**Final Report**

**IE 5316**

**Simulation on a system (Chick-fil-A) by using SIMIO software**

**Problem definition:**

Chick-fil-A is one of the largest American fast food restaurant chains located at the Student Union building of Texas Tech University. This place is very popular especially during the lunch hour to the students because of its food quality and service. To maintain the quality of the service the system needs to be designed for maximum efficiency. Real-time data for two consecutive days (2 hours each day) was collected for the system. The data needs to be fitted to an appropriate distribution. Depending on the distribution a queuing model needs to be established. A simulation will be built to simulate the system by using SIMIO software in different terminating conditions. The result of the simulation needs to be compared with the actual output from the system. Further suggestions are required to improve the system. After collecting and fitting the data with the goodness of fit test, the following problem can be suggested:

This is M/M/5 model where Inter-Arrival time is exponentially distributed with mean 27 seconds and service time is exponentially distributed with mean of 100 seconds with the single queue and 5 servers in the system.

**Objective:**

The objective of this project work is to observe a system and collect real-time data. This data will be used to analyze the performance of the system by using SIMIO software. The result of the simulation will be compared to the actual result derived from queueing theory. Some objectives of this study are as follows:

1. Building a simulation model from real-life data obtained from an actual system (Chick-fil-A at Student Union)
2. Describing the key aspects of the system
3. Fitting the collected data to a statistical distribution model and building a simulation model by using SIMIO
4. Determining the desired accuracy of the simulation by alternating the number of replication and suggesting alternate model configurations to improve the efficiency of the system.

**Data collection and analysis:**

Arrival times and serving times Data were collected for 2 consecutive days. Lunchtime was chosen to observe the system. The system has a single queue and five servers. The data are tabulated in an excel file which is attached to this file.

We have observed 15 customers picked randomly with an interval of 32 seconds between each customer's arrival time and calculated their actual waiting time in the queue and system. Also, we collected the size of the queue lengths while each customer leave the system (refer to the attached excel file “Output Data”). On the other hand, we have collected a total of 100 customers data to calculate mean Inter-arrival time and collected 100 customers service time to calculate mean service time. Since we have picked a total of 15 customers data who have got served with a mean inter-arrival time of 32 seconds between them so the arrival rate and the service rate will be different from the actual arrival and service rate data which has been calculated in the attached excel file. As a result, an average number of customers in queue and system, average waiting time in queue, and system (for 15 customers) will vary accordingly from the actual data (with 100 customers mean IAT and 100 customers mean ST).

Inter arrival times and serving times in five servers of the system follow the exponential distribution curve which was determined by using StatFit software. We used the Chi-square test to confirm the goodness of fit. From the distribution curve we see the inter-arrival time and service time follows an exponential distribution.

![Graphical user interface

Description automatically generated with low confidence]()![Chart, histogram

Description automatically generated]()

**Figure 1**: Exponential distribution and goodness of fit for the inter-arrival time

A picture containing text, receipt

Description automatically generated

**Figure 2**: Chi-square test for goodness of fit of the serving time in five servers

Graphical user interface

Description automatically generated

**Figure 3**: Exponential distribution fit for the service time in five servers

**System’s key aspects and assumptions:**

The simulation model has been created based on the 100 data that was collected from the Chick-fill restaurant in Texas Tech SUB for 2 consecutive days from the restaurant opening time till lunch closing time. We run the model for 4 hours including the initial transient period to find the terminating simulation. We have run the model as a terminating simulation by considering the initial transient period and not including the warm-up period.

We also compared our simulation results with actual observed data and discussed the discrepancies below and run 4 replications as well. We also examined the system by increasing the mean Inter-Arrival rate by 25% and 50% and explained the situations on below. Additionally, if any extra servers are added then can the system handle it, also been discussed below.

**Initial Conditions for actual simulation run:**

Based on 100 collected data, Customers mean Inter-Arrival time is exponentially distributed with 27 seconds, with 5 servers each are also exponentially distributed with mean service time 111, 79, 107, 105, 100 seconds respectively. There is one exit point. Each server’s input buffer capacity is considered as zero. During the data collection, we also observed that, among 5 servers, 3 servers were also out of service when the queue length was less than 7. So, we have described those 3 servers’ uptime and break time by using triangular distribution in the reliability logic section by getting the observations of their minimum, mode, and maximum time.

We also defined the Average customers in Queue and Average time in Queue by using the output arrival node. Then, we have simulated for 4 hours.

The picture simulation model of the system is given below:

![Chart

Description automatically generated]()

**Figure 4**: Simulation model for the system

|  |  |  |
| --- | --- | --- |
| **Metrics** | **Actual Observed Data** | **Simulation Model Data** |
| Average Customers in System | 13 | 12.7761 |
| Average time in System | 365 seconds | 336.5927 seconds |
| Average Customers in Queue | 8 | 8.1098 |
| Average time in Queue | 210 seconds | 206.5753 seconds |

From the above table, we can say that a very small number of discrepancies have been found while running the simulation model data and then compared with the actual observed output data in the “Average time in the system” and “Average time in queue” matrices. Since we have taken actual observed output data based on 15 customers (picked randomly) whereas we run the model based on 100 data (customers) so this discrepancy arises. Also, their unit is in seconds so we can ignore this tiny discrepancy as well. Hence, all metrics validate our simulation model.

We have already validated our model against the actual observed data but while adding replications in our model, we see that 950 replications with smaller confidence intervals provide results that are not close enough to the actual observed output data of the system. This is because the initial transient period is included in the model and the warm-up period hasn't been considered. As the number of replications increases to 25, 50, 200, and 950, the half-width becomes smaller, and the estimations converge but do not improve a lot. Hence, more replications can increase our confidence on the results. To further improve the results, a longer run time can be used with more additional actual data.

**Increasing the demand by 25% and 50% and observation of the system:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metrics** | **Actual Observed Data** | **Simulation Model Data** | **Customers mean IAT increase by 25%** | **Customers mean IAT increase by 50%** |
| Average Customers in System | 13 | 12.7761 | 5.1631 | 3.8559 |
| Average time in System | 365 seconds | 336.5927 seconds | 188.1103 seconds | 156.9601 seconds |
| Average Customers in Queue | 8 | 8.1098 | 1.4375 | 0.5054 |
| Average time in Queue | 210 seconds | 206.5753 seconds | 52.2728 seconds | 21.1565 seconds |

From the above table, we can say that if mean IAT increases by 25% and 50% then still the system looks stable. Since there are 5 servers and customers mean IAT increases so it doesn’t affect on server’s utilization capacity as well but other metrices like Average customers in system, queue and Average time in system, queue gradually decreases from the actual observed data (see the attached excel “actual observed output data”).

**Adding additional resources and observation of the system:**

If the customers mean IAT increases by 50% and we add extra two servers’ then we can see that still the system is stable and these two additional servers 6 and server 7’s utilization capacity is only 22.1507% and 10.7708% that is if the system gets an unexpectedly huge number of customers, then these two extra servers can easily handle the system’s load alongside with other 5 servers.

Similarly, average customers in the system, queue, and average time in the system, queue decreases significantly compared to the actual observed data. The table is given here:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Servers** | **Utilization %** | **Average Customers in System** | **Average time in System** | **Average Customers in Queue** | **Average time in Queue** |
| 1 | 52.2125 | 3.3663 | 137.6850 seconds | 0.029 | 1.2055 seconds |
| 2 | 59.7305 |
| 3 | 37.4742 |
| 4 | 47.1957 |
| 5 | 26.6227 |
| 6 | 22.1507 |
| 7 | 10.7708 |

**Two alternative configurations to improve the system:**

**(1) First configuration:**

Since in our system we saw that among 5 servers, 3 servers were also out of service when the queue length was less than 7. The other two servers (server 2 and server 4) continuously operated. So we increased the server capacity from one to two for those two servers and simulated for four hours. We observed a significant decrease in all the servers utilization, average time and number in the system, and average time and number in the queue.

![Chart

Description automatically generated]()

**Figure 5**: Simulation model with increasing server capacity for server 2 and server 4

The comparison between the two model is shown in the table below:

|  |  |  |
| --- | --- | --- |
|  | Simulation  model | Alternate  configuration |
| Average number in system | 12.7761 | 4.0735 |
| Average time in system | 336.5927 | 116.4429 |
| Average number is queue | 8.1098 | 0.0994 |
| Average time in queue | 206.5753 | 2.8509 |
| Server 1 utilization | 58.5866 | 46.1463 |
| Server 2 utilization | 92.1841 | 63.6048 |
| Server 3 utilization | 51.6133 | 33.2105 |
| Server 4 utilization | 92.4171 | 52.0461 |
| Server 5 utilization | 43.4329 | 15.8987 |

**(2) Second configuration:**

|  |  |  |
| --- | --- | --- |
|  | Simulation  model | Alternate  configuration |
| Average number in system | 12.7761 | 6.3627 |
| Average time in system | 336.5927 | 171.4895 |
| Average number is queue | 8.1098 | 0.0126 |
| Average time in queue | 206.5753 | 0.3425 |
| Server 1 utilization | 58.5866 | 50.7862 |
| Server 2 utilization | 92.1841 | 76.8180 |
| Server 3 utilization | 51.6133 | 42.7386 |
| Server 4 utilization | 92.4171 | 64.7514 |
| Server 5 utilization | 43.4329 | 37.0386 |

In our system, we saw the five servers were used to take the order and payment. Food is also delivered in the same server, which increases the serving time. In the second configuration, we used three servers (servers 6 to 8) for taking orders and payment. After payment, the customer will go to a single queue and wait there for the food. The food will be delivered from one of the five servers (servers 1 to 5). This configuration decreases all the server's utilization, average time and number in the system, and average time and number in the queue. The comparison is given here:

![A picture containing text, sky, map

Description automatically generated]()

**Figure 6**: Simulation model with increasing number of servers for food order and payment

**Sensitivity analysis:**

**(1) Sensitivity analysis of the main model:**

**Chart, bar chart

Description automatically generated**

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(a) number in system

**Figure 7**: Response sensitivity of main model

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 8**: Response sensitivity of main model for 2% increase in arrival rate

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 9**: Response sensitivity of main model for 2% decrease in arrival rate

From the sensitivity test, we see time in the system and number in the system have a positive correlation with all the input parameters except arrival rate. For the main model (figure 7) we can see if the arrival rate changes 1 unit the time in system and number in system will change 161 units and 27500 units respectively in a negative direction. When the arrival rate is increased 2% (figure 8), time in the system and number in the system does not decrease significantly. But for 2% decrease in arrival rate significantly changes time and number in the system (figure 9).

**(2) Sensitivity analysis of the *alternate model - 1* (increased server capacity):**

**![Chart, bar chart

Description automatically generated]()**

(a) Time in system

**![Chart, bar chart, histogram

Description automatically generated]()**

(b) number in system

**Figure 10**: Response sensitivity of ***alternate model 1*** (increased server capacity)

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 11**: Response sensitivity of ***alternate model 1*** for 2% increase in arrival rate

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 12**: Response sensitivity of ***alternate model 1*** for 2% decrease in arrival rate

Figure 10, 11 and 12 show the response sensitivity for the ***alternate model 1*** in which the capacity for server 2 and 4 was increased. Here also, we see time in the system and number in the system has a positive correlation with all the input parameters except arrival rate.

Figure 10 shows if the arrival rate changes 1 unit the time in the system and number in the system will change 0.85 units and 700 units respectively in a negative direction. When the arrival rate is increased 2% (figure 11), time in the system and number in the system does not change significantly. But for 2% decrease in arrival rate significantly changes time and number in the system (figure 9).

**(2) Sensitivity analysis of the *alternate model - 2* (increased server):**

**![Chart, bar chart

Description automatically generated]()**

(a) Time in system

**![Chart, bar chart

Description automatically generated]()**

(b) number in system

**Figure 13**: Response sensitivity of ***alternate model 2*** (increased server)

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 14**: Response sensitivity of ***alternate model 2*** for 2% increase in arrival rate

![Chart, bar chart

Description automatically generated]()

(a) Time in system

![Chart, bar chart

Description automatically generated]()

(b) number in system

**Figure 15**: Response sensitivity of ***alternate model 2*** for 2% decrease in arrival rate

Figure 13 shows the response sensitivity of ***alternate model 2***. Arrival time and server 6 has negative correlation with the time in system and number in system. For one unit of change in arrival rate the time in system changes 95 unit and number in system changes 15000 unit. Figure 14 and 15 shows the response for 2% increase and 2% decrease in arrival rate respectively. From the plots we can see 2% decrease in arrival rate has a significant effect on the response of time in system and number in system.

**Steady state analysis:**

Steady state analysis has been calculated based on 100 customers arrival and service rate and their arrival and service rates are exponentially distributed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | **Utilization** | |
| **Servers** | **Arrival rate (in seconds)** | **Service rate (in seconds)** | **Steady State Utilization (in %)** | **Simulation Model Utilization (in %)** |
| **Server 1** | **0.00544** | **1/111** | **60.38** | **60.36** |
| **Server 2** | **0.0107** | **1/79** | **84.53** | **84.64** |
| **Server 3** | **0.00507** | **1/107** | **54.25** | **54.24** |
| **Server 4** | **0.007915** | **1/105** | **83.11** | **83.45** |
| **Server 5** | **0.004** | **1/100** | **40.01** | **40.01** |

**Conclusion:**

We made a model of **Chick-fil-A** using **SIMIO** simulation software based on real-time collected data for two consecutive days. The model has five servers in which order placement, payment and delivery is done. By doing data analysis we saw the inter-arrival time and server time followed exponential distribution. After running the simulation, we compared the result to values obtained by using the queuing theory. We also demonstrated the number of replications needed to be used to get the desired result. To improve the system, we proposed two alternate configurations in which we increased server capacities and server number for order placement and payment. We observed that ***alternate model 1*** in which we increased the capacity from one to two for server 2 and server 4 gave the best results. ***alternate model 2*** also gave a better result than the main model but adding three additional servers may increase the cost of the system. Also, it needs a larger area for the food restaurant. From our analysis, we conclude that by increasing the server capacity we improve the system.