Optimization and Simulation of the Flow of Passengers Based on Time Series Analysis[1]

WANG, Sicheng
Shandong University,
School of Mathematics and Statistics,
Weihai, P. R. China
wangsc@mail.sdu.edu.cn

Abstract

To maximize security while minimizing inconvenience to passengers in airport seems conflicting. Our goal is to analyze the flow of passengers at the security checkpoint, and optimize the passenger throughput without reducing security. We analyze the time series to determine the proper probability distributions, by which we can generate more data. We select candidate distributions by empirical Cumulative Distribution Function (CDF) plots and Cullen & Frey graph, then we fit the data to candidate distributions and visually evaluate the goodness of fit. We eventually choose the best distribution after applying Kolmogorov-Smirnov test. Due to two types of passengers (Pre-Check and regular), we particularly assume a mixture distribution of two normal distributions, and estimate the parameters. We model passengers passing through checkpoints as abstract servers. We develop a core algorithm describing behavior of such servers. The algorithm takes three inputs: arrival time, service time, and the number of servers. And it generates the start and end time of service. We define the waiting time, which equals to the end time minus the arrival time, as main object of our interest. We simulate the flow of passengers based on our model and the generated data. By comparing results of different parameters, we develop some modifications to the current process to improve passenger throughput and reduce variance. Finally we perform a more detailed sensitivity analysis and discuss how we should accommodate changes. We simulate different traveler styles by generating data with different means and standard deviations. Then we provide some improved modifications to increase the efficiency of civil aviation security system.

Keywords: *cumulative distribution function, simulation, flow of passengers.*

AMS Classification: 93A30

[1]Citation: Wang S., *Optimization and Simulation of the Flow of Passengers Based on Time Series Analysis*, "An. Univ. Spiru Haret, Ser. Mat.-Inform.", 13(1), pp. 5-19, 2017.

1. ***Introduction***

At present, there is a tension between desires to maximize security while minimizing inconvenience to passengers for airports. During the year 2016, the US Transport Security Agency (TSA) has been the target of major criticism of extremely long airlines. As a result of this public attention, TSA has invested in several modifications to their equipment and control procedures and has increased staff in more congested airports [1, 9]. But there have also been incidents or unexplained and unpredicted long lines at other airports, including airports that normally have short wait times. This high variance in checkpoint lines can be extremely costly to passengers as they decide between arriving unnecessarily early or potentially missing their scheduled flight.

Our task is to develop models to identify potential bottlenecks that disrupt passenger throughput, then put forward creative solutions that both increase checkpoint throughput and reduce variance in wait time, next consider how cultural differences may impact the way in which passengers process through checkpoints as a sensitivity analysis, finally propose policy and procedural recommendations for the security managers based on our models.

2. ***Symbols, definitons and assumptions***

Table 1: Symbols used in the model

Symbols	*Definitions*
$X(i)$	*The arrival interval of the i-th passenger*
$T(i)$	*Service time of the i-th passenger*
$A(i)$	*The arrival time of the i-th passenger*
$S(i)$	*The start time of the service for the i-th passenger*
$E(i)$	*The end time of the service for the i-th passenger*
$WT(i)$	*The waiting time of the i-th passenger*
$RT(i)$	*The remaining time of the i-th passenger*

2.1. Assumptions

- We assume the end time of one service is exactly the arrival time of its successor.
- Passengers are always able to select the most efficient server.

3. ***Models***

3.1. Analyze the time series

Choice of candidate distributions [2]: In order to determine the distribution of given data, first of all, we draw histogram and cumulative distribution function plots of our sample data to get preliminary observation of its distribution. In addition to empirical plots, descriptive statistics may help to choose candidates to describe a distribution among a set of parametric distributions, especially the skewness and kurtosis, we draw Cullen and Frey graph which shows the relationship between square of skewness and kurtosis of different common distributions (For some distributions, such as normal, uniform, logistic and exponential, there is only one possible value for the skewness and the kurtosis, the distribution is represented by a single point on the plot. For other distributions, areas of possible values are in a line, such as the gamma and lognormal distributions; or larger areas like the beta distribution). By comparing the position of our observation with that of other distributions, we can select candidate distributions.

Fit of distributions by maximum likelihood estimation: Once selected, one or more parametric distributions may be fitted to the data set. Under the i.i.d. sample assumption, by maximizing the likelihood function defined as [4]:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{1}$$

After, we obtain the parameters of candidate distributions. Qualitatively, we draw four classical goodness-of-fit plots (Cullen and Frey 1999). Quantitative, we can apply the Kolmogorov–Smirnov test (K–S test) to get the p-value of those distributions. At last we can determine the most fitting distribution according to the combination of the results of K-S test and above-mentioned figures. We take two time series of examples.

Pre-Check Arrival Times and ID Check Process Time 1: As shown in Fig. 1 & 3, our observation samples and bootstrapped values are more likely one of the beta distribution family (Transformed beta, Burr, Pareto). As Fig. 2 & 4 shows, the distribution of ID check process time 1 is most likely one of normal, lognormal, gamma and beta distribution. And one shouldn't simply assume that service time be subject to exponential distribution.

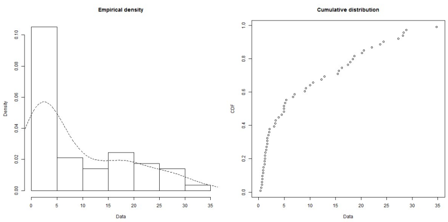

Figure 1. *Pre-Check Arrival Times*

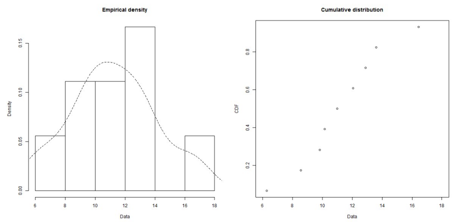

Figure 2. *ID Check Process Time 1*

Fig. 5 shows comparisons of sample data with candidate distributions and comparisons of candidate distributions in histogram and theoretical densities, Q-Q plot, empirical and theoretical CDFs and P-P plot. Visually we may not determine which one is better in Normal, Pareto and Transformed Beta. That's why we need K-S test.

The p-value means that we can accept the original hypothesis with the p-value as the minimum confidence level; the confidence level represents the probability of the type I error (the hypothesis is true but we reject the hypothesis). The greater p-value is, the greater risk of rejecting the real hypothesis.

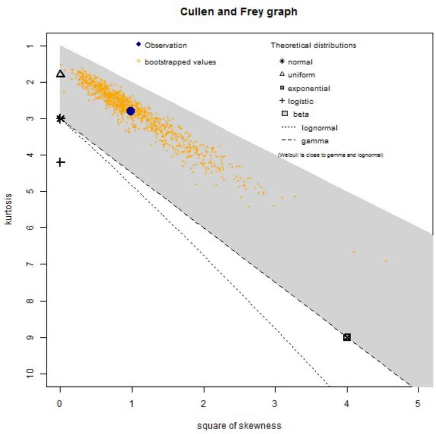

Figure 3. *Cullen and Frey graph of TSA Pre-Check Arrival Times*

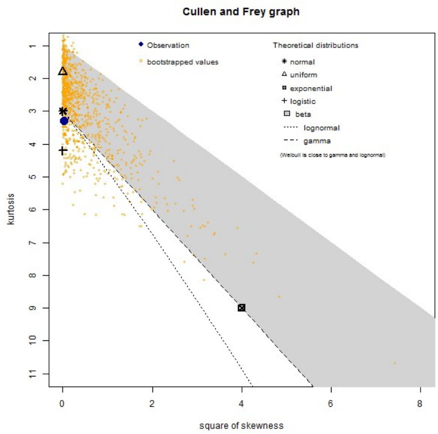

Figure 4. *Cullen and Frey graph of ID Check Process Time 1*

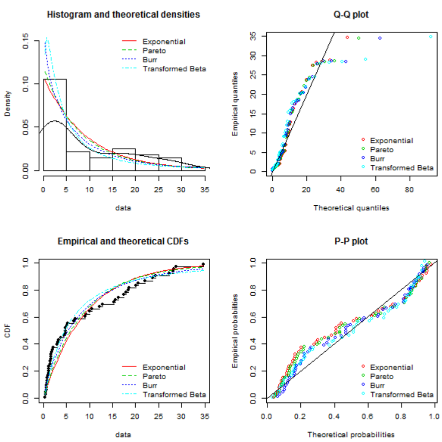

Figure 5. *Goodness-of-fit of Pre-Check Arrival Times*

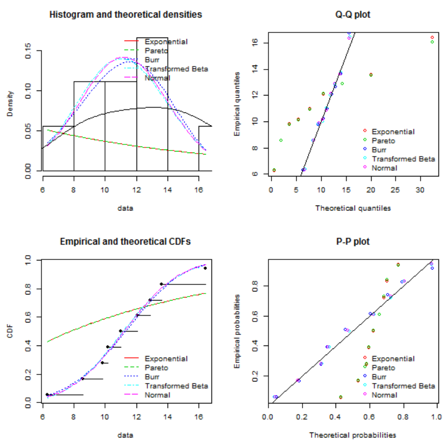

Figure 6. *Goodness-of-fit of ID Check Process Time 1*

Table 2: Goodness-of-fit statistics p-value of candidate distributions (exponential, pareto, burr, transformed beta, normal)

Candidate distributions	Exp	Pareto	Burr	Transformed Beta
p-value of Pre-check	0.0416	0.0811	0.322	0.3347
p-value of Regular	0.1404	0.0406	0.539	0.3535
p-value of ID check 1	0.052	0.0521	0.999	0.9997
p-value of ID check 2	0.0649	0.0642	0.976	0.9382

For Pre-Check arrival intervals, it should be considered subjecting to Transformed Beta distribution. But we will choose the seemly lesser optimal Burr distribution. Because the difference is small, and the Transformed Beta distribution has more parameters it means we are more likely over-fitting it. For regular arrival intervals, we should choose Burr distribution, too.

For both ID checking time, the best result is Normal distribution. However, the time should always be positive, so there might be theoretical risks of choosing Normal distribution as ID checking times distribution. But we eventually decide to use it because the probability of such cases are rather small, and Normal distributions are the simplest.

Now, the problem here is that the amount of current data is too small, obtaining more data will not be difficult. Even if we use a sufficiently meticulous and rigorous method, we cannot rule out that randomness leads to such a result, we can only say that, at a certain confidence level, the results will be satisfying, but not absolutely correct, this is not because defects of our methods, but due to the lack of data. Therefore, as a contracted team, we demand the TSA or the airport to collect more data.

Following, we will use Burr distribution as the basis for the simulation of arrival times and normal distribution to simulate time of ID check process.

Millimeter Wave Scan Time Intervals:

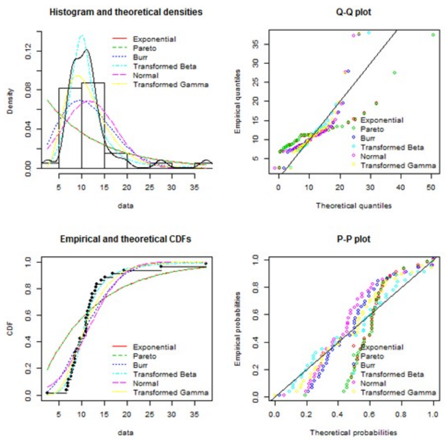

Figure 7. *Goodness-of-fit of Milimeter Wave Scan Time Intervals*

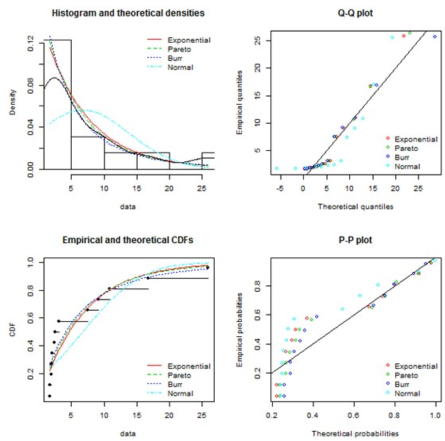

Figure 8. *Goodness-of-fit of X-ray Scan Time Intervals*

We can tell that the millimeter wave scan time intervals are subjecting to Transformed Beta distribution. As we can see from Fig. 8, X-ray scan time is subject to Exponential distribution or Burr distribution.

"Time to Get Scanned" Time in Zone B: For the "time to get scanned" however, we found it to be a multimodal distribution. Due to 2 types of passengers (Pre-Check and regular), we particularly assume it a mixture distribution of two normal distributions, and estimate the parameters.

As shown in Fig. 7 & 8, the black curve is the density curve of the time to scan all passengers. And the mean value of the red curve is smaller than the green one, we can deduce that the red curve represents the time of Pre-Check passenger while the green one represents the regular pax passenger. As the parameter estimated below in table 3, however, the proportion of the density curve of red one is 0.722, which indicates that there are approximately 72.2% passengers who enroll in the Pre-Check program while only 27.8% passengers get regular check. This might seem to contradict the 45% of Pre-Check passengers, but with such little data, we cannot sufficiently deny the assumption.

Table 3: Parameter Estimation of the time distribution

	lambda	mean value	sigma
Pre-Check	0.722	23.375	42.268
Regular	0.278	9.549	13.975

3.2. **Passengers passing through checkpoints as abstract servers**

Core Algorithm (Parallel Connection of Identical Servers):

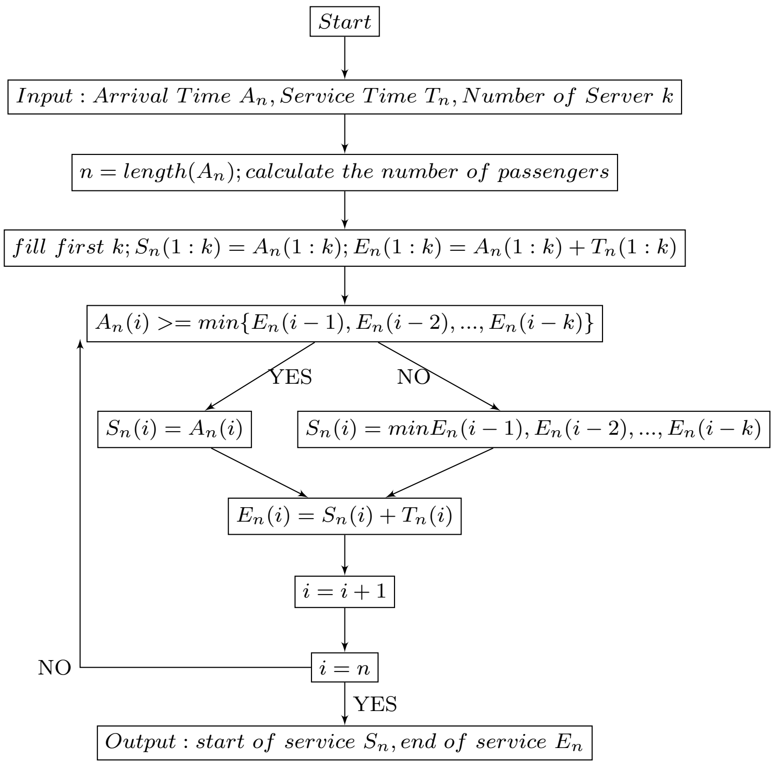

Figure 9. *Core Algorithm (Parallel Connection of Identical Servers)*

The algorithm takes 3 inputs: arrival time, service time, and number of servers (Fig. 9). And it generates start and end time of service. Assuming the end time of one service is exactly the arrival time of its successor, we build a series model.

Series connection:

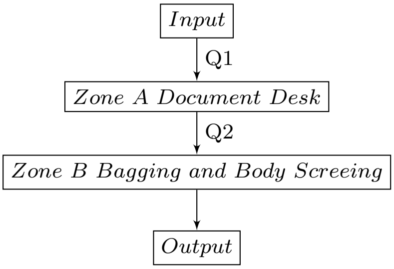

Figure 10. *Flow Chart of Zone A & B (Series Connection)*

The relationships simply are:

$$
\begin{aligned}
&X_A(i) \Rightarrow A_A(i), T_A(i) \Rightarrow S_A(i), S_A(i) \overset{T_A(i)}{\Rightarrow} E_A(i);\\
&E_A(i) = A_B(i), T_B(i) \Rightarrow S_B(i), S_B(i) \overset{T_B(i)}{\Rightarrow} E_B(i)
\end{aligned} \tag{2}
$$

Then we define waiting time and remaining time, particularly regard waiting time as a main object of our interest.

$$
\begin{aligned}
WT(i) &= S_A(i) - A_A(i)\\
RT(i) &= E_A(i) - A_A(i)
\end{aligned} \tag{3}
$$

However, this is not generalized enough. Because we specifically need each parallel servers be identical, we proposed a more general algorithm without such limits. For effective simulations, we will use the algorithm of identical servers.

4. ***Simulations***

We simulate 20,000 passengers in airport, however, based on the airport traffic data from Chicago Department of Aviation [6], there are above 200,000 passengers per day. And according to the given data, we will set 2 servers in Zone A, 1 line for Pre-Check and 3 lines for regular pax in Zone B.

We randomly generate arrival interval with 45% Pre-Check passengers using Burr distribution. In Zone A, we generate service time using normal distribution. In Zone B, we separately generate service time of these 2 types of passengers using normal distribution with different parameters. There is no applicable data of Zone C and Zone D. However, with proper amount of data, if given a probability from Zone B to Zone D, we can easily simulate them in the same manner.

Algorithm 1 Pseudocode of Core Algorithm: Parallel Connection of Identical Servers

Require: Arrival time, A_n; Time of service, T_n; Number of servers, k;
Ensure: Start of service time, S_n; End of service, E_n;

n=length(An); //calculate the number of passengers
for $v = 1$ to k **do**
    $S_n(i) = A_n(i);$
    $E_n(i) = A_n(i) + T_n(i);$
end for
for $i = k + 1$ to n **do**
    if $A_n(i) >= min(E_n(i-1), E_n(i-2), ..., E_n(i-k)$ **then**
        $S_n(i) = A_n(i)$
    else
        $S_n(i) = minE_n(i-1), E_n(i-2), ..., E_n(i-k);$
    end if
    $E_n(i) = S_n(i) + T_n(i);$
    $i = i + 1;$
end for

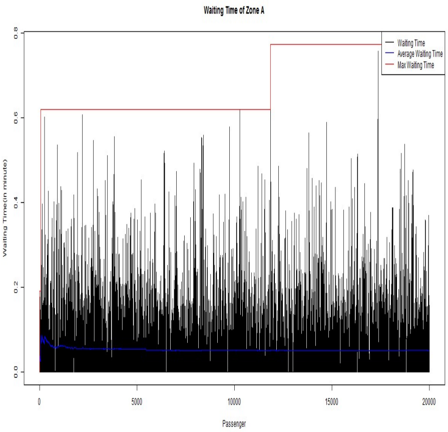

Figure 11. *Waiting: Zone A*

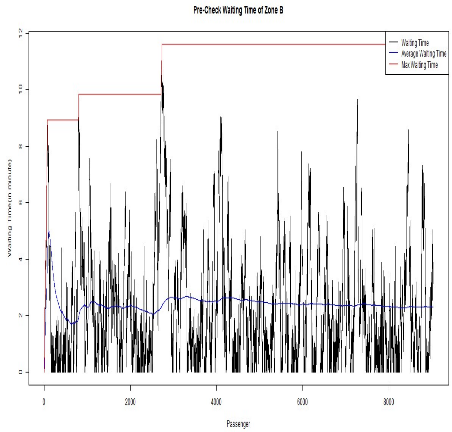

Figure 12. *Pre-Check: Zone B*

What's more interesting is that, in real life cases, the waiting time doesn't begin from zero, due to the fact the airport is opened all day, so people should expect much longer waiting time.

- The waiting time in zone A is fluctuant and lacks regularity with big variances.
- Surely, the maximum waiting time is rising with the number of passengers getting larger, and people who are unfortunately waiting for a long time are more likely to complain.

When people leave Zone A (i.e. after ID check process), we assume that they immediately arrive at Zone B, and ready to get scanned, that is, the ending time of service A equals the arrival at service B. And we discussed before that, for Pre-Check and Regular passengers, there are different distributions of service time in Zone B, so we simulate these people separately in zone B. (We can do this because in the beginning, when we generate these two types of passengers, we create new variables for labeling them.)

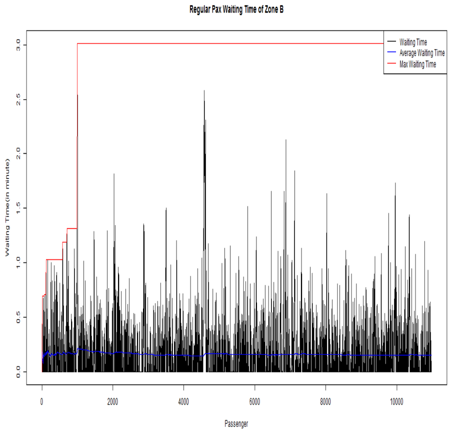

Figure 13. *Regular Pax Waiting Time of Zone B*

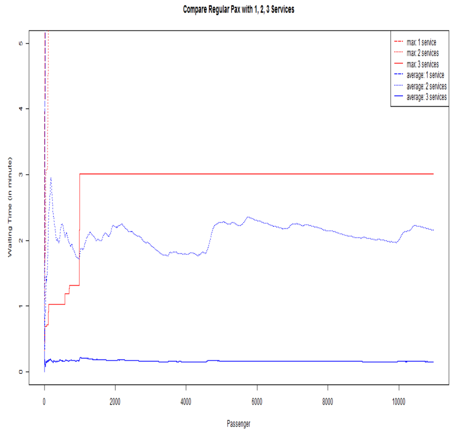

Figure 14. *Compare Regular Pax with 1, 2, and 3 Services*

The waiting times, average waiting times and max waiting times behave alike in the figure above. But with a much higher average of waiting time and maximum of waiting time. (Notice that the max waiting time is as high as 12 minutes!)

The waiting times and average waiting times and max waiting times behave alike in the figure above. But notice that whichever of waiting time (max or average), the regular pax is much smaller than Pre-Check, because 45% of Pre-Check passengers only take 25% of Scanning services, even though there the average service time is smaller.

As shown in figures above, we can easily tell that increase of the number of services will dramatically ameliorate the waiting time.

5. ***Results***

5.1. **Identify bottleneck**

Our simulation clearly shows that the bottleneck is in Pre-Check lines in Zone B. Also, we want to discuss this in a general case. We define service intensity:

$$Service\ Intensity = \frac{Average\ Service\ Time}{Average\ Arrival\ Time\ Interval} \tag{4}$$

Clearly, if the average service time is significantly longer than the arrival time interval, i.e. the service intensity is significantly larger than 1, based on the results of our simulation, one should expect an accumulation in waiting time and it' s very likely that a bottleneck exists.

With further data, we can firstly calculate the input time interval, and by our simulation algorithm we can calculate the output time interval as input interval of next service, and determine where the bottleneck is.

5.2. Modifications

Our goal is to increase throughput and reduce variance, since the lesser waiting time on average, the more throughput, so it could be rewritten as: decrease average waiting time, reduce variance.

1. Encourage more Pre-Check passengers and increase Pre-Check servers
In our model of service time in Zone B, the average service time of pre-check passenger is lesser than that of regular pax, let the number of servers be the same, we simulate the waiting time between pre-check and regular pax.

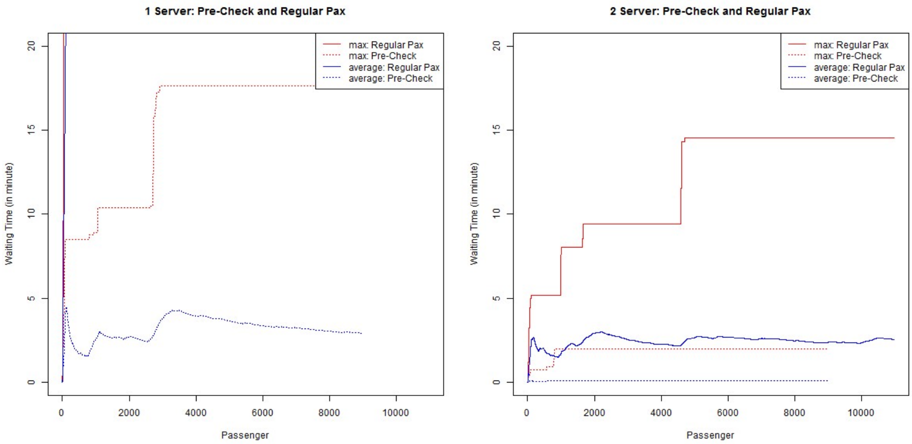

Figure 15. *Compare between Pre-Check and Regular Pax*

With the same number of servers, both max waiting time and average waiting time of regular pax is significantly larger than Pre-Check. With the same 4 servers, we recalculate the situations for 2 lanes for Pre-Check and 2 lanes for regular pax, see table 4.

Table 4: Total 4 lanes of scanner, different waiting times (minutes)

	Scheme1		Scheme2	
	Pre-Check 1	Regular 3	Pre-Check 2	Regular 2
max of server	11.92	3.01	1.35	13.94
average of server	2.56	0.15	0.07	2.16
max of Scheme	11.92		13.94	
average of Scheme	1.24		1.22	

Which means, without any extra cost, simply by rearrange the servers, we can reduce both average waiting time.

2. Increase the number of staff and enhance staff training

To increase the number of staff is equivalent to increase the number of servers; the advantages of training staff have two aspects: 1. Reduce the average service time; 2. To make the process more standardized, reduce the average standard deviation of service time. From the figure 16 we can see great changes brought by the change of number of servers. We change the standard deviation of the average service time and the service time. In the case of pre-check, we keep 1 server. We set the average service time from 15, 20, 25, 30 and the standard deviation from 7, 9, 11.

Table 5: Average waiting time and maximum waiting time

	Service Time							
	Waiting Time(Average)				Waiting Time(Max)			
Standard Deviation	15	20	25	30	15	20	25	30
7	0.1	0.3	3.46	285.17	1.46	4.72	3.44	544.85
9	0.13	0.4	3.75	283.19	2.32	3.65	16.24	543.73
11	0.15	0.42	4.65	292.56	2.37	4.86	19.64	555.81

- In all cases, the average waiting time and the maximum waiting time will increase with some rate in the average service time. In most cases, the average waiting time and the maximum waiting time increase with the standard deviation of the service time. Therefore, the training of staff to improve efficiency, reduces the average service time and service time standard deviation can effectively reduce the waiting time.

Table 6: Standard deviation of Waiting time

	Service Time			
	Waiting Time(Average)			
Standard Deviation	15	20	25	30
7	0.36	0.93	4.53	154.66
9	0.44	0.99	4.86	151.6
11	0.46	0.99	6.19	157.91

- The variance of the waiting time of the passengers (the square of the standard deviation) always increases with the increase of the average service time and the standard deviation of the service time. Therefore, the same approach can reduce the variance of waiting time.

3. Set coordination center

It should be noted that our simulations assume that passengers are always able to select the service desk which is the most efficient, but in reality the passenger cannot always go where he should go due to various factors. The airport is not necessarily timely adjusted according to the situation. In order to make the best results of the program, on the one hand, the airport needs to have good monitoring, that is, one timely bottleneck detection mechanism, on the other hand, the airport should adjust according to various situations in a timely manner, so it is necessary to introduce a coordination center mechanism.

4. Different kinds of passengers should be diverted

Such as whether the passenger has baggage, or whether the passenger will take international flight or not.

- **American**: known for deeply respecting and prioritizing the personal space of others, and there is a social stigma against cutting in front of others.
  Analysis: Respecting personal space will increase the length of the queue, which leads to an increase of arrival interval; Never cutting means the queue is prone to be more stable, which means that variance will be smaller, but might be a good optimal of efficiency.
- **Swiss**: known for their emphasis on collective efficiency.
  Analysis: Emphasis on collective efficiency means Swiss might be less selfish and focus more on average waiting time, and passengers will collaborate with each other to achieve a more optimal flow.
- **Chinese**: known for prioritizing individual efficiency.
  Analysis: Prioritizing individual efficiency means Chinese might focus on max waiting time, because complaints of individuals waiting will be amplified, and chaos is more likely to generate, which means larger variance.

We simulate different traveler styles by generating data with different means and standard deviations. We add a constant disturbance and a normal disturbance to the arrival interval, so the changes can be controlled.

The principle is as follows:

Due to the fact that it is complex to control the mean value and variance of the arrival interval by changing the modifying of the Burr distribution. So we consider setting the new variableX_i' by the equation: $X_i' = X_i + \varepsilon + C$, ε is a stochastic variable meeting the normal distribution $N(0, \delta^2)$, C is constant. The mean value and variance of the new variable can be calculated by the equation below:

$$\begin{aligned} E(X_i') &= E(X_i) + E(\varepsilon) = E(X_i) + C \\ Var(X_i') &= Var(X_i) + \delta^2 \end{aligned} \tag{5}$$

Then we apply control variables method to simulate the arrival interval of different passenger flow. First and foremost, we control the variance and change the mean value to simulate the situation that the passenger flow increases while the variance is almost unchanged. What' s more, we control the mean value and change the variance to simulate the situation that the passenger flow is almost unchanged while the variance of the passenger flow is apparently changed.

The results are as follows:

Table 7: Standard deviation of Waiting time

	Average Arrival Interval(s)			
	Waiting Time(Average) (min)			
s.d. of arrival interval(s)	9.9	10.9	11.9	12.9
12.74	11.8	2.88	0.9	0.52
14.74	11.31	2.83	0.87	0.59
16.74	12.12	3.72	1.05	0.63
18.74	12.04	2.28	1.11	0.63

With the arrival interval increasing, the waiting time (both average and max) and its standard deviation are drastically decreased, note that it's a tipping point if the service intensity is 1 (i.e. average arrival interval is equivalent to 10.90), if the passenger arrives 1 second faster on the average, that is, the traffic intensity increases by about 10% at this time, the waiting time and variance will become nearly four times larger than the original one.

An increase in the arrival interval variance within a reasonable range has relatively little effect on the waiting time and its variance. An increase in the variance of the arrival interval is likely to lead to an increase in waiting time and the variance of waiting time, but not always. Therefore, we need to pay special attention to the tipping point where service intensity equals 1. If the airport doesn't make timely adjustment at that point, the congestion will accumulate quickly. The specific approaches are:

- Establish real-time monitoring mechanism with particular attention to short-term peak traffic; make early warning for critical case.
- When the situation arises, as previously discussed, additional manpower should be added immediately to improve the situation quickly and to provide guidance to incoming passengers.

5.3. **Policy and procedural recommendations for the security managers based on our model**

1. The regular pax check time is quite greater than the Pre-Check time, so the Pre-Check program is supposed to be highly advocated.
2. The average serving time is rather long if the staff is unskilled, and the number of servers is rather small, so the transportation security agency is supposed to increase the number of staff and servers or enhance the level of service, i.e., reduce the service time and reduce the variance.

3. There is a rather serious problem in the arrangement for the servers. Approximately 45% passengers are Pre-Check passengers while there is only 1/4 of lanes for scanning. So the agency is supposed to rearrange the rate of servers for Pre-Check and regular passengers to a proper ratio.
4. There is a tension between remaining service efficiency and reducing waiting time, since service efficiency reaches the highest when the service intensity no less than 1, but 1 is exactly the tipping point. The agency must establish a coordination center to adjust the servers structure and dispatch the policemen at any time to solve unexpected emergencies.
5. Set up fast-track for the passengers without baggage or with few baggage to be screened in order to avoid unnecessary procedures and save waiting time.
6. Separate the passengers for domestic and international flights due to the rather different service time to avoid bottlenecks to minimize the average waiting time and maximum waiting time. The agency can carry out diversion and control in advance according to the predicted number in order to effectively reduce the congestion.
7. Divert the passengers and separate the flow of people into domestic and international passengers to control the service mean rate and variance.

6. ***Evaluation of our models***

6.1. **Strengths**

- We develop rigorous methods to detect the most interesting distributions of intervals.
- We first discuss special situations of series and parallel connection. From the particular to the general, then we consider the general simulation method of server model because the most general model is composed with series and parallel connection model. We establish a general simulation method of server model.
- The average and maximum waiting time can be easily affected by different parameters such as the mean value and variance of arrival time, the service intensity of checkpoints and so on. We apply control variable method to discuss the arrival interval by simulation.
- We combine the simulation results with real conditions to suggest policies and procedural recommendations for security manager.

6.2. **Weaknesses**

- Due to complicated distributions of arrival intervals, our methods rely on the simulation, but there are no analytical solution.
- Except the parameters mentioned, the stochastic phenomena can also influence the waiting time which are not mentioned in our model.

7. ***Future Work***

Addition to only caring about reducing the waiting time and its variance, we should take cost into consideration. With even more target function, there must be more conflictions of different variables. It will be more valuable if we can solve such problems. Also, it is crucial to develop analytic methods, for that simulation is not necessarily efficient.

Acknowledgements

This work was fully supported by National Training Program of Innovation and Entrepreneurship for Undergraduates and the Foundation of Shandong University.

References

1. Mao, X., Wu, Z., *The Optimizing of the Passenger Throughput at an Airport Security Checkpoint*, "Open Journal of Applied Sciences", 7, 485-501, 2017.
2. Cullen, A. C., Frey, H. C., *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Springer, Berlin, 1999.
3. Kovcs, G., Harmati, I., Kiss, B., *Methods for Airport Terminal Passenger Flow Simulation*, "International Journal of Mathematics & Computers in Simulation", 6(6), 529-541, 2012.
4. Delignette-Muller, Marie Laure, and C. Dutang, *fitdistrplus : An R Package for Fitting Distributions*, "Journal of Statistical Software", 64(4), 1-34, 2015.
5. Dutang, Christophe, V. Goulet, and M. Pigeon, *An R Package for Actuarial Science*, "Journal of Statistical Software", 25(7), 1-37, 2008.
6. Leone, Kelly, R. Liu, *Improving Airport Security Screening Checkpoint Operations in the US via Paced System Desgin*, "Journal of Air Transport Management", 17(2), 62-67, 2011.
7. Chawdhry, Pravir K., *Risk Modeling and Simulation of Airport Passenger Departures Process*, "Winter Simulation Conference", 2009.
8. Flight Data, Air Traffic Data, http://www.flychicago.com/OHare/EN/AboutUs/ Facts/Pages/Air-Traffic-Data.aspx.
9. Optimizing the Passenger Throughput at an Airport Security Checkpoint, ICM2017, http://www.comap.com/undergraduate/contests/mcm/contests/2017/problems/2017_ICM_Problem_D.pdf