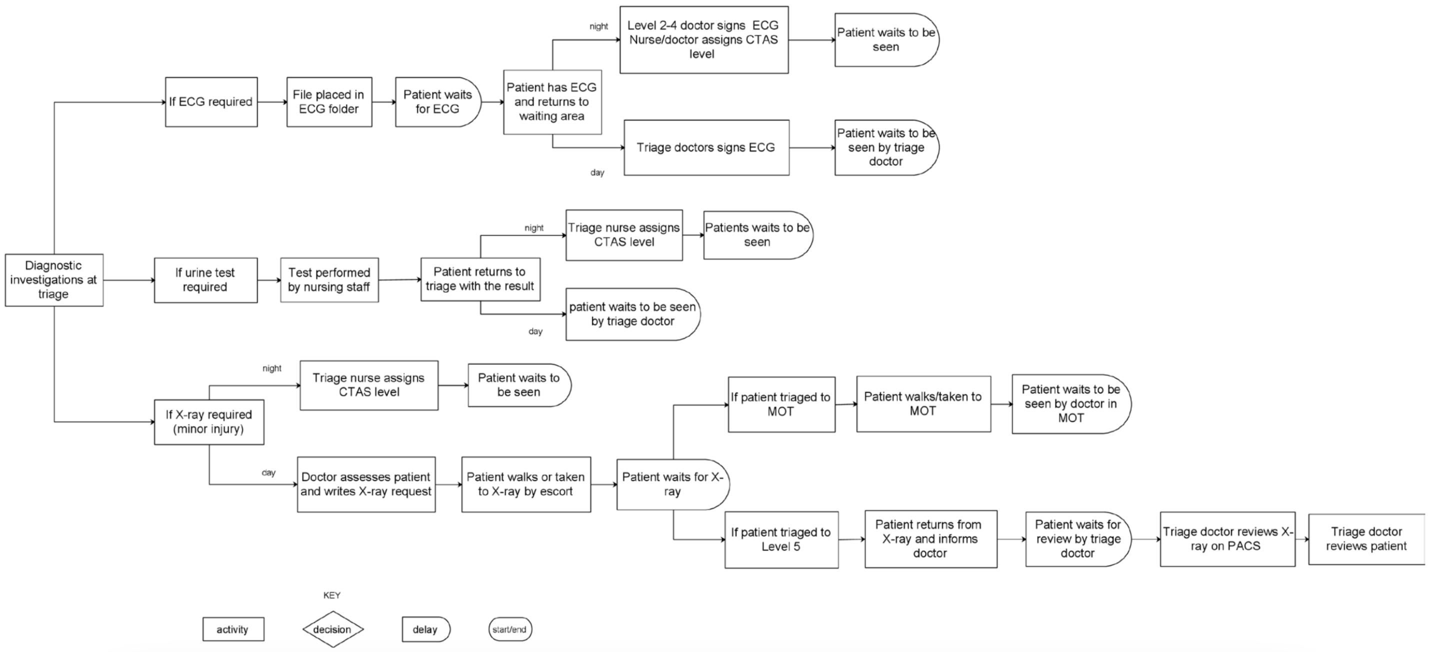

Fig. 2. Subprocess 1: Diagnostic Investigations at Triage

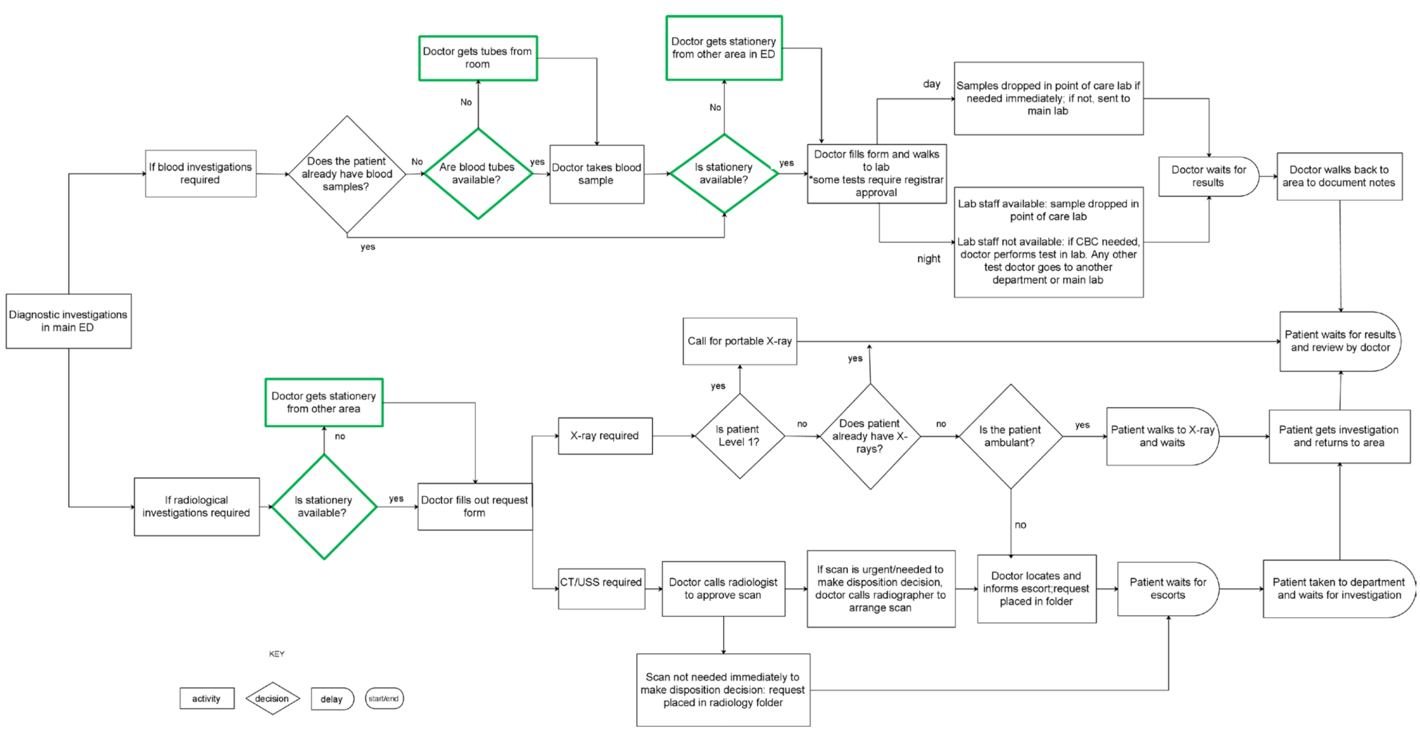

Fig. 3. Subprocess 2: Diagnostic Investigations at Main ED

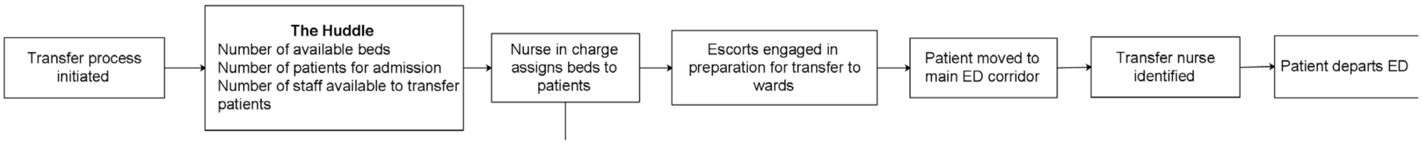

Fig. 4. Subprocess 3: Transfer to Ward

C. Subprocess 2: Diagnostic evaluations at main ED

The diagnostic evaluations at the main ED, however, are different from that at the Triage department. These investigations include 2 prominent tests: Blood and Radiological investigations.

For the blood investigation/test, a patient is sent to a laboratory for blood collection. A separate fixed-capacity resource is created for blood tubes. If the blood tubes are exhausted, the replenishment process is undertaken immediately. The implementation of this process in the project doesn't assume any wait times for replenishment, however, the patient does need to wait for a doctor or admin staff to be assigned to them.

The radiological investigations on the other hand have two options: an X-Ray or a CT Scan. Any of the scans in this process need to be first approved by a radiologist. In the implementation, however, admin staff is used to model radiologists. This is done only to maintain simplicity in the subprocess. The process can be made complex by adding a skill level to the doctors that can act as radiologists. Once the scans are approved, the machines are allocated from a fixed capacity resource, and the service is performed according to a Triangular distribution.

It is important to note that a different process for day and night is not implemented for this project.

D. Subprocess 3: Transfer process

The transfer process is initiated at the last process in the simulation model where a patient if required to be observed over a long period of time, is transferred or admitted into a ward. All the procedure dealing with the assignment of beds and nurses is done in this subprocess.

This project implementation does not model any further than this subprocess.

IV. Setup and Statistics

A lot of resources were curated to complete the implementation of the above-described simulation model. Some of these resources were kept constant over the course of the simulation to reduce complexity, rest were constantly changed for the purpose of analysis. The setup and resource statistics used in this project implementation can be divided into various classes,

A. Hospital Resources

Hospital resources may include staff that works at different departments or machinery and laboratory equipment. These items are termed resources since Simpy provides a class named *Resource* to define items in a simulation model with the same basic concepts. These resources can be *requested* or *released* as per requirement. In order to request a resource, Simpy utilizes the keyword *yield*. Yield puts in a request for the resource and assigns it a task once it becomes available.

- Staff: This includes doctors, nurses, administrative staff, and consultants.
  Over the course of the implementation, the staff is *requested* as per requirement and *released* when the process ends. The request method *yields* and assigns a task once the requested resource is available.

- Machinery: This includes all the testing machines such as ECG, CT, or X-Ray machines.

Machines are implemented as fixed-capacity resources. That is, their capacity is fixed to either 5 or 10 over the course of the simulation and cannot be changed by the user (the developer, however, can modify the capacity).

- Laboratory equipment and Medication: This includes all the blood tubes and medication supply. These resources are defined as *Containers* with an initial fixed capacity, changeable over time. A simulation environment accesses these containers with the help of *get()* method. If the resources are exhausted, they can be easily replenished with the help of the *put()* method.

The changes made to the capacities of every hospital resource are discussed in detail in the Analysis section of the project report.

B. Arrival and Inter-Arrival Times

Hospital data is very difficult to obtain or generalize over a large number of patients. In this project implementation, we limit ourselves to utilizing historic arrival and inter-arrival rates that are either constant for all the patients or vary for patients from different CTAS levels [2].

The Canadian NACRS Emergency Department data [2] provides us with provincial data on the Total number of ED visits per CTAS category (CTAS I-III or CTAS IV-V) over a period of 6 months. From this data, a constant arrival rate can be calculated as,

$$\textbf{Arrival rate (mins)} = \frac{\textit{Total ED visits in BC in 6 months}}{\textit{Time Period}} \tag{1}$$

Further, in order to increase the complexity of the simulation model, different arrival rates for CTAS I-III and CTAS IV-V category patients were also considered from the same data. In this implementation, a prediction of the patients' CTAS level was made in order to generate an inter-arrival time.

The distribution of the arrival rate, as referred from historic data and research papers [3], is assumed to be Poisson distribution. Hence, the inter-arrival times are Exponentially distributed over the course of the implementation.

For half of the simulation systems analyzed (2 out of 4), the arrival times were kept constant. For the other half, a combination of inter-arrival times for CTAS I-III and CTAS IV-V patients was considered.

C. Service Times

Expected duration/service times of various processes implemented in the project were taken from a previously published paper by Jones [3]. The service times utilized in this paper are Triangular distributions that specify the minimum, maximum, and mode of the duration of a particular process or service. A few modifications to these values were made for this project implementation. Table II shows all the distributions and values used to model various processes in the ED.

V. Analysis

The simulation program is run for 4 systems with different system variables, for 5 runs each. These systems are run with Common Random Numbers (CRN) with respect to every run. Both the within-replication and across-replication results for each simulation system are obtained and listed.

Patient Care Process	Distribution	Parameters
Registration	Triangular	3.0, 4.0, 8.0
Assessment Time	Triangular	4.0, 6.0, 8.0
Triage	Triangular	4.0, 6.0, 8.0
X-Ray	Triangular	10.0, 18.0, 30.0
Urine Test	Triangular	5.0, 7.0, 12.0
ECG	Triangular	45.0, 55.0, 60.0
Blood Test	Triangular	5.0, 7.0, 12.0
CT	Triangular	45.0, 55.0, 60.0
Medication Administration	Triangular	1.0, 2.0, 3.0
Minor Procedure	Triangular	4.0, 6.0, 10.0
Consultation (CTAS I)	Triangular	10.0, 15.0, 30.0
Consultation (rest)	Triangular	5.0, 10.0, 30.0
Bed Release	Triangular	30.0, 50.0, 90.0
Review	Triangular	1.0, 2.0, 3.0
Replenishment of containers	Triangular	1.0, 2.0, 3.0
Transfer	Triangular	4.0, 7.0, 9.0

TABLE II
SERVICE TIME DISTRIBUTIONS

Parameters	System 1	System 2	System 3	System 4	System 5
Simulation Time Length	43800 min	43800 min	43800 min	43800 min	43800 min
Arrival rate	CTAS I-III: 1.6 CTAS IV-V: 1	CTAS I-III: 1.6 CTAS IV-V: 1	1.3	2.7	2.7
No. of Doctors	100	50	100	100	200
No. of Nurses	100	50	100	100	200
No. of Admin Staff	70	35	70	70	140
No. of Consultants	10	5	10	10	20
No. of Beds	10	10	10	10	10
No. of ECG Machines	5	5	5	5	10
No. of CT Machines	5	5	5	5	10
No. of X-Ray Machines	5	5	5	5	10
Performance Measure	LOS	LOS	LOS	LOS	LOS

TABLE III
PARAMETER VALUES FOR DIFFERENT SIMULATION SYSTEMS

A. *System 1*

Simulation System 1 has the given simulation parameters and statistics in Table III. Please note, the service times of various processes in the simulation remain as they were declared in the section Setup and Statistics.

In this system, the patients do not have any decision-making ability. The decision to undergo a particular diagnostic investigation is taken with the help of random number generator functions *(random.randint)*. In addition to this, a patient can be subject to multiple diagnostics investigations, i.e. a patient who underwent a blood test may also be prescribed a radiology scan.

This system is run for a total of 43800 minutes, i.e. one month, and the average number of patients processed completely, i.e. from arrival to departure, in every run is 48,357.

According to the statistics generated by this simulation system, as shown in Table IV and Figure 5, the performance of all the runs are similar and their confidence levels overlap. If the simulation is run for longer than one month time period, Run 1 or Run 3 may perform better than the rest.

This simulation run, however, tells us that with the assumed parameter values in Table III, the average LOS of patients is 1.35 hours with a confidence interval (C.I.) half-interval of 0.004 (taken from across replication data). The confidence interval of this LOS result can be given as,

Number of Run	Seed	Number of Patients	Within Replication Data
1	100	48000	$\overline{Y_1} = 1.35 \quad S_1^2 = 0.48 \quad H_1 = 0.006$
2	121	48000	$\overline{Y_2} = 1.34 \quad S_2^2 = 0.45 \quad H_2 = 0.006$
3	146	48000	$\overline{Y_3} = 1.35 \quad S_3^2 = 0.47 \quad H_3 = 0.006$
4	258	48000	$\overline{Y_4} = 1.35 \quad S_4^2 = 0.47 \quad H_4 = 0.006$
5	41	48000	$\overline{Y_5} = 1.34 \quad S_5^2 = 0.46 \quad H_5 = 0.006$
Across Replication Data			$\overline{Y} = 1.35 \quad S^2 = 0.00003 \quad H = 0.004$

TABLE IV
STATISTICS FOR SIMULATION SYSTEM 1

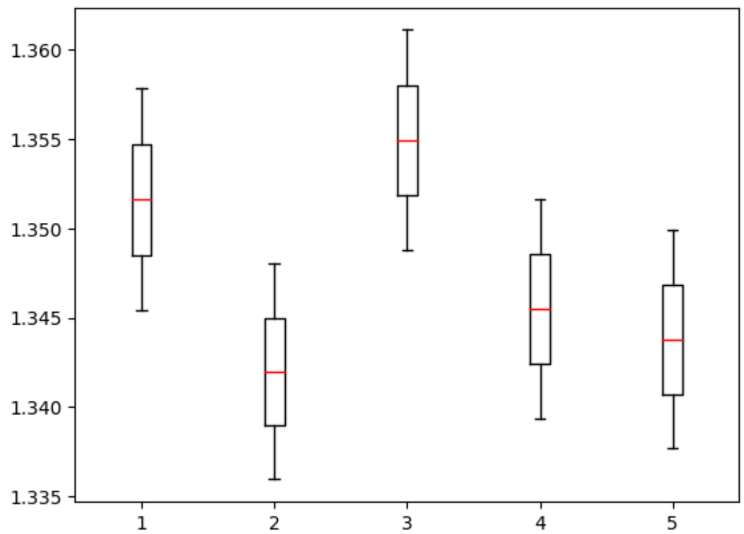

Fig. 5. Confidence Intervals for Simulation System 1 (x = Simulation run; y = LOS)

$$C.I.\ Upper\ bound = 1.35 \tag{2}$$
$$C.I.\ Lower\ bound = 1.34 \tag{3}$$
$$LOS = 1.35 \pm 0.004 \tag{4}$$

B. System 2

Simulation System 2 has the given simulation parameters and statistics in Table III.

This system is run for a total of 43800 minutes, i.e. one month, and the average number of patients processed completely, i.e. from arrival to departure, in every run is 833. A significant reduction in the number of resources is created for this simulation system.

A reduction in the number of patients processed can be explained by the backlog created by the reduction in the number of available human resources, including doctors, nurses, and admin staff. This clearly tells

Number of Run	Seed	Number of Patients processed	Within Replication Data
1	100	800	$\overline{Y_1} = 3.7 \quad S_1^2 = 6.3 \quad H_1 = 0.17$
2	121	800	$\overline{Y_2} = 5.7 \quad S_2^2 = 15.8 \quad H_2 = 0.24$
3	146	800	$\overline{Y_3} = 4.3 \quad S_3^2 = 9.54 \quad H_3 = 0.21$
4	258	800	$\overline{Y_4} = 4.5 \quad S_4^2 = 11.3 \quad H_4 = 0.21$
5	41	800	$\overline{Y_5} = 3.6 \quad S_5^2 = 5.88 \quad H_5 = 0.19$
Across Replication Data			$\overline{Y} = 4.40 \quad S^2 = 0.77 \quad H = 0.76$

TABLE V
STATISTICS FOR SIMULATION SYSTEM 2

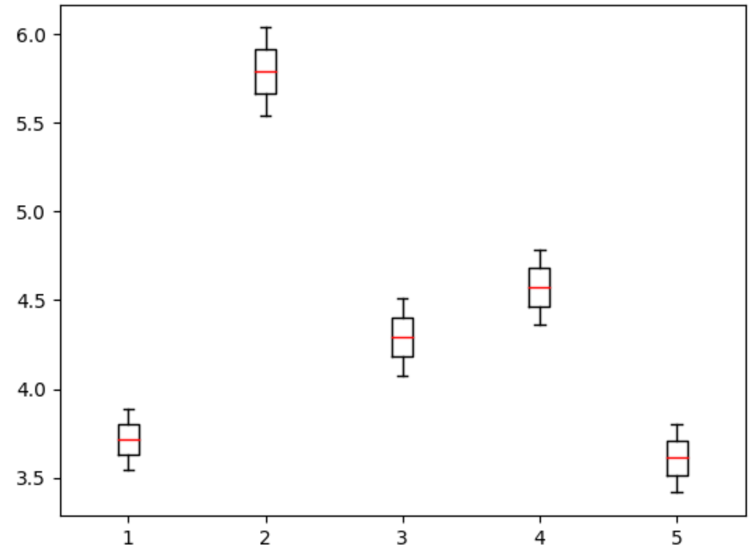

Fig. 6. Confidence Intervals for Simulation System 2 (x = Simulation run; y = LOS)

us that there is an acute shortage of staff since the current amount of doctors and nurses is not sufficient to handle the expected patient volume in one month. This shortage of staff not only leads to delayed care, but the patients already processed also saw an increase in their LOS periods.

According to the statistics generated by this simulation system, as shown in Table V and Figure 6, the average LOS of patients is 4.4 hours with a confidence interval (C.I.) half-interval of 0.76 (taken from across replication data). However, the performance of Run 2 stands out significantly with no overlap with other runs.

The confidence interval of this LOS result can be given as,

$$\textit{C.I. Upper bound} = 5.16 \tag{5}$$

$$\textit{C.I. Lower bound} = 3.63 \tag{6}$$

$$\textit{LOS} = 4.4 \pm 0.76 \tag{7}$$

C. System 3

Simulation System 3 has the given simulation parameters and statistics in Table III. This system is run for a total of 43800 minutes, i.e. one month, and the average number of patients processed completely, i.e. from arrival to departure, in every run is 47,259.

In this simulation system, all the parameters are similar to that of System 1 with the exception of the arrival rate. The arrival rate in System 3 is 1 customer per minute. Here, no distinction between different CTAS patients is made, i.e. every patient arrives with a constant arrival rate. This is unlike how the arrival rates were set in Systems 1 and 2.

No clear reduction in the number of patients can be seen. This result can indicate that distinguishing between different patients may not be the right characteristic at an ED-level simulation. A major explanation behind this result can also be that CTAS-I patients are immediately taken to Section 5 (Resuscitation) of the main map, hence creating a faster treatment pathway. For all the other patients, the treatment flow remains the same. A prominent difference in LOS may be seen once the decision to admit a patient is made. This will overall increase the LOS of patients who need to be observed for a longer period of time and hence should stay in the ED.

According to the statistics generated by this simulation system, as shown in Table VI and Figure 7, the average LOS of patients is 1.34 hours with a confidence interval (C.I.) half-interval of 0.004 (taken from

Number of Run	Seed	Number of Patients	Within Replication Data
1	100	47000	$\overline{Y_1} = 1.34 \quad S_1^2 = 0.45 \quad H_1 = 0.006$
2	121	47000	$\overline{Y_2} = 1.34 \quad S_2^2 = 0.45 \quad H_2 = 0.006$
3	146	47000	$\overline{Y_3} = 1.33 \quad S_3^2 = 0.45 \quad H_3 = 0.006$
4	258	47000	$\overline{Y_4} = 1.33 \quad S_4^2 = 0.44 \quad H_4 = 0.006$
5	41	47000	$\overline{Y_5} = 1.34 \quad S_5^2 = 0.46 \quad H_5 = 0.006$
Across Replication Data			$\overline{Y} = 1.34 \quad S^2 = 0.00001 \quad H = 0.004$

TABLE VI
STATISTICS FOR SIMULATION SYSTEM 3 (X = SIMULATION RUN; Y = LOS)

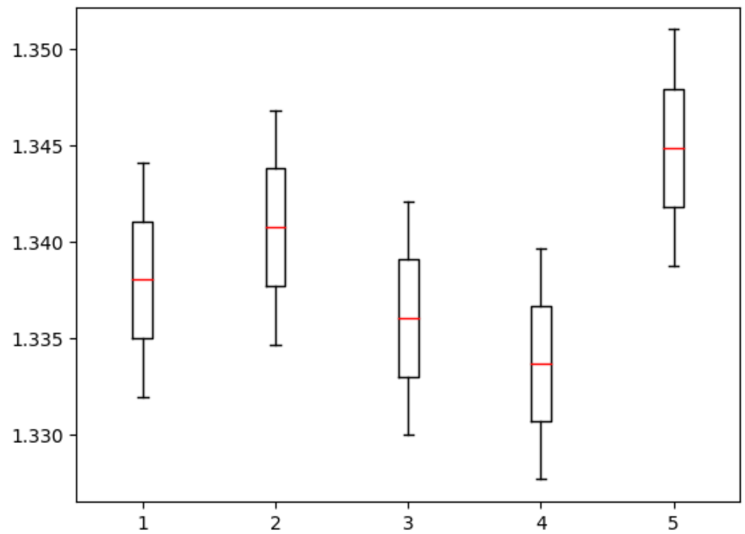

Fig. 7. Confidence Intervals for Simulation System 3

across replication data). The performance of all the runs in this system is similar and their confidence levels overlap. If the simulation is run for longer than one month time period, Run 5 may perform better than the rest.

The confidence interval of this LOS result can be given as,

$$\textit{C.I. Upper bound} = 1.34 \tag{8}$$

$$\textit{C.I. Lower bound} = 1.33 \tag{9}$$

$$\textit{LOS} = 1.34 \pm 0.004 \tag{10}$$

D. System 4

Simulation System 4 has the given simulation parameters and statistics in Table III. This system is run for a total of 43800 minutes, i.e. one month, and the average number of patients processed completely, i.e. from arrival to departure, in every run is 1693.

In this simulation system, all the parameters are similar to that of System 1 with the exception of the arrival rate. The arrival rate in System 4 is 2.7 customers per minute. Here, no distinction between different CTAS patients is made, i.e. every patient arrives with a constant arrival rate. This is unlike how the arrival rates were set in Systems 1 and 2. This system is similar to System 3 with an exception of an increased arrival rate.

A significant reduction in the number of patients can be seen and can be explained by the backlog created by the increased arrival rate for a constant number of available human resources, including doctors, nurses, and admin staff. This clearly tells us that with an increased arrival rate, an increase in the number

Number of Run	Seed	Number of Patients	Within Replication Data
1	100	1600	$\overline{Y_1} = 4.25 \quad S_1^2 = 11.49 \quad H_1 = 0.17$
2	121	1600	$\overline{Y_2} = 4.22 \quad S_2^2 = 9.40 \quad H_2 = 0.14$
3	146	1600	$\overline{Y_3} = 3.69 \quad S_3^2 = 7.69 \quad H_3 = 0.14$
4	258	1600	$\overline{Y_4} = 4.70 \quad S_4^2 = 12.84 \quad H_4 = 0.15$
5	41	1600	$\overline{Y_5} = 4.19 \quad S_5^2 = 10.10 \quad H_5 = 0.15$
Across Replication Data			$\overline{Y} = 4.21 \quad S^2 = 0.12 \quad H = 0.31$

TABLE VII
STATISTICS FOR SIMULATION SYSTEM 4

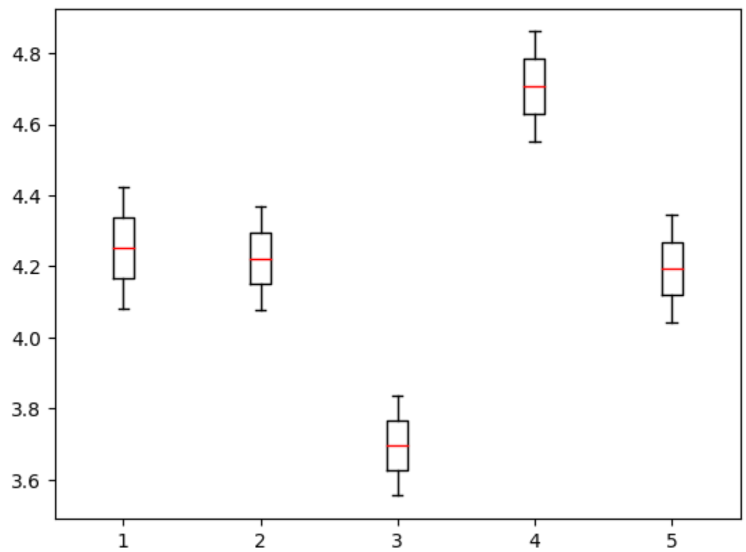

Fig. 8. Confidence Intervals for Simulation System 4 (x = Simulation run; y = LOS)

of resources is necessary since the current amount of doctors and nurses is not sufficient to handle the expected patient volume in one month. This shortage of staff with an increased arrival rate not only leads to delayed care, but the patients already processed also saw an increase in their LOS periods.

According to the statistics generated by this simulation system, as shown in Table VII and Figure 8, the average LOS of patients is 4.21 hours with a confidence interval (C.I.) half-interval of 0.31 (taken from across replication data). However, the performance of Run 4 stands out significantly with no overlap with other runs.

The confidence interval of this LOS result can be given as,

$$\textit{C.I. Upper bound} = 4.53 \tag{11}$$

$$\textit{C.I. Lower bound} = 3.90 \tag{12}$$

$$\textit{LOS} = 4.21 \pm 0.31 \tag{13}$$

E. System 5

Simulation System 4 has the given simulation parameters and statistics in Table III. This system is run for a total of 43800 minutes, i.e. one month, and the average number of patients processed completely, i.e. from arrival to departure, in every run is 98300.

In this simulation system, all the parameters are similar to that of System 1 with the exception of the arrival rate. The arrival rate in System 4 is 2.7 customers per minute. This system is similar to System 4 with a significant increase in the number of resources to overcome the backlog created in the previous

Number of Run	Seed	Number of Patients	Within Replication Data
1	100	90000	$\overline{Y_1} = 1.33 \quad S_1^2 = 0.44 \quad H_1 = 0.004$
2	121	90000	$\overline{Y_2} = 1.33 \quad S_2^2 = 0.43 \quad H_2 = 0.004$
3	146	90000	$\overline{Y_3} = 1.32 \quad S_3^2 = 0.43 \quad H_3 = 0.004$
4	258	90000	$\overline{Y_4} = 1.33 \quad S_4^2 = 0.43 \quad H_4 = 0.004$
5	41	90000	$\overline{Y_5} = 1.33 \quad S_5^2 = 0.44 \quad H_5 = 0.004$
Across Replication Data			$\overline{Y} = 1.33 \quad S^2 = 0.000007 \quad H = 0.002$

TABLE VIII
STATISTICS FOR SIMULATION SYSTEM 5

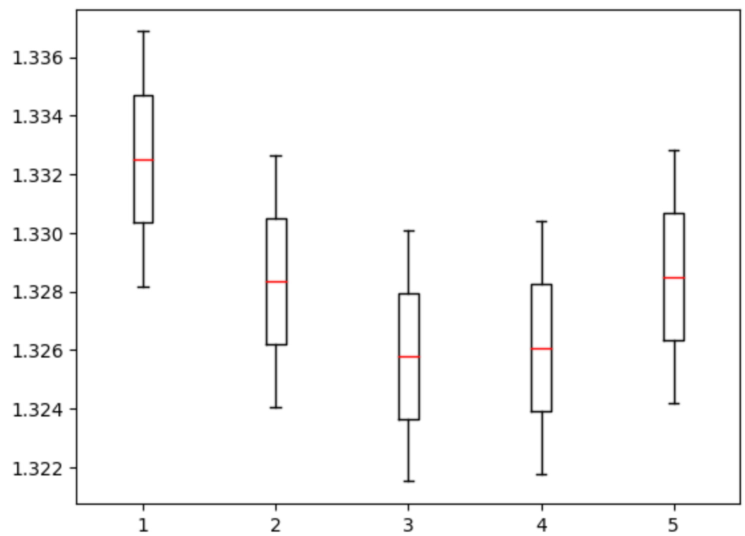

Fig. 9. Confidence Intervals for Simulation System 5 (x = Simulation run; y = LOS)

simulation system. This increase is also made for an initially fixed number of lab equipment and machinery such as ECG, CT, and X-Ray machines.

With an increase in the number of resources such as doctors, nurses, admin staff, and other lab equipment, a significant improvement in the number of processed patients can be seen (when compared with System 4). This clearly tells us that with an increased arrival rate of patients, a demand to increase the available human resources, including doctors, nurses, and admin staff, or even other lab equipment and machines maybe be inevitable. Therefore, an increase in the number of resources is necessary to overcome the bottlenecks seen in System 4.

According to the statistics generated by this simulation system, as shown in Table VII and Figure 8, the average LOS of patients is 1.33 hours with a confidence interval (C.I.) half-interval of 0.002 (taken from across replication data). The performance of all the runs in this system is similar and their confidence levels overlap. If the simulation is run for longer than one month time period, Run 1 may perform better than the rest.

The confidence interval of this LOS result can be given as,

$$C.I.\ Upper\ bound = 1.33 \tag{14}$$

$$C.I.\ Lower\ bound = 1.32 \tag{15}$$

$$LOS = 1.33 \pm 0.002 \tag{16}$$

F. Length of Stay

The LOS data obtained from the above simulation systems when plotted as a histogram display a multi-modal distribution. The reason behind this multi-modal observance of LOS can indicate multiple things,

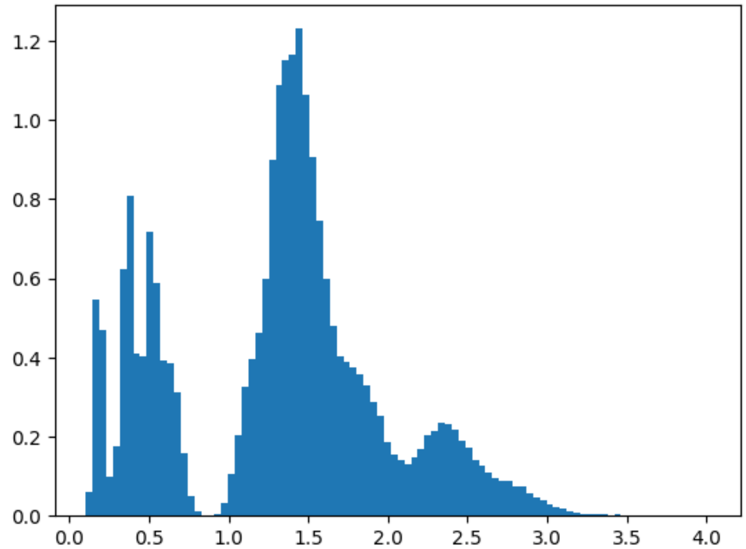

Fig. 10. Histogram of LOS of patients (x = LOS (hours); y = Frequency) (Sampled from System 4 Run 1)

1) Heterogeneity of the system: A simulation system with multiple classes of input parameters (CTAS I-V patients) or multiple subsystems (refer to section Methodology) may introduce such heterogeneity in the system performance measure, here, LOS.

A very interesting observation that can be made here is that few patients from CTAS-I are treated within minutes, since due to the random nature of the simulation they may not be assigned any diagnostic investigations or consultations. Similarly, a few of the CTAS-V patients are referred to local health clinics early in Section 1 of triage. This supposedly fast treatment of patients can be seen on the graph in Figure 10 where the data at the very beginning of the distribution indicates lower LOS.

2) Non-linearity of the system: A non-linear system is a system in which the LOS factor majorly depends on the availability of the processing parameters such as doctors, nurses, etc. Real-world patient flow in an ED is also representative of a non-linear system. Even with the presence of a triage system such as CTAS, non-linearity may be seen in the system. For example, if two patients at higher but similar CTAS enter the ED at the same time, one of these patients may have to wait longer than the other.

VI. Conclusion

The project report describes the methodology behind the implementation of a simulation program to model the flow of patients in an ED. The simulation is built in Python programming language with the help of the Simpy simulation library. The process map is divided into various departments at an ED department accompanied by three subprocesses, including diagnostic investigations and consultations.

The project also reports the statistics obtained from five different simulation systems, each run for five rounds using a CRN respectively. The statistics generated from these five systems tell us the mean LOS values (in hours) for patients in the modeled ED. Along with this, the statistics provide us with confidence intervals for both within-replication and across-replication data. The reported statistic values (across-replication) for mean LOS for all the five systems are: $1.35 \pm 0.004, 4.4 \pm 0.76, 1.34 \pm 0.004, 4.21 \pm 0.31,$ and 1.33 ± 0.002. These values show that while the arrival rate of patients (irrespective of being constant or varied according to CTAS level) is around 1.6, a team of 100 doctors, 100 nurses, 70 admin staff, and 10 consultants with 5 of each ECG, CT, or X-Ray machines can efficiently manage the patient flow without any serious bottlenecks (shown in Systems 1 and 3). However, as soon as the arrival rate of the system increases, with the same amount of available resources, a series of bottlenecks arise (shown in

System 4). This leads to not enough patients being able to complete their procedure or having extremely large LOS, about 4.4 hours in the ED. In addition to this, if the number of resources is somehow reduced or removed from the simulation, the system again starts to show serious problems in terms of efficiency and patient flow management (shown in System 3). These problems are however removed by a judicious increase in the number of resources with the increase in the flow/arrival rate of patients in the ED. With more available resources and machines, the significant problem of patient flow bottleneck is resolved and more patients are processed with lesser LOS values.

For every simulation system, five runs with different seed values were performed. For Systems 1, 3, and 5 the confidence intervals of the mean LOS values overlap significantly. Hence a proper conclusion cannot be drawn unless these simulations are run for a much longer period than one month. However, for Systems 2 and 4, at least one among the five runs (Run 2 for System 2 and Run 4 for System 4) stands out from the rest of the runs, without any overlap in the confidence levels. It can thus be said that these runs are better representations of the simulation model built for this project for their respective parameter values.

The LOS values obtained from these simulation systems seem to follow a multi-modal distribution with 3 peaks. This behavior of the performance measure can be explained by listing the presence of heterogenous subsystems in the simulation model, namely Triage, main ED, and inpatient departments or CTAS-I-V patients, that follow different paths while in the simulation system.

The purpose of performing this simulation is to recognize the cause-and-effect of various resources, their availability, and the complex interaction of different processes in a real-life scenario. This simulation was built to recognize potential areas of problems and their solutions. From the statistics obtained from these five simulation systems, it can be concluded that an increased need for more resources, especially doctors and nurses, is an inevitable requirement of the time. All of this is needed to significantly reduce the patient's LOS in an ED, and their waiting times while making sure that EDs are not overcrowded and provide the best care as soon as a patient enters the hospital.

The simulation model built is attached to the project report in form of *.py* and *.ipynb* files for further reference.

VII. Future Work

To further improve the simulation model created for this project, various changes and additional features can be included. These changes are not limited to,

- Addition of a skill level for doctors and nurses can be introduced that confines these resources to a specific department. This can make sure that a significant amount of resources is exclusive to the busiest departments in the ED.
  This can also decrease the waiting time of patients during the earlier critical phase of their flow into the ED.
- A complexity increase can be introduced to the simulation system by defining the duration of every human resource's shift. This can ensure that the workers in the real-life scenario are not exhausted. In addition to this, it might be interesting to see how resources can work in coordination during the edge cases, such as shift end and shift begin periods.
- Further diagnostic evaluations can be introduced to the simulation system. This can include but are not limited to, sonography, ultrasound, MRI, etc. These extra diagnostics can provide alternatives to the original investigations in case of the unavailability of those machines during an emergency.
- Analysis of different simulation systems for their values of waiting room length or server (doctors, nurses, or machines) utilization can also be performed. The simulation program written for this project has functions that can already perform such analysis.

Acknowledgment

The author would like to thank Dr. Wu for an amazing semester!

References

[1] L. De Freitas, S. Goodacre, R. O'Hara, P. Thokala, and S. Hariharan, "Qualitative exploration of patient flow in a Caribbean emergency department," BMJ Open, vol. 10, no. 12, p. e041422, Dec. 2020, doi: https://doi.org/10.1136/bmjopen-2020-041422.

[2] "NACRS emergency department visits and lengths of stay — CIHI," www.cihi.ca. https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay

[3] M. Jones, "Using discrete event simulation to improve the patient care process in the emergency department of a rural Kentucky hospital.," doi: https://doi.org/10.18297/etd/709.