Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		1234567

An ICM Paper Made by Team 1234567

Summary

People around the world are complaining about the traffic jams every $day^{(1)}$. Some people may spend more than 3 hours on their way from office to home everyday on their cars, which makes them frustrated. And our problem is to investigate the methodsto optimize the design of toll plazas on highways, especially the area after merging,which is also a possible location where vehicles can be stuck in.

Before establishment of our models, we list some assumptions to make the real lifescenario easier to model.

And then we start to analyze the existing models, from which we conclude their strengths and weaknesses. By investigating their characteristics, we get the inspiration to form our two new models: Control Time Model (CTM) andWaiting Area Model(WAM). In these two new models, we introduced a way of conducting control of the departure time of the vehicles at the tollbooths. This kind of control is basically preventing the vehicles which are merging in the same lane from leaving the booths simultaneously, which may be a hidden danger for traffic accident or leads to a traffic jam. And we?ll continue to calculate the size and shape of the merging area according to our control methods and some assumptions. After that, we introduce a method for finding the optimal merging pattern based on both mathematical proof and computer simulations.

After that, we run some simulations to find out the throughput, risk and cost of different models, which are based on some statistical laws for real life situations. Then we compare these three models in all factors with the help of statistical hypothesis testing, and conclude that the CTM is the best in general. The next section is about some slight modification under different conditions such as including the self-driving car, and the different arrangements in terms of merging patterns when the proportion of different types of tollbooths vary.

For the following section, we test our model by investigating the sensitivity of construction cost and throughput (per hour) in terms of some variables included in our model, in order to justify the reliability of our model from different perspectives.

Finally, we end our report by the conclusion part followed by strengths and weaknesses analysis.

Contents

1 Introduction

1.1 Background

These are increasingly challenging times, many problems can not be solved well even by the team. In this case, seeking a more reasonable team strategy and applying a more appropriate team plan can stimulate everyone's energy to the greatest extent and tide over the difficulties together. As a traditional form of team activities, sports team is no exception. With the development of science and technology, more optimized team strategy can be used to greatly improve the existing achievements on the basis of the original team members, so that the sports teams can obtain better team grades. Based on the performance of the players on the field, the corresponding data can be used to judge the tacit understanding between the players and the fit between the players and the team, so as to provide the direction for the team to strengthen and adjust in the off-season. The results of team optimization strategies in sports groups can also be referenced by other groups in society. Taking good use of the strategy, the team's ability of handling problems can be improved to a higher level.

1.2 Restatement of the problem

We are required to explore the construction of a strong team in competitive team sports, and we choose to study about football team formation and optimization. We get the data of performance of Huskies last season, based on it, not only the behavior and the interaction leading to scoring but also the team's dynamic changes throughout the season can be acquired, then we build a network model. The model will show the performance indicators of teamwork and team performance. Generally speaking, it reflects the structure, configuration, and dynamics of team work. After finishing the above work, we're able to give suggestions both on adjustment for the whole team and direction of changes for specific players. After the research on the Huskies team is completed, we will generalize the results to the more general team formation problem. We will talk about the construction of a strong team that can solve complex problems and the performance issues needing to be considered for the team, based on our model. In short, our work is just next three tasks:

- Using the network model, we build a model to reflect the structure, configuration, and dynamics of team work at multiple scales.
- Based on the results of our model, we provide strategies on adjustment of the whole team and corresponding suggestions for specific players.
- We should generalize the model, deliver our opinions on the construction of a strong team able to work out complex problems.

1.3 Literature review

In the early 1950s, Charles Reep had pointed out that "The key to scoring goals and winning games is to move the ball forward and back as quickly as possible", thus indirectly began the long ball movement of soccer. And Jordi Duch, Joshua S. Waitzman and Luís A. Nunes Amaral (2010), holds that the composition of the team determine their odds of success. Later in 2013, Ricardo Duarte(2013) construct a model to assess the dynamics of team–team and player–team synchrony in professional association football, the model quantized the relationship between people and the teams they are in, the fits of the players are depend their position on the gound. Put the $[0,100]$ convert to $[-\pi,\pi]$, and let θ_k, the phase time-series obtained with Hilbert transform, indicate palyer's movement in radians $[-\pi,\pi]$, where k = 1,2...,n. so that the cluster phase time-series can be calculated as:

$$r(t_i) = \frac{1}{n}\sum_{k=1}^{n} exp(i\theta_k(t_i))$$

and:

$$r(t_i)_1 = atan2(r(t_i))$$

where $i = \sqrt{-1}$, $r(t_i)$ $r(t_i)_1$are the resulting cluster phase in complex and radian form. So that the relative phase time-series, ϕ_k, between each player and the group cluster phase is then equal to:

$$\phi_k(t_i) = \theta_k(t_i) - r(t_i)$$

with the mean relative phase $\bar{\phi}_k$ for every player k, with respect to the team caculated from:

$$\bar{\phi}'_k = \frac{1}{N}\sum_{1}^{N} exp(i\phi_k(t_i)) \quad \bar{\phi}_k = atan2(\phi'_k)$$

where N is the number of time steps t_i, ϕ'_k and ϕ_k is the mean cluster phase in complex and radian $[-\pi,\pi]$ form. Note that ϕ_k gets the phase shift of the movement associated with the team behavior $r(t_i)$. For steady synchronization, it can be used to compare whether a person's movement has the same mean phase as a team's movement, thus determinie the relative phase relationship between the movements.

Finally, synchronization's continuous degree of the team ,the cluster amplitude $\rho_{group,i}$at every time step t_i can be described as:

$$\rho_{group}(t_i) = |\frac{1}{n}\sum_{k=1}^{n} expi(\phi_k(t_i) - \bar{\phi}_k)|$$

where $\rho_{group,i} \in$ [0,1]and group synchronization's average degree is computed as

$$\rho_{group} = \frac{1}{N}\sum_{i=1} N\rho_{group,i}$$

we can judge whether the group is in synchronization by caculating ρ_{group}, the larger the value of the ρ_{group}, the better of the fitness. This model can be used as a test for how well a player fits into the team.

2 Assumptions

To make our problems easier to deal with, we make the following basic assumptions, each of which is properly justified.

- All of the statistic and the analysis are based on the player's performance on the ground. That means we don't take the outside factors into consideration. (some player's performance can be affected by accident, which will lead to the bad game state.)
- According to the statistic we have got, we regard the numbers of passing between two players as the standard of the tacit understanding. The more frequent the passing between two people, the better relationship they have and so as the tacit understanding.

- Since we don' t konw weather the ball went in or not, when juding the player's abilitity of scoring, we take the shooting times instead of the actual goals he got this seaso as the critera n.
- The number and success rate of cooperation on the field represent the relationship between two players. A good relationship between two players can promote the team to achieve better results

3 Notations

Notation	Definition	Unit
X_{ij}	$EventOrigin_x$ of the player(ij correspond the player's field position, eg.X_{F1})	N
Y_{ij}	$EventOrigin_y$ of the player(ij correspond the player's field position, eg.Y_{F1})	N
$C_{i\to j}$	The number of times i pass to j	N
Q_{ho}	The number of gay people infected	N
Q_{he}	The number of heterosexuals infected	N
Q_{nd}	Number of drug users in year n	N
Q_p	The number of pregnant women	N/A
N_y	Number of sexual encounters per year	N
P_{ucho}	Gay people who do not use condoms to get infected	N/A
P_i	The probability of getting infected without using a condom	N/A
P_{uche}	Heterosexual people do not use condoms to get infected	N/A
P_p	Proportion of pregnant women	N/A
P_a	The percentage of pregnant women infected	N/A
η	Percentage of drug users	N/A
ε	The rate of AIDS among drug users	N/A
δ	The percentage of babies infected	N/A

4 ... Models ...

We first analyze the four tasks and identify the algorithm using cellular automata as the core of our model, Our model includes a model of free transmission and a model based on which drug intervention is introduced, The implementation of these models for cellular automata is to add the properties of each "cell" and adjust the relevant parameters, we will put the code in the appendix, In order to make the expression of the model more intuitive, we use the form of column differential equations in the paper to describe the influence of each factor on the dependent variable,But in essence, the idea of cellular automata and differential equation modeling is still the same.

On the basis of our own thinking, we also cited some existing models to make our work smoother, including the model of population growth, the model of virus response to drugs.The introduction of these models can make our results more accurate

4.1 Passing network model model(PNM)

We consider the influence of several factors and establish a Free propagation model to reveal the development principle of the AIDS under on intervention.

4.1.1 Assumptions

We only present the eleven people with the most passes and catch on the picture. Since substitutes play less and have less data, they are not representative enough. For the sake of simplifying data processing and data visualization, we just leave people with the most passes and catches on the picture. We won't take long-term or frequent injuries into consideration so that the data reflects exactly the normal level of every player. The range of players' activities on the court is limited and fixed, and their frequency of occurrence at other positions is lower or just zero. In football games, players have their own responsibilities and corresponding activities in the team. For example, the position of the defender and goalkeeper is relatively backward, and the Midfielder is relatively forward. Based on this, we can use the average of the coordinates of the player's active position in all matches as the player's coordinates on the visualization. In the game, important members of the team or star players will control more management and opportunities for passing. The important members need to get more passes to improve the team's performance in the game. Based on this, we can use the number of passes related to a player to characterize his importance in the team

4.1.2 Passing network model construction

In task 1, there is no consideration about the intervention, so when we construct model about the expected rate of change in the number of AIDS, we only take the factors above. And the longevity of the man who is diagnosed with AIDS is relatively short. This issue will be discussed in the following. By exploring the influence of those factors mentioned above, we can predict the future rate of change in the number of AIDS in South Africa.

$$\frac{dQ_i}{dt} = \sum_{n}^{n+1} Q_o + \sum_{n}^{n+1} Q_e + \sum_{n}^{n+1} Q_{nd} - Q_{(n-1)d} + \sum_{n}^{n+1} Q_p\delta - \sum_{n}^{n+1} P_{ua}$$

In this model :

- Q_o indicate the number of gay people infected in this year
- Q_e indicate the number of heterosexuals infected in this year
- Q_n Q_{n-1} indicate the number of people in year n andyear n-1, we will construct model about the Natural population growth in the following part.
- Q_p indicate the number of baby born in this year
- δ is the percentage of the babies who are born with HIV

Form the expression, we can find that

$\sum_{n}^{n+1} Q_o + \sum_{n}^{n+1} Q_e$ indicate the infected member increased by sexual contact

$\sum_{n}^{n+1} Q_{nd} - Q_{(n-1)d}$ indicate the infected member increased by injecting drug

$\sum_{n}^{n+1} Q_p\delta$ indicate the infected member increased by mother-to-fetus transmission

$\sum_{n}^{n+1} P_{ua}$ indicate the infected member reduced because death, since there is no way to cure AIDS now, so dying is the only way to decrease the infected member.

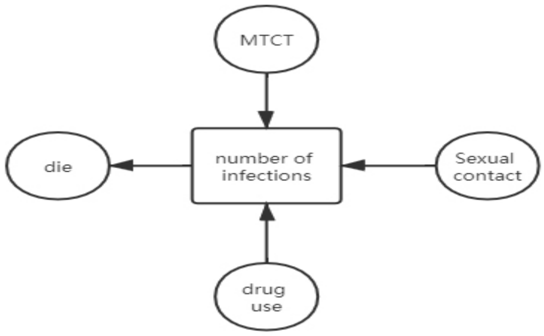


Figure 1: Four main factors affect the number of infections

Number of sexually transmitted infections

The nunmber of sexually transmitted infections an be expressed as

$$Q = Q_o + Qe = \sum_{n}^{n+1} Q_{ho} * (1-(1-P_{ucho}P_i)^{N_y}) + \sum_{n}^{n+1} Q_{he} * (1-(1-P_{uche}P_i)^{N_y})$$

In this exprssion:

- Q_{ho} indicate the number of homosexuals
- Q_{he} indicate the number of homosexuals.
- N_y is the number of sexual encounters per year, which is assumed to be same between homosexuals and homosexuals. We will change it in the sensitivity analysis part.
- P_{ucho} P_{uche}indicates the probabilities gay people and heterosexual people who do not use condoms to get infected. According the research, P_{ucho}, the percentage is so high that it means people use condoms only aiming for preventing them from having babies. And we can assume the percentage can be relatively higher in the gay people. We will also test this in the sensitivity analysis part
- P_i is the the probability of getting infected without using a condom.

Every sex act between susceptible people can cause the spread of AIDS. And the probability of infection is equal to $P_{ucho} * P_i$, After having sex a hundred times a year, we can conclude that the probability of not contracting AIDS is equal to $(1-P_{ucho} * P_i)^{N_y}$, This is a binomial distribution experiment, And the probability of getting AIDS is$1-(1-P_{ucho}P_i)^{N_y}$

N_y is a constant. Considering the social pressure and different social status, these two group can act diffrently theoretically speaking. and in the futher work, we will introduce social factors to estimate people's eager for sexual life. The higher the pressure is, the more likely they want turn to their sexual partner to have relax. And although homosexuality is recognized in South Africa,Gay people still don't get a fair deal in many social situations. We can see later, weather the change of N_y is vital to our model.

the infected member increased by injecting drug

the infected member increased by injecting drug can be expressed as

$$\sum_{n}^{n+1} (Q_n - Q_{n-1})\eta\varepsilon$$

In this exprssion:

- η is the percentage of drug users,
- ε is the rate of AIDS among drug users

Drugs spread AIDS mainly by sharing syringes. However, studies on individuals or small groups exchanging syringes are too complicated and lack relevant data. So we assume that the rate of HIV in the drug users is constantε So we can predict the number of people infected with HIV due to the increase in drug use if we know the change in the number of drug users each year. In the course of studying changes in the number of drug users each year,we consider the year-over-year changes in the number of drug users in South Africa. For this reason, Based on whether or not to consider the government to curb the situation of drug abuse we established two models of the change in the number of drug users. In the following discussion, to simplify the model, let's assume that the proportion of drug users remains the same. η is the percentage of drug users.

the infected member increased by mother-to-fetus transmission

the infected member increased by mother-to-fetus transmission can be expressed as

$$\sum_{n}^{n+1} Q_p\delta = Q_n * P_p * P_a * \delta$$

In this exprssion:

- δ is the probability of producing AIDS babies
- P_p is the proportion of the total population who are pregnant
- P_a indicates the rate of infection among pregnant women

To calculate the increase in the number of AIDS babies caused by mother-to-child transmission, we first obtain the Qy of the proportion of pregnant women in the population. Pregnant women are divided into pregnant women and pregnant women. The babies born by HIV-negative pregnant women are naturally also HIV-negative, while HIV-positive pregnant women are likely to give birth to HIV-positive and HIV-negative babies. Therefore, we checked the data to find that the proportion of HIV-positive babies in the two infants was δ. Therefore, the number of new AIDS babies generated by mother-to-child transmission in the population is the total number the proportion of pregnant women t the proportion of AIDS patients in pregnant women a the probability of producing AIDS babies δ. Considering that our inspection period is one year, multiple births or multiple births in one year are not considered.

the infected member reduced because death

We count deaths between susceptible populations and patients separately. Because of the short life span of patients after the onset of AIDS, generally in two or three years, we can consider AIDS and its induced diseases as the primary cause of patient death compared to natural death factors, and ignore the natural death factors of AIDS patients. The average survival time of AIDS patients after infection is x years, so when the patient's illness time reaches x years, we judge it as death and clear all its data. The deaths of susceptible people are determined by the average life expectancy and natural mortality. Every year, among the susceptible people, the death tolls of all ages are determined according to the proportion of deaths of all ages in South Africa. The sum of the two is the total number of deaths.

population growth

Our population growth model is estimated by a logistic retarded growth model,which can be expressed as

$$\frac{dP}{dt} = r * (a - P/K) * P$$

Among them, P is the total number of people in the area at time t, r is the maximum growth rate of the population in the area, and K is the upper limit of the population load in the area. The general solution is $P(t) = r/[ce^{-rt} + r/K]$, and c is a constant determined by the initial value. Divide P to the left of the equal sign, which is

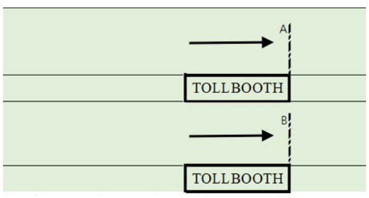
$$\frac{1}{p} * \frac{dp}{dt} = a + bp$$

where a = r, b = -r / K. We can use the method of least squares estimation to substitute South Africa's data from 2000 to 2018. Regression calculation shows that r = 0.263727 and K = 112703846, that is, the maximum annual growth rate of South Africa is 26.3727%, and the upper limit of population carrying is 112703846. At this time, c = 4.144387586816374*10^{220} can be calculated, and the annual number of people can be calculated by$P(t) = r/[ce^{-rt} + r/K]$.The process of regression analysis can be found in the appendix.

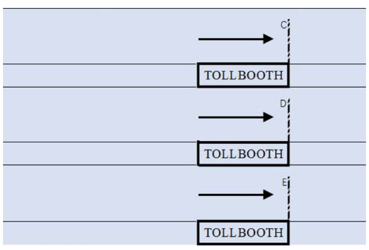
4.2 Control Time Model (CTM)

Given aforementioned deficiencies of existing merging pattern, we propose a new model, partially based on the current one. Instead of having all the vehicles moving and merging at their own discretion, control time model will control the departure time of vehicles to ensure a smooth and safe emerging process.

Specifically, for situations where two booths merge into one lane, the second vehicle will only be allowed to proceed t0 seconds after the first vehicle moves forward. The time t0 is defined as the control time. Similarly, for situations where three booths merge into one lane, the third vehicle will be allowed to proceed t0 seconds after the second vehicle moves forward, whilst the second vehicle t0 seconds after the first vehicle, as shown in Figure 1



(a) Either A or B is open.

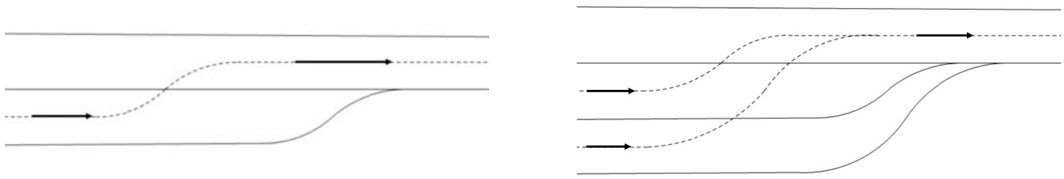


(b) Either C, D or E is open.

Figure 2

In this way, the regulated merging of the vehicles into another lane would be more efficient than the situation where vehicles are proceeding without regulation, for drivers should take time to make decision when multiple booths merge into one lane simultaneously, let alone the risk for doing so.

We model that vehicles start with constant acceleration as until reach the maximum speed vm in the straight path. They then immediately starts merging into their prescribed road in two consecutive tangent circle arcs.



(a) (b)

Figure 3: Two or three tollbooth egress lanes merge into one lane

Now we evaluate the "appropriate" control time. When emergency happens, one vehicle take severe brake action with acceleration ab, after response time tres, the posterior vehicle take severe brake action with the same acceleration. Consider the distance of the two vehicles (t = 0 is the time when emergencies happen):

$$d = v_m t_{rea} + \int_0^t v_{posterior}(t)dt - \int_0^t v_{previous}(t)dt$$

where

$$v_{previous} = \begin{cases} v_m, & t \leq t_{rea} \\ v_m - a_b(t - t_{rea}), & t_{rea} < t \leq t_{rea} + \frac{v_m}{a_b} \\ 0, & t_{rea} + \frac{v_m}{a_b} \leq t \end{cases} \tag{1}$$

Solve the equation $d \geq 0$, we get $t_0 \geq tres$. By saying "appropriate" control time, we take t_0= $1s \geq t_{res} = 0.8s$ (Lee et al, 2002).

After describing the control time model, we first calculate the throughput of the toll plaza.

Let i_k be the number of booths corresponding with the k^{th} lane. For example, if the first three booths merge into the first lane, we say $i_1 = 3$.

By computer simulation (details are included in Appendix), we have t_{ik} close to $\frac{T}{i_k}$ when i_k $\ll \frac{T}{t_0}$ (especially when the variance of service time is small), and t_{ik} increases as i_k increases, and the increase speed is quite large when $i_k \geq \frac{T}{t_0}$. where tik is the averaged timefor a car to come into k_{th} lane.

Therefore, to maximize the efficiency, we'll keep i_k as small as possible, whereas the throughput comes suboptimal if $i_k \geq \frac{T}{t_0}$ for some k. The optimal throughput is closed to $\frac{B}{T}$ car(s) per second.

By substituting B by 8, the expectation of human service time T by 15. According to our optimization method above, when we keep the i_k as small as possible, which is in other words, let them distributed evenly(e.g. 2,3,3). We will get the optimal throughput closed to $\frac{B}{T} \approx 0.533$ car per second.

Then we calculate the size of the plaza, by doing so we discuss the value of D.

Define $f : \{1, 2, \ldots ,\text{B}\} \to \{1, 2, \ldots ,\text{L}\}$ which is non-decreasing, and $f(i) = j$ if and only if vehicles from i_{th} booth merge into j_{th} lane. It obviously follows that $i_k = \| f^{-1}(\{\text{k}\})\|$. We call the mapping f a "merging pattern".

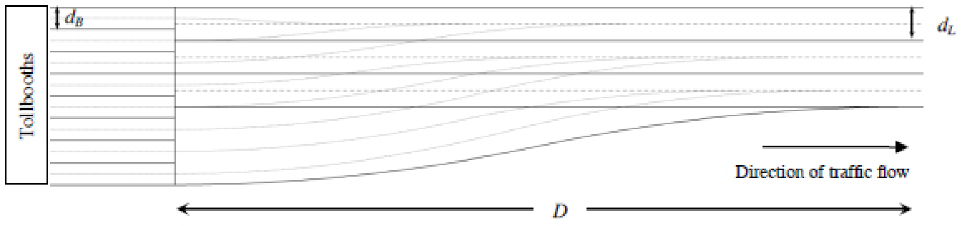


Figure 4: B = 8 tollbooth egress lanes merge into L = 3 lanes

By geometric relations, we have

$$(\frac{D-\frac{v_m^2}{2a}}{2})^2+(r-\frac{(i-\frac{1}{2})d_B-(j-\frac{1}{2})d_L}{2})^2=r^2$$

Observed from above relation, a certain (i, j) prescribes a lower bound of D, and for a certain merging pattern, the distance between the tollbooth and the end of the plaza that we eventually employ is the greatest lower bound of D. The less the greatest lower bound is, the smaller the size will be, and so will the cost of plaza. By property of quadratic function, for a certain j, the greater i is, the greater D is. We have

$$\sum_{k=1}^{j} i_k d_B - j d_L \geq (\sum_{k=1}^{j} i_k + 1)d_B - (k+1)d_L$$

since $d_L \geq d_B$. That implies if j is changed into j + 1 whilst i is changed into i + 1, D will get smaller. Therefore, the global maximum must be obtained at one of the local maximum points, and then we let x be

$$x=\max_{j\in\{1,2,\ldots,L\}}\left\{\frac{(\sum_{k=1}^{j} i_k-\frac{1}{2})d_B-(j-\frac{1}{2})d_L}{2}\right\}$$

And for given B and L, the optimal choice of merging pattern f would be the case if we can minimize x to be

$$x_{min}=\min_{\{i_k\}\in\{\|f^{-1}(\{k\})\|:\forall f\}} x$$

After trying out some pairs of small (B,L), we induce that for any given small number B, L (which can be put in practical use) satisfying $B \geq L$, there exists a merging pattern f_m such that

$$x=\frac{(B-\frac{1}{2})d_B-(L-\frac{1}{2})d_l}{2}:=x_0$$

Since for any ik we have the relation $f(B) = L$, which implies above expression will appear in every calculation of x for all i_k, hence $x_{min} \leq x_0$, hence the existence of x_0 guarantees ”=”, providing us with the optimal merging pattern to minimize D.

By substituting B = 8, L = 3, d_B = 2.5m, d_L = 3.75m, r = 115m, v_m = 15m/s, a =2.78m/s^2, the minimum of D is 105.5m, when (i1, i2, i3) $\in$ {(3, 3, 2), (3, 2, 3), (2, 3, 3)}.

4.3 Waiting Area Model(WAM)

After completing the above model, something seems to be deficient also appears in our model. In the control time model, vehicles may need to wait at the booths even after the service finished, while there are a lot of vehicles queuing after it. So we wonder that whether it is better to build a waiting area at a plaza (as shown in Figure 5), so that the vehicles whose service has finished can go to the waiting area, then the consecutive vehicle can be serviced. So in this model, the staff spending more time serving instead of doing nothing while the vehicles are just waiting at the booths

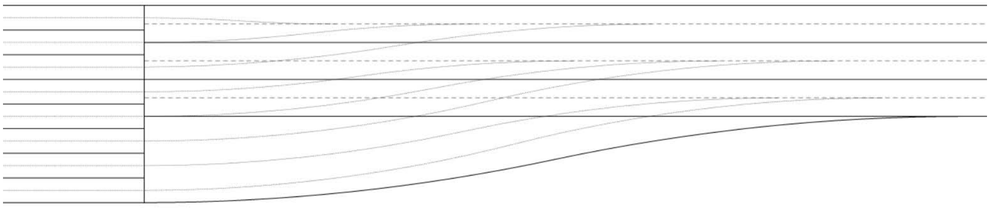


Figure 5: B = 8 tollbooth egress lanes merge into L = 3 lanes

We will test and compare this model with control time model in the following sections,to see whether it is efficient or not.

5 Analysis and Results

5.1 Throughput

Throughput (θ) here is defined to be the number of vehicles per hour passing the point where the end of the plaza joins one of the the L outgoing traffic lanes. Here we discuss only one outgoing lane merged from two lanes in human-staffed tollbooth area in heavy traffic where vehicles come continuously during heavy traffic so that the cashier work continually, as shown in Figure 6.

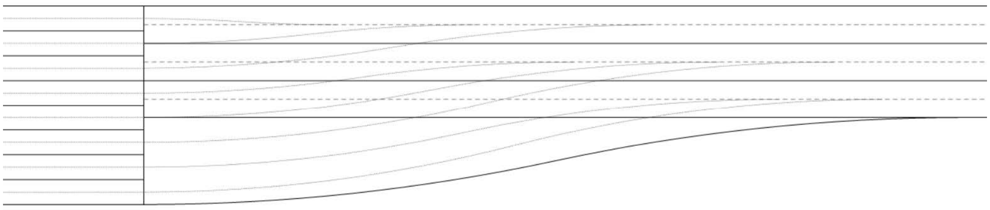


Figure 6: 8 tollbooth egress lanes merge intoL = 3 lanes in heavy traffic

5.1.1 Throughput of existing model

The toll collection service time is a random variable which can be seen as a sum of random variables including time needed by driver to search money or credit vehicled, time needed by cashier to take related record, etc. According to the central limit theorem, we can assume that the service time of ith vehicle ($t_{h,i}$) of the each lane follow normal distribution with μ = 15 and σ^2 = 4 independently and identically.

6 Strengths and Weaknesses

6.1 Strengths

- First one...
- Second one ...

6.2 Weaknesses

- Only one ...

Memorandum

To: Heishan Yan
From: Team XXXXXXX
Date: October 1st, 2019
Subject: A better choice than MS Word: LaTeX

In the memo, we want to introduce you an alternate typesetting program to the prevailing MS Word: **LaTeX**. In fact, the history of LaTeX is even longer than that of MS Word. In 1970s, the famous computer scientist Donald Knuth first came out with a typesetting program, which named TeX ...

Firstly, ...

Secondly, ...

Lastly, ...

According to all those mentioned above, it is really worth to have a try on LaTeX!

References

[1] Abdel-Aty, M. A., & Radwan, A. E. (2000). *Modeling traffi c accident occurrence andinvolvement*. Accident Analysis & Prevention, 32(5), 633-642.

[2] Corwin, I., Ganatra, S., & Rozenblyum, N. (2005). *A Single-vehicle Interaction Model of Traffic for a Highway Toll Plaza*. The UMAP Journal, 26(223), 299-315.

[3] Friedman, D. A., & Waldfogel, J. (1995). *The administrative and compliance cost of manual highway toll collection: evidence from Massachusetts and New Jersey*. National Tax Journal, 217-228.

Appendix A: Further on LaTeX

To clarify the importance of using LaTeX in MCM or ICM, several points need to be covered, which are ...

To be more specific, ...

All in all, ...

Anyway, nobody **really** needs such appendix ...