M/M/1 Queuing Model

Garland Fire Department: Station 4

Coleman Strickland, Capt USMC

ABSTRACT

The purpose of this project is to build a M/M/1 queuing model for the Garland Fire Department (Station 4) in order to analyze utilization rates, mean times in the system, and average queue lengths for the city's fire and medical first responder system.

Table of Contents

Background

The City of Garland is part of the Dallas-Fort Worth Metroplex, contains over 235,000 citizens, and consists of three counties including Dallas, Collin, and Rockwall. With a city of this magnitude it is critical to ensure a first responder system is capable of supporting a multitude of emergency incidents and service each incident in a timely manner. The Garland Fire Department contains 11 fire stations to service all fire and medical emergency incidents. Station 4 (District 4) was opened in 2002 and is located in the southern portion of Garland. Station 4's equipment consists of Ambulance 4 (A4) and Truck 4 (T4). Station 4 is capable of supporting any incident throughout the city; however, priority is given to incidents occurring in District 4. This project will formulate and analyze a M/M/1 queuing model to evaluate the utilization rates, mean service times, and average queue lengths for Station 4's equipment from 2011-2016.

Assumptions and Data Description

Assumptions for this project are outlined below and are based on speaking with firemen at Station 4 as well as analyzing the data set. A layout of the data set is outlined in Table 2 and was obtained utilizing the str() function in R Studio.

Assumptions:

1. Units cannot service two incidents at one time.
2. Incident numbers cannot be utilized more than once.
3. Incidents cannot be denied service by Station 4.
4. If the equipment is available, it will service the next incident accordingly.
5. Alarm times are recorded when the dispatcher receives the incident.
6. Dispatch times are recorded when the station receives the incident.
7. Arrival times consist of the moment the unit arrives on scene.
8. Clear times are recorded when the unit is available to receive another incident.
9. The next incident arrival time is based on the next incident dispatch time.
10. A unit begins service according to the dispatch time.
11. A unit concludes service according to the clear time.
12. Inter-arrival times are calculated based on the dispatch times.
13. Both inter-arrival times and service times follow an exponential distribution.
14. All times calculated do not include preparation, cleaning, maintenance, or any other activity that the unit must accomplish before the next incident.

Data Frame Column Name	Variable Type	Description
Incident	Integer	Incident number
Alarm_Time	Character	Time incident arrived at dispatch
Incident.Type.Code	Numeric	Code for incident type
Incident_Type	Character	Description of incident type code
Apparatus	Character	Equipment dispatched
Dispatch_Time	Character	Equipment Dispatch Time
Arrival_Time	Character	Equipment Arrival Time
Clear_Time	Character	Equipment Clear Time
District	Integer	Location of incident by district
Number of Observations		**26,478**
Number of Variables		**9**

Table 2: Description of Garland Fire Department District 4 Data Set

Data Compilation

Incident data for this project is received though a Fire Analyst for the City of Garland. R Studio is utilized to compile and clean the data for further processing and is also used for building the queuing model. Data for this project ranges from 2011 to 2016 and contains only incident data for T4 and A4. The data contains all necessary information to build a M/M/1 queuing model to include the dispatch time, arrival time, and clear time. A formal definition and description of the model is provided in the Model Building section of this report. Lastly, the data includes the district in which the incident is located and the type of incident that has occurred.

After compiling the data, there are a few items of note. There are instances in which the incident numbers are duplicated for both equipment types without a district number, this error is recorded for 29 incidents. For example, incident number 1109240 was recorded on 05/30/2011 at 01:40:11 and was located in District 1, however, the same incident was also recorded on 09/24/2011 at 23:11:27 and the district is not specified. Because incident numbers are unique to each incident it is concluded that this is an anomaly in the data. In another example, incident number 1119523, features an occurrence in which the error occurred twice for one piece of equipment (A4) and only once for another (T4). Occurrences of this type arose a total of 29 times for T4 and 42 times for A4 in a data set containing 19,882 incidents. When combining the incidents for Station 4 together, these errors occurred 57 times across all incidents. The data reflected in Table 3.1 includes these incident errors; therefore, the total number of incidents from 2011-2016 will differ from the sum of the total number of incidents specifically in District 4 and outside of District 4. Additionally, Table 3.2 contains a list of incident numbers with these discrepancies.

Another item of note is the disparity between the alarm time and dispatch time. These times should only differ by matter of seconds; however, there are multiple occasions in which the difference was greater than five minutes. For example, the difference between the alarm time and dispatch time of incident number 1103115 was over nine minutes. After averaging the differences between the alarm time and dispatch times over all 19882 incidents, the average difference is 0.112 seconds, therefore, the model will use the dispatch times to calculate the inter-arrival times of an incident vice the alarm time because the difference between the two times is not significant.

Table 3.1 outlines statistics of the data given by T4, A4, and both pieces of equipment combined. The average number of incidents per year is calculated by taking the "Total Incidents (2011-2016)" value and dividing by the difference in the number of years. The average number of incidents per day is calculated by taking the "Total Incidents (2011-2016)" value and dividing by the total number of days between 2011 and 2016.

Calculation of average response time involves taking the difference between the arrival time and dispatch time and placing these values in a response time vector. Values that are less than or equal to zero are considered erroneous data entries and values greater than 60 minutes occurred only 4.38 percent of the time, therefore, these values were dropped from the vector. Lastly, the mean() function in R Studio is applied to the vector to obtain the values located in Table 3.1. Similarly, the time at each incident is obtained by calculating the difference between the clear time and arrival time. Values less than zero are considered erroneous data entries and are eliminated from the vector. Values greater than 720 minutes occurred 0.02 percent of the time; therefore, these values are also eliminated from the vector. The average inter-arrival time was calculated by obtaining the difference between the dispatch time of the next incident and the dispatch time of the current incident. Again, values less than zero are considered erroneous and values greater than 720 minutes occurred 2.06 percent of the time, therefore, these values were eliminated.

District 4 Data Statistics (2011-2016)			
Incidents			
	Station 4	T4	A4
Total Incidents (2011-2016)	19882*	12428**	13801**
Total Incidents in District 4	13681	9011	10339
Total Incidents Outside of District 4	6144	3388	3420
Average Number of Incidents per Year	3976.40	2485.60	2760.20
Average Number of Incidents per Day	10.89	6.80	7.56
Average Times			
	Station 4	T4	A4
Response Time (min)	4.83	4.90	4.77
Time at Incident (min)	28.36	18.50	36.76
Time until Next Incident (min)	113.34	199.36	200.99

Table 3.1: District 4 Data Statistics (2011-2016)

* Number of total incidents including the 57 discrepancies when combining T4 and A4.
** Number of total incidents including the 29 errors for T4 and 42 errors for A4.

Incident Numbers with Discrepancies	
1108108	1205845
1109240	1208865
1116320	1210438
1116371	1217110
1116393	1217213
1119523	1217462
1120048	1221655
1121061	1302566
1121573	1302695
1121574	1406480
1201938	1418627
1202766	1418632
1204163	1418645
1204511	1610114
1204934	

Table 3.2: Incident Numbers with Discrepancies

Model Building: M/M/1

A queuing system consists of entities arriving (i.e. customers, patients, etc.) and receiving a service at single or multiple stations (Kelton, 24). If the entity arrives and there are no service stations available the entity will wait in a queue for the next available server. A M/M/1 model processes entity inter-arrival and service times according to an exponential distribution and only one server is available to service an entity at a time. Inter-arrival times are defined as the average time until the next entity arrives for service. Similarly, the service time is defined as the average amount of time the server utilizes to serve one entity. Furthermore, a M/M/1 model has an infinite queue length resulting in customers never being denied service; they will wait for the server to become available. Figure 4 illustrates a M/M/1 queuing model with inter-arrival times of rate λ and service times of rate μ along with the formulas for calculating λ and μ. The M/M/1 model is stable if $\lambda < \mu$, the model is servicing entities faster than they are arriving.

$$\lambda = \frac{1}{\text{average interarrival time}} \qquad \mu = \frac{1}{\text{average service time}}$$

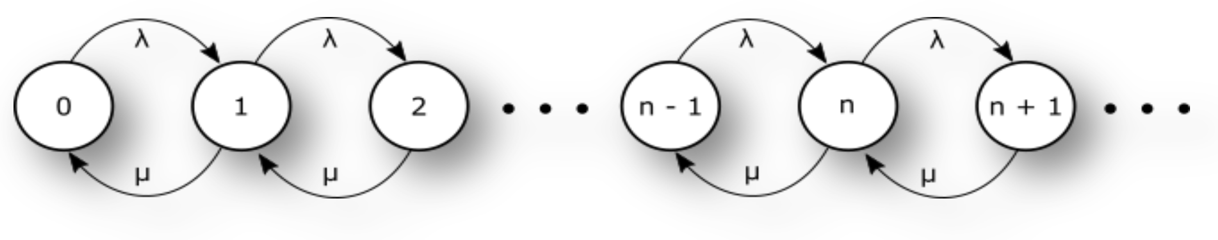

Figure 4: M/M/1 Queuing Model
Image Retrieved from: https://www.grotto-networking.com/figures/BBDelayBlocking/StateTransitionMM.png

The M/M/1 model is chosen for this project because the equipment at Station 4 can only service one incident at a time and customers will not be denied service if the server is unavailable, assumptions 1 and 3. Moreover, inter-arrival and service times follow an exponential distribution, assumption 12. Finally, all incident arrivals are serviced on a first-in, first-out (FIFO) process, assumption 4.

The project utilizes the "Queuing" package in R Studio. The description of the package states, "Provides a versatile tool for analysis of birth and death based Markovian Queuing Models and Single and Multiclass Product-Form Queuing Networks" (Canadilla, description file). The package is capable of handling multiple Markovian models to include a M/M/1 queue. The package is utilized to calculate the server utilization rate, average queue length, and mean time in system for the model.

Model Input Parameters

The M/M/1 queuing model is chosen for this project for two reasons: the inter-arrival and service times for incidents in District 4 follow an exponential distribution and the equipment at Station 4 can only serve one incident at a time. Before inputting the model parameters and input data, the inter-arrival and service times are calculated and plotted to ensure they meet the above criteria.

To calculate the inter-arrival time, the difference between the next time dispatch time and the current dispatch time is calculated and the average is taken over the vector. Similarly, the calculation of the average service time for the equipment involves computing the difference between the current dispatch time and the current clear time, and finding the average. Table 5.1 illustrates these calculations for incident number 1100006 and 1100008.

Inter-Arrival and Service Time Calculations		
Incident Number	1100006	1100008
Dispatch Time	01/01/11 1:28:34	01/01/11 1:40:39
Clear Time	01/01/11 1:53:31	01/01/11 1:53:08
Service Time (min)	24.95	12.48
Next Dispatch Time	01/01/11 1:40:39	01/01/11 2:14:12
Inter-Arrival Time (min)	12.08	33.55

Table 5.1: Inter-Arrival and Service Time Calculations

Figure 5.1 and Figure 5.2 illustrates the inter-arrival times and service times for incidents over the entire data set for both T4 and A4 together. An exponential distribution density curve with mean equal to the average time calculated is added to the plot to illustrate the goodness-of-fit relation of the data set and the exponential distribution. The inter-arrival times match the exponential curve almost perfectly and the assumption that the inter-arrival times are exponential is confirmed. The service times do not match the exponential distribution curve as well as the inter-arrival times; therefore, assumption 13 is necessary for model building.

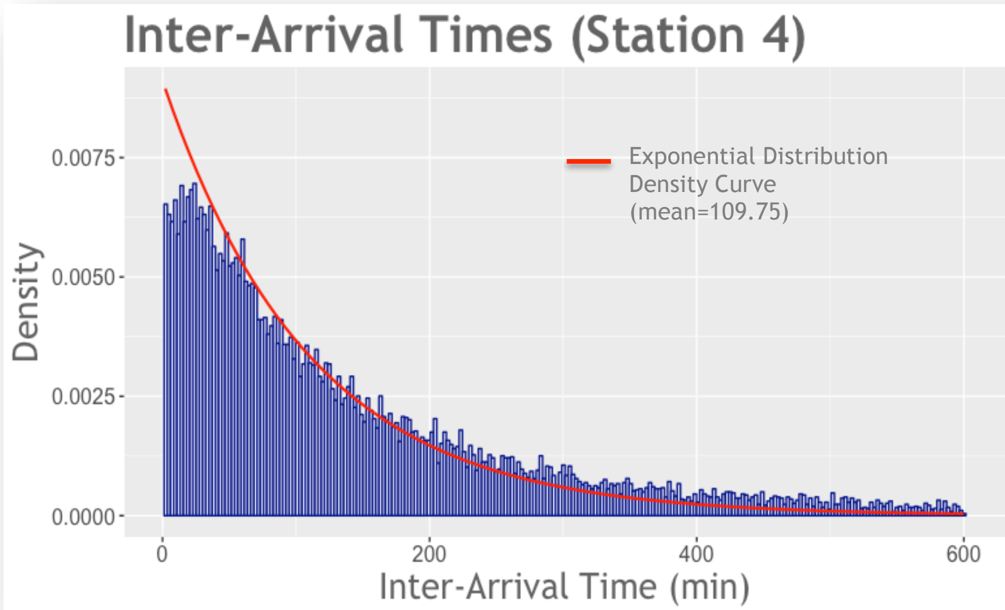

Figure 5.1: Plot of Inter-Arrival Times

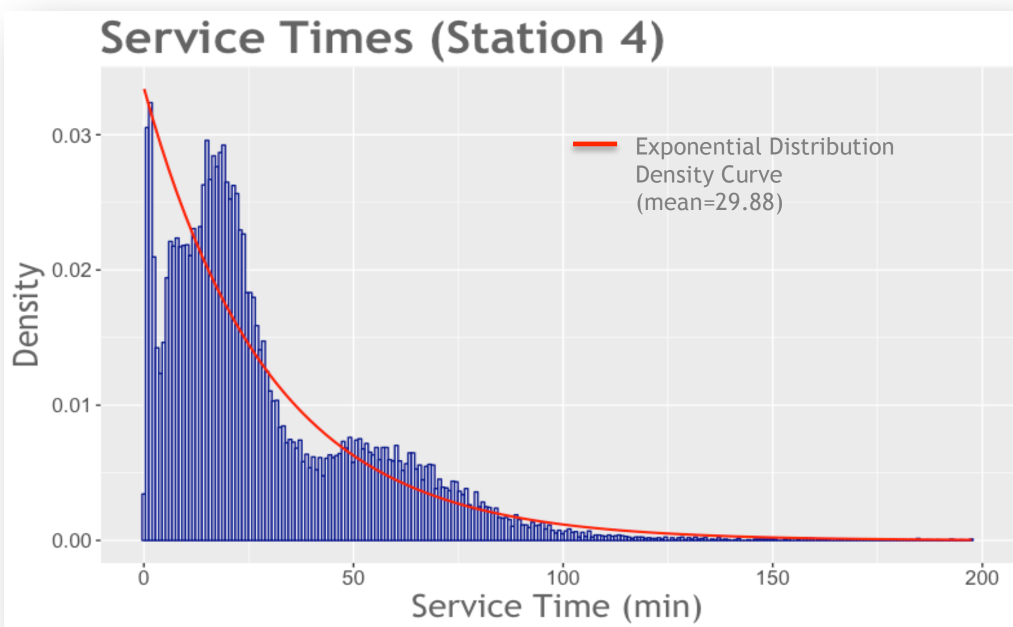

Figure 5.2: Plot of Service Times

Table 5.2 represents the mean inter-arrival time and mean service time for Station 4 (A4 and T4 combined) as well as the mean inter-arrival and mean service time when treating A4 and T4 as separate servers. Calculating individual times for each piece of equipment allows one to compute the utilization rates, mean queue lengths, and service times for both pieces of equipment separately. Table 5.2 also includes the inter-arrival and service time averages for each year of the data set. All model inputs utilize the inter-arrival and service times displayed in Table 5.2. In all instances, $\lambda < \mu$; meeting the stability conditions for the model for all input parameters. Finally, it is noted that the mean inter-arrival time decreases from 2011 to 2016, which should contribute to an increased utilization rate from 2011 to 2016.

Station 4: A4 and T4 Combined							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
Mean Inter-Arrival Time (min)	110.89	118.23	118.37	109.55	106.24	98.65	109.75
Mean Service Time (min)	28.93	32.58	29.69	30.96	29.26	28.39	29.88
A4							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
Mean Inter-Arrival Time (min)	190.11	210.38	207.19	194.49	195.42	183.36	196.27
Mean Service Time (min)	36.73	42.71	38.79	40.28	37.45	35.73	38.51
T4							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
Mean Inter-Arrival Time (min)	199.15	214.42	204.36	195.48	187.82	179.56	195.37
Mean Service Time (min)	19.82	20.24	19.37	20.31	20.78	20.82	20.27

Table 5.2: Inter-Arrival and Service Times by Year

Model Output and Analysis

Output data for the model focuses on three values: utilization rate, mean time an entity is in the system, and the mean queue length. Table 6 depicts the output data for each year represented by Station 4, and A4 and T4 separately.

Station 4: A4 and T4 Combined

	2011	2012	2013	2014	2015	2016	*Total*
Number of Incidents	3176	2929	3016	3368	3538	3797	***19824***
Utilization Rate (%)	26.09	27.56	25.09	28.26	27.54	28.77	***27.22***
Mean Time in System (min)	39.15	44.98	39.64	43.15	40.37	39.85	***41.06***
Mean Queue Length	0.092	0.105	0.084	0.111	0.105	0.116	***0.100***

A4

	2011	2012	2013	2014	2015	2016	*Total*
Number of Incidents	2245	2120	2115	2382	2381	2515	***13758***
Utilization Rate (%)	19.32	20.3	18.72	20.71	19.16	19.49	***19.62***
Mean Time in System (min)	45.53	53.59	47.73	50.8	46.33	44.38	***47.91***
Mean Queue Length	0.046	0.052	0.043	0.054	0.045	0.047	***0.048***

T4

	2011	2012	2013	2014	2015	2016	*Total*
Number of Incidents	1932	1838	1866	2061	2262	2440	***12399***
Utilization Rate (%)	9.95	9.44	9.48	10.39	11.06	11.59	***10.38***
Mean Time in System (min)	22.01	22.36	21.39	22.67	23.37	23.55	***22.62***
Mean Queue Length	0.011	0.009	0.009	0.012	0.014	0.015	***0.012***

Table 6: Output Data

A Portland Fire and Rescue Service Delivery System Study conducted in 2006 used the following formula to calculate the Unit Hour Utilization (UHU) rate to estimate the percentage of time a unit is occupied by emergency calls. The study states, "UHU measures the percent of a unit's time that is spent running calls" (TriData, 77).

$$UHU = \frac{(number\ of\ calls) * (average\ call\ duration\ in\ hours)}{(8{,}760\ hours\ per\ year)}$$

The study found that a medical unit with a UHU between 15-25 percent was optimal to maintain acceptable response times to emergency incidents. Moreover, if the UHU was above 40 percent the unit was not available to service the next emergency incident 40 percent of the time. Similarly, the study found that optimal UHU rates for fire units (trucks and engines) were between 5-15 percent. In addition, the study stated, "If a [fire] unit is out of its station on a call more than 10 percent of the time, then it is unlikely to meet response time goals of 90 percent of calls in four minute travel times" (TriData, 7). The study found that a UHU rate above 45 percent results in units not being available with enough frequency; therefore, response times begin to deteriorate as the city must dispatch calls to surrounding districts to support emergency incidents the current unit cannot service.

Relating the UHU rate to the utilization rate found from implementing the M/M/1 model, the output data supports that the individual units at Station 4 are between the optimal utilization rate ranges found in the Portland study. Furthermore, the utilization rate for Station 4 as a whole never rises above the 45 percent threshold; therefore, response times in District 4 do not suffer from over utilization of the equipment. The utilization rates found in the M/M/1 model conclude that Station 4 and its emergency response equipment combined are operating at the ideal utilization rate and that no additional units are recommended to be placed inside of District 4 to support the inter-arrival times of emergency incidents.

Figure 6.1 plots by year the incident count, utilization rates, and average time the entity is in the system for Station 4. The utilization rate average over the five-year period was 27.22 percent with a mean time in system of 41.06 minutes. It is important to note the service rate over the year range remained relatively constant even though the number of incidents increased by an average rate of 3.84 percent per year. According to the North Central Texas Council of Governments, the population for the City of Garland was 226,876 in 2010 and estimated to be 234,710 in 2017. This is an increase of approximately 0.4 percent or 1,120 citizens per year, which could be contributing to an increase in incident counts per year.

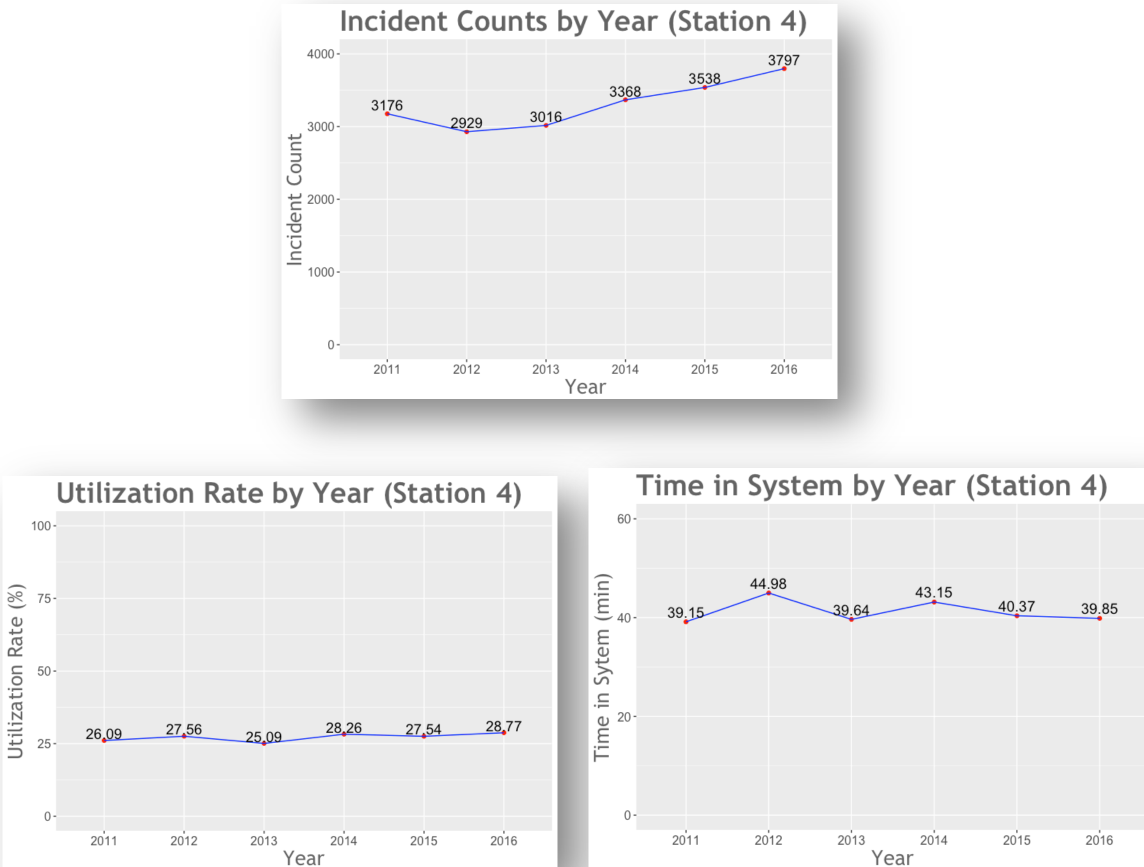

Figure 6.1: Station 4 Incident Counts, Utilization Rate, and Time in System

Figure 6.2 and Figure 6.3 show the incident count, utilization rate, and mean time in system per incident for T4 and A4 separately. The total utilization rate for A4 over the five-year period is 19.62 percent while T4 achieves a total utilization rate of 10.38 percent. The utilization rate for A4 remains reasonably stable as the incident count per year increases, contributing to a decrease in the mean arrival-time for each incident as depicted in Table 5.2. Additionally, the utilization rates for A4 achieve the optimality conditions for medical units as outlined in the Portland

Study who found optimal utilization rates for medical units should fall between 15-25 percent.

The utilization rate for T4 breaches the 10 percent threshold for optimal fire unit utilization rates as outlined in the Portland study. T4 shows a consistent increase in its utilization rate beginning in 2012, reaching a maximum of 11.59 percent in 2016. Referring to Table 5.2, the mean inter-arrival time of emergency incidents for T4 has decreased by 20 minutes from 2011 to 2016. Simultaneously, the time in system for each incident T4 services has increased by 1.5 minutes since 2011. These two actions have resulted in T4's utilization rate increasing by almost 1.5 percent from 2011 because the incident arrival rate is increasing while T4's ability to service each incident is also increasing.

Figure 5.2: A4 Incident Counts, Utilization Rate, and Time in System

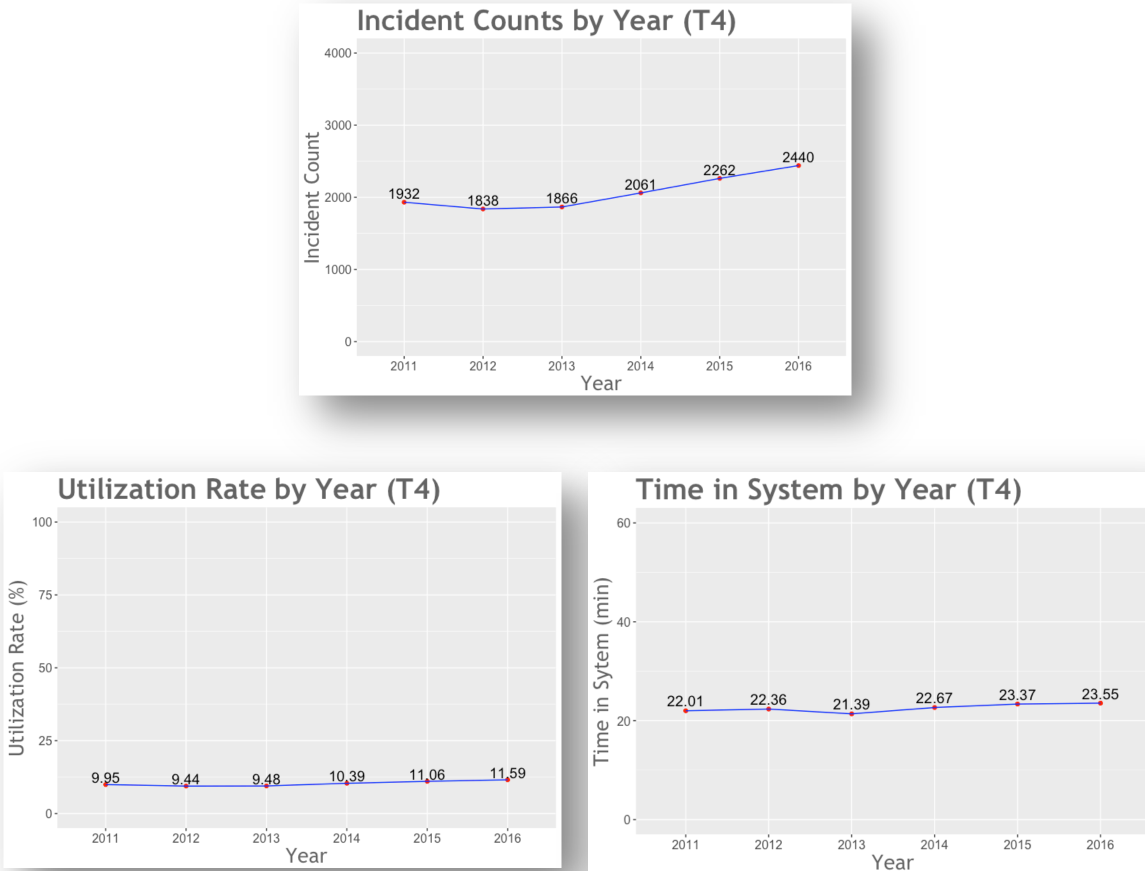

Figure 5.3: T4 Incident Counts, Utilization Rate, and Time in System

Transition times were also analyzed from 2011 to 2016 to determine if an increased transit time is contributing to the increase in service time for each incident. Once again, the model's service times were calculated by taking the difference between an incident's clear time and it's dispatch time. This calculation takes into account the transition time for the equipment to arrive on scene. Figure 5.4 displays the average transition times for Station 4, and A4 and T4 separately. These times were calculated by taking the difference between the incident arrival time and the incident dispatch time in order to compute the amount of time it takes the unit to leave Station 4 and arrive at the location of the emergency.

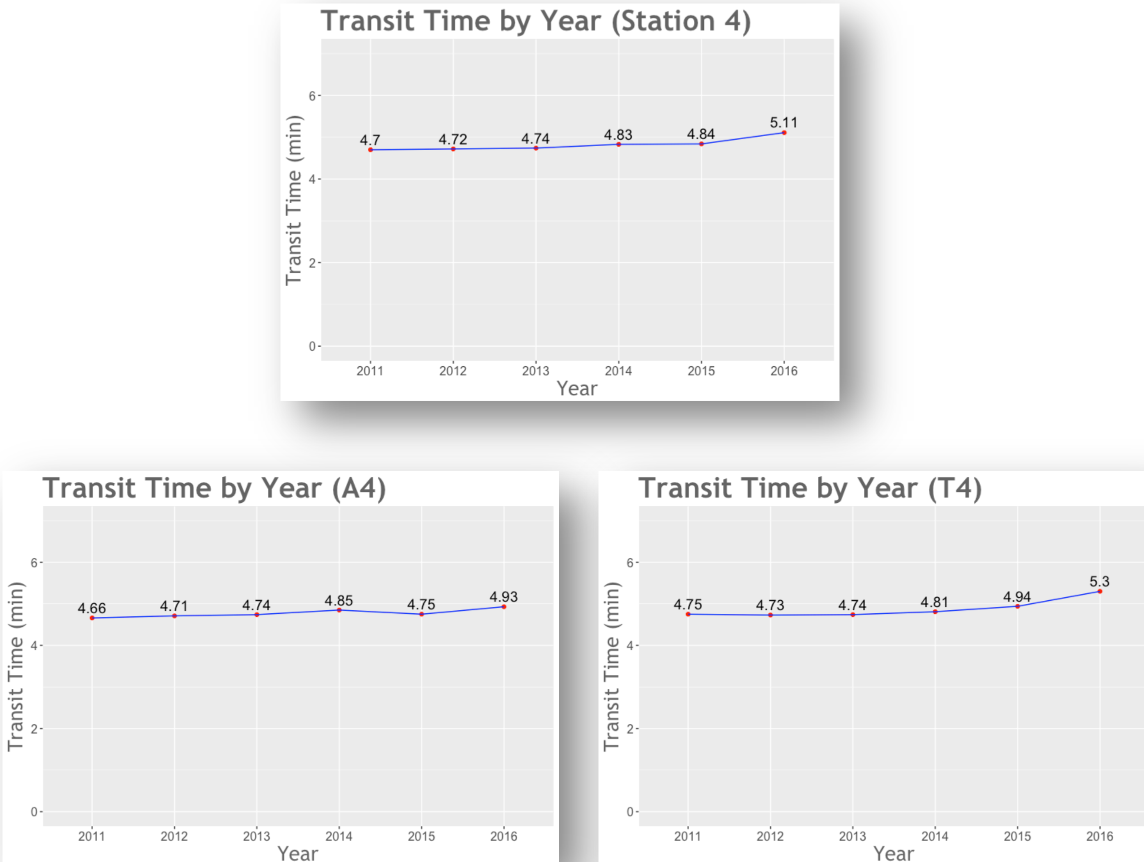

Figure 5.4: Station 4, A4, and T4 Transition Times

The transition time for each piece of equipment has increased over the five-year period. T4 has seen the most substantial increase of over 30 seconds since 2011. The increase in T4's transition time could be contributing to its increased server utilization rate and service time. With incident arrivals becoming more frequent and an increasing transition time interval contributing to a longer service rate, it is expected that T4's utilization percentage will continue to increase in the coming years if future data reflects historic trends.

Finally, the average queue length of the server reaches a maximum of 0.116 in 2016. A mean queue length below one ensures that Station 4 is servicing emergency incidents without multiple incidents waiting for service in the queue. If units are not available the City of Garland will leverage other districts or other cities for additional help to facilitate incoming emergency incidents; however, the model shows that Station 4's utilization rate is sufficient enough to keep the server available to service the next incident.

Conclusion

Pros of this model include it's simple design which contributes to it's ability to accept any data set with dispatch and clear times, and build a quality model to be implemented quickly and efficiently. Additionally, the model maintains the capacity to look at an individual fire station as a whole or a fire station's individual equipment separately. The benefit of this modification allows the user to determine which piece of equipment is being utilized the most and which equipment is contributing to the fire station's overall performance.

One drawback to the current model is modification is necessary in order to examine utilization rates of multiple fire stations at one time. For example, to see the utilization rates across the entire city of Garland, one would need to implement a different Markovian model to handle multiple servers at one time. While this model would be more complex, it would allow the user the ability to see the entire city of Garland's fire and medical first responder system at one time, and determine which stations are utilized the most and which stations are under utilized. Revision of the current M/M/1 model to accommodate this modification would be significant and would require a data set containing similar data for all 11 fire districts in Garland. Additionally, it would be necessary to confirm assumption 13, inter-arrival times and service times follow an exponential distribution, to ensure the M/M/1 model is appropriate for the data set.

From this model it can be concluded that Station 4 as a whole has continued to maintain an optimal utilization rate without an increase in equipment to help service incoming emergency incidents. The model meets all required metrics of a M/M/1 queue and accurately calculates the intended metrics set out to obtain: utilization rate, incident time in system, and average queue length. Overall, the model has proved the optimality conditions put forth by the Portland study and returns valuable feedback for investigating utilization percentages and service time rates for Station 4 and its associated equipment.

Works Cited

Kelton, David, Jeffrey Smith, and David Sturrock. *Simio and Simulation: Modeling, Analysis, Applications*. Simio LLC, 2014. Print.

Pedro Canadilla, Queueing, R 3.3.2, (March 2017)

"Portland Fire and Rescue Service Delivery System Study." *https://www.portlandoregon.gov/ fire/article/388484*. TriData, a Division of System Planning Corporation, March 2006

"2017 Population Estimates". North Central Texas Council of Governments (NCTCOG). April 2017