Efficient dispatching rules on double tracks with heterogeneous train traffic

SHI MU and MAGED DESSOUKY*

Daniel J. Epstein Department of Industrial and Systems Engineering
University of Southern California
Los Angeles, CA 90089-0193
Phone: (213) 740-4891   Fax: (213) 740-1120
maged@usc.edu

*Corresponding author

Abstract

The most natural and popular dispatching rule for double-track segments is to dedicate one track for trains traveling in one direction. However, sometimes passenger trains have to share some portions of the railway with freight trains and passenger trains are traveling faster and faster nowadays. The major drawback of this dedicated rule is that a fast train can be caught behind a slow train and experience significant knock-on delay. In this paper, we propose a switchable dispatching policy for a double-track segment. The new dispatching rule enables the fast train to pass the slow train by using the track traveled by trains in the opposite direction if the track is empty. We use queueing theory techniques to derive the delay functions of this policy. The numerical experiments show that a switchable policy can reduce the fast train knock-on delay by as high as 30% compared to a dedicated policy. When there are crossovers at the middle of the double-track segment, our proposed switchable policy can reduce the delay of the fast trains by as high as 65%.

1. Introduction

Railway operation is an effective mode to transport both people and goods. The freight transportation industry uses railway to move goods because it is relatively cheap compared to trucks and air; people take trains because it is typically faster and safer than cars and yet less expensive than airplanes. Nowadays, passenger trains are traveling faster and faster. However sometimes passenger trains have to share some portions of the railway with freight trains, because in some places, it is not possible to expand the railway infrastructure. For example, in the downtown area of Los Angeles, freight railway companies like BNSF and Union Pacific share rail tracks with Metrolink and Amtrak which are passenger train companies. Also in China, the lower grade of their high speed lines (usually running at speeds between 120 and 160 mph) comprise both passenger and freight trains. Mixed usage of infrastructure is also very common in central European countries. Since the passenger trains travel much faster than freight trains, if a passenger train is caught behind a freight train, the passenger train will experience severe delay. Even if the track is dedicated for passenger trains, they might travel at different speeds, with the faster passenger train being blocked by a slower passenger train.

To determine if the current infrastructure can handle certain traffic loads or to compare different dispatching rules, it is imperative to have a robust delay estimation technique that accurately predicts the travel time delay given different operating conditions. It is common to see double-track railway segments between two major intersections, the most natural and popular dispatching rule is to dedicate one track for trains traveling in one direction. The major drawback of this simple rule is that a fast train can be easily caught behind a slow train thus experiencing significant knock-on delay. In this paper, we focus on minimizing the knock-on/secondary delay, not the existing/primary delay. We propose a "smarter" dispatching policy and estimate the knock-on delays of this policy for fast and slow trains on double-track segments.

There has been some prior work in delay estimation of railway operations.

Frank (1966) studies delay on a single track. He considers a single train speed and deterministic travel times. Chen et al. (1990) present a technique to estimate the delay for different types of trains on a single track. Harker et al. (1990) extend Chen's model to partially double-track sections. Ozekici et al. (1994) use Markov chain techniques to study the knock-on delays under different dispatching policies and arrival patterns of passengers. They construct a transition matrix to calculate the expected delay. Higgins et al. (1998) also present an analytically based model to compute the delays for individual trains and track links in urban railway network. Yuan et al. (2007) optimize the capacity utilization of stations by using a new stochastic model of train delay propagation. The new analytical model estimates the train delays caused by route conflicts and late transfers. The delay discussed in Yuan et al. (2007) is associated with pre-determined timetables, whereas the delay in this paper is referring to simple knock-on delay.

Queueing theory has also been applied to study the train delays. Schwanhäußer (1974) applied queueing theory to develop a smooth model to assess expected waiting times of trains on a rail line. His model considers a minimum headway requirement, train priorities, and the level of initial delays. Greenberg et al. (1988) use queueing models for predicting dispatching delays on single-track, low speed rail segments. The arrivals of trains are assumed to be a Poisson Process. In their work, the sidings are assumed to have infinite capacity. Huisman et al. (2001) study the delay of fast trains being caught behind slower ones. The running times of trains are obtained by solving a system of linear differential equations. The queueing model belongs to the class of a $G/G/\infty$ resequencing queue. Huisman et al. (2002) propose a solvable queueing network model to estimate the mean delays at the network level using closed form expressions.

Besides queueing theory, simulation models have also been used to determine delay time for more realistic rail operations. Dessouky et al. (1995) built a simulation model to analyze the delay of trains in a complex rail network. Their model considers the track speed limits, headways, train lengths and acceleration and deceleration rates of trains. This work is extended by Lu et al. (2004). Hooghiemstra et al. (1998) build

a simulation model for planners to study network-wide dynamic properties under different timetables. Yalçınkaya and Bayhan (2009) use a simulation modeling and response surface methodology approach to determine the optimal settings such as headway to minimize passenger waiting times. Medanic and Dorfman (2002) propose a discrete event model using a travel-advance strategy (TAS) to schedule trains on a single line. This approach is expanded in Dorfman and Medanic (2004) for general networks and incorporated in a simulation model of an actual real system. Li et al. (2008) present an alternative simulation model that implements an algorithm that makes use of global information of the system state to schedule trains. The simulation models can be at different scales and closely imitate real situations. However, compared to analytical models, simulation models generally lack insight on input/output relations. Simulation methods also require more modeling effort and higher computation times.

In this paper, we propose a dispatching policy over double-track railway segments that are used by both fast and slow trains. The new dispatching rule enables the fast train to pass the slow train by using the track traveled by trains in the opposite direction if the track is empty. Thus our dispatching rule is more flexible than the existing dedicated policy. Then we use queueing theory techniques to calculate the delay functions of this policy. We analytically compare our dispatching policy with a dedicated policy. Though there are some research on the topic of automatic re-routing/dispatching of trains of different speed-profiles (see Kuckelberg 2011 and Weymann 2011), to our best knowledge, no policies similar to ours have been studied in the literature nor the delay functions of such a policy.

The rest of the paper is organized as follows. Section 2 introduces the dispatching problem and two dispatching polices. Section 3 presents the delay functions of the two dispatching policies. Several numerical experiments are discussed in Section 4 and Section 5 expands the findings for double-track segments with crossovers. Finally, Section 6 summarizes the findings of this paper.

2. Problem statement and dispatching policies

Without loss of generality, consider a double track east-west segment of a railroad line between two major intersections as shown in Figure 1. Trains arrive at either end of the track segment according to a Poisson process. There are two types of trains: fast trains (e.g., passenger trains) and slow trains (e.g., freight trains). The arrival of fast trains and slow trains are independent Poisson processes. The Poisson process, instead of fixed timetables, is used to model train arrivals for four reasons: (1) There is evidence in the literature showing that in highly congested railway corridors, the arrivals and departures at intermediate stations may be approximated by a Poisson process (Goverde, 2005 and Yuan, 2006); (2) Before new railway infrastructures become fully available, the eventual timetables might be unknown. Thus a timetable with Poisson arrivals can be used to represent various possible future timetables; (3) Train services that are highly responsive to passenger demand may have Poisson arrivals; (4) Poisson-process is particularly valid if there is a significant share of freight trains which do not run according to pre-determined timetables.

We define the following variables:

$\lambda_{E,f}$:   The arrival rate of eastbound fast trains.

$\lambda_{E,s}$:   The arrival rate of eastbound slow trains.

$\lambda_{W,f}$:   The arrival rate of westbound fast trains.

$\lambda_{W,s}$:   The arrival rate of westbound slow trains.

λ_{E} :   The arrival rate of eastbound trains, $\lambda_E = \lambda_{E,f} + \lambda_{E,s}$.

λ_{W} :   The arrival rate of westbound trains, $\lambda_W = \lambda_{W,f} + \lambda_{W,s}$.

D: The length of the double track segment.

S_f :   The speed of a fast train.

S_s :   The speed of a slow train.

T_f :   Minimum running time of fast trains on track segment, $T_f = D/S_f$.

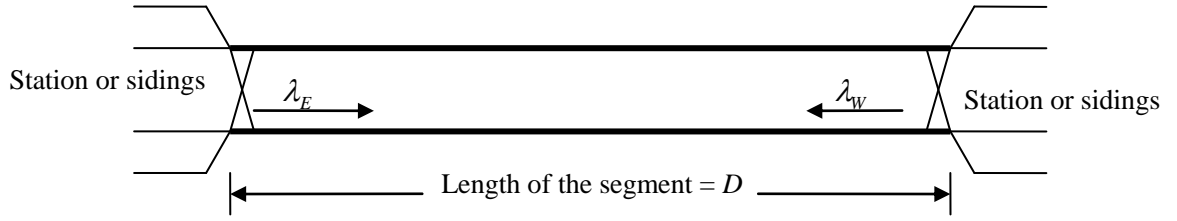T_s :   Minimum running time of slow trains on track segment, $T_s = D / S_s$.

Figure 1. Double-track railroad segment

The upper and lower tracks of the segment are bi-directional. Thus any eastbound and westbound trains can travel on either track segment as long as no deadlock is caused. In this research, the length of the train and the headway between consecutive trains are not considered (e.g. train length and headway are both assumed to be zero). The effect of train length and headway on total train delay traveling in the segment would be negligible if the length of the track segment is much longer than the length of the train. Another main reason for making such assumptions is to keep the analytical model tractable and the goal of this paper is to develop an analytical model to compare the difference between the existing dedicated policy and the proposed new policy. Clearly, ignoring these assumptions underestimates the absolute magnitude of the delay (e.g., if minimum headway is considered, a train might be held back at the station until the minimum distance to the proceeding train is satisfied, thus the knock-on delay of trains will increase), but we are interested in the relative comparison of delay for the two policies. Mu and Dessouky (2011) conducted simulation studies where train length and headway are considered. The results show that, with the consideration of train length and headway, the proposed new policy dominates the existing dedicated policy by the same magnitude as we show by the analytical model in this paper. We also assume infinite waiting areas at both ends of the track segments and the switching of trains from one track to the other takes a negligible amount of time.

The dispatching policy on this double-track segment controls the movement of trains for both directions. Different dispatching rules will result in different delays for

fast and slow trains. The delay time is calculated as:

$$\text{Delay} = \text{Completion time} - \text{Arrival time at segment} - \text{Minimum running time}$$

The minimum running time is defined to be the fastest traveling time of the train assuming there are no other trains in the network. Delay can occur when either a train arrives at the segment and it has to wait for the track to be cleared from being occupied by trains traveling in the opposite direction or a fast train catches up with a slow train traveling in the same direction so that the fast train has to travel at the speed of the slow train. Next we are going to describe two dispatching policies whose delay functions are derived in Section 3.

The easiest dispatching policy for this trackage configuration is to dedicate one track to one direction. Without loss of generality, let the upper track be dedicated to trains traveling westbound and the lower track be dedicated to trains traveling eastbound. Under this dispatching rule, all trains will be able to start traveling on the track as soon as they arrive at the beginning of the segment. Thus the slow trains will not be delayed and the delays of the fast trains are solely caused by reaching and following a slow train in the same direction.

The major drawback of the dedicated policy is that if the arrival rates are high then the fast trains are likely to be blocked by slow trains. This type of delay will dominate even further if the speed difference between the fast and slow trains is relatively large. Here we propose a more flexible dispatching rule, called switchable policy, which is designed to reduce the likelihood of a fast train being blocked by a slow train. The switchable policy will allow a track segment to be used by trains travelling in two directions. Though similar "flying overtaking" is practiced in Switzerland and Germany, in most railway systems, allowing a track to be traveled in both directions may be considered risky (though interlocking mechanism assures no additional risks arise) and thus is not frequently used. However, as the traffic demand increases, some busy railway segments will experience severe congestions and delays. For a busy railway corridor where building new infrastructure is almost impossible or extremely expensive, incorporating new dispatching rules like the switchable policy

may be the best way to increase the throughput capacity. The basic idea of the switchable policy is to allow the fast train to use the other track while some slow trains in the same direction are occupying the dedicated track. However, occupying the dedicated track for the reverse direction would block all trains from travelling in the reverse direction. Thus we need a rule to determine when the fast train is allowed to switch to the other track. Again, let the upper track be the designated track for westbound trains and let the lower track be the designated track for eastbound trains. Then the upper (lower) track would be the track with reverse direction for the eastbound (westbound) trains.

Switchable policy

1. All slow trains travel on their designated track. Upon arrival, if the designated track is not occupied by trains travelling in the opposite direction, the slow train starts travelling on the designated track immediately. If the designated track is occupied by trains travelling in the opposite direction, the slow train waits for the opposing moving trains to finish traveling on the segment before proceeding.

2. When a fast train arrives, if no slow train has arrived within $(T_s - T_f)\sigma$ time units from now, the fast train starts travelling on the designated track if there are no opposing moving trains on the track segment. If there are opposing moving trains travelling on the designated track, the fast train waits for the opposing moving trains to finish travelling on the segment before proceeding.

3. When a fast train arrives, if a slow train has arrived within $(T_s - T_f)\sigma$ time units from now, then the fast train will try to use the track designated for the reverse direction. The fast train will use the reverse direction track only if it is empty. If the reverse direction track is occupied by trains travelling in either direction, the fast train will use the designated track.
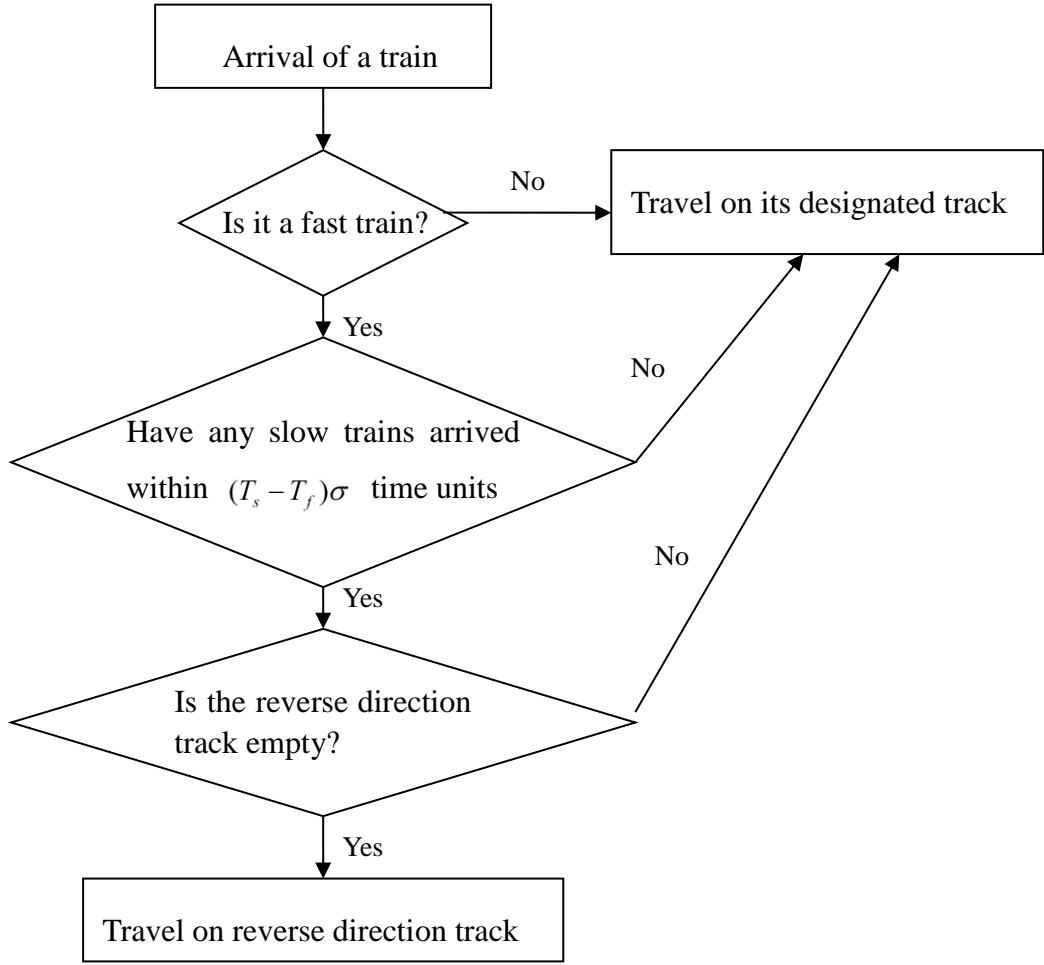
Figure 2. Switchable policy

The switchable policy (also illustrated by the flowchart in Figure 2) enables the fast train to switch to the reverse direction track if two conditions are satisfied: 1. the fast train will be blocked by a slow train in the same direction if it travels on the designated track; 2. the reverse direction track is empty. When condition 1 is satisfied, the fast train will try to switch to the reverse direction track and only when condition 2 is satisfied, the fast train will finally switch. We use the time interval between the arrival of the fast train and the nearest slow train ahead of it as the indicator on whether or not the fast train will be blocked by a slow train. If the time interval is smaller than $T_s - T_f$, the fast train will catch up with the slow train on the designated track before it reaches its destination. Thus, it will be blocked. However, if the fast

train catches up with the slow train near the end of the designated track, the delay will not be significant and in this case, it is probably better to keep the fast train on the designated track. So we define a parameter σ, which is between 0 and 1, to be multiplied to $T_s - T_f$ to be the threshold interval. When σ is set to 1, it means that as long as the fast train can catch up with the slow train on the designated track at some point, the fast train can switch; when σ is set to be 0, it means the fast train will not be allowed to switch even if it arrives at the segment just behind a slow train. Increasing σ from 0 to 1 moves the threshold catching up point from the start point to the end point on the designated track.

We note that a bundling policy of switching consecutive fast trains to the other track can further reduce the train delay, but this is not considered in the analytical model since it would make it intractable. Mu (2011) extended the switchable policy in this paper to a more complicated policy which considers the bundling of trains. Simulation results show the effectiveness of the bundling policy in further reducing the train delay.

3. Delay functions

Before we show our derivations of the delay functions of the switchable policy, we first review the delay functions of the dedicated policy derived by Ren et al. (1996) and Huisman et al. (2001).

3.1. Dedicated policy

Let $F_s(t)$ and $F_f(t)$ be the cumulative density functions (CDF) for the minimum running times of the slow and fast trains. In our case, the minimum running times are constant thus we have:

$$F_s(t) = \begin{cases} 0 & t < T_s \\ 1 & t \geq T_s \end{cases}$$

$$F_f(t) = \begin{cases} 0 & t < T_f \\ 1 & t \geq T_f \end{cases}$$

Let $R_f(t)$ be the CDF of the running time of the fast train. The travel time on the track segment is defined to be:

Travel time = Completion time from the segment − Arrival time at the segment

The slow trains have a delay of zero under this dispatching policy (e.g. slow trains run at a deterministic running time equal to their minimum running time.). The running time of fast trains has a CDF which is derived in Ren et al. (1996) and Huisman et al. (2001) as:

$$R_f(t) = F_f(t)\exp(-\lambda\int_t^\infty (1 - (p_f F_f(x) + p_s F_s(x)))dx)$$

Here $\lambda = \lambda_E$ if the fast train is traveling eastwards; p_f and p_s denote the proportions of fast and slow trains, respectively,

$$p_f = \lambda_{E,f}/\lambda_E \quad \text{and} \quad p_s = \lambda_{E,s}/\lambda_E$$

The expected delay of fast trains is calculated as:

$$Delay_f = \int_0^\infty (1 - R_f(t))dt - T_f$$

The derivation of expected delay of fast trains is rather straight forward for the dedicated policy. We do not need to look into the busy period to calculate the expected delay of the fast trains.

3.2. Switchable policy

To derive the train delay under the switchable policy, the first step is to assess the probabilities that an arriving train sees the various states of the two-track system. Then conditioning on the states of the two-track system when the train arrives, the delay function is derived for that specific condition. The expected delay of trains is finally obtained by considering the delay under all the conditions.

3.2.1. Proportions of time of the various states

Without loss of generality, we assume that the westbound and eastbound arrival rates are the same, so are the fast and slow train proportions. That is,

$$\lambda = \lambda_E = \lambda_W, \quad \lambda_f = \lambda_{E,f} = \lambda_{W,f} \quad \text{and} \quad \lambda_s = \lambda_{E,s} = \lambda_{W,s}$$

For the switchable policy, trains traveling eastbound on the lower track (upper track) are said to be traveling in the designated (reverse) direction and trains traveling westbound on the lower track are said to be traveling in the reverse (designated) direction. Let $P_{l,E}$ ($P_{u,E}$), $P_{l,W}$ ($P_{u,W}$), and $P_{l,0}$ ($P_{u,0}$) be the proportion of time that the lower (upper) track is occupied by eastbound train, westbound trains and by no trains, respectively. Since the eastbound and westbound arrival rates are the same and the dispatching policy is symmetrical for the eastbound and westbound trains, in the long term equilibrium, we have the following:

$$P_{l,E} = P_{u,W}, \quad P_{l,W} = P_{u,E} \quad \text{and} \quad P_{l,0} = P_{u,0}$$

Let P_D denote the proportion of time that a track is occupied by trains traveling in the designated direction, P_R denote the proportion of time that a track is occupied by trains traveling in reverse direction, and P_0 denote the proportion of time that a track is occupied by no trains. The designated directions for the lower and upper track are eastbound and westbound, respectively; the reverse directions for the lower and upper track are westbound and eastbound, respectively. Thus we have the following:

$$P_D = P_{l,E} = P_{u,W}, \quad P_R = P_{l,W} = P_{u,E} \quad \text{and} \quad P_0 = P_{l,0} = P_{u,0}$$

To better illustrate the track status associated with the variables representing the various proportions of time, Figure 3 shows the track states associated with $P_{l,E}$, $P_{l,W}$ and $P_{l,0}$. If the lower track is occupied by trains travelling in a designated track, there are three possible states for the upper track: idle, occupied by trains travelling eastbound and occupied by trains travelling westbound. If the lower track is occupied

by trains in the reverse direction (e.g. a train from upper track switches to the lower track), the upper track has to be occupied by trains travelling in the designated direction of the upper track. Lastly, if the lower track is idle, the upper track can be either idle or occupied by trains travelling in the designated direction of the upper track.

$P_{l,E}$ represents the proportion of time that the tracks are in any of the following three states:
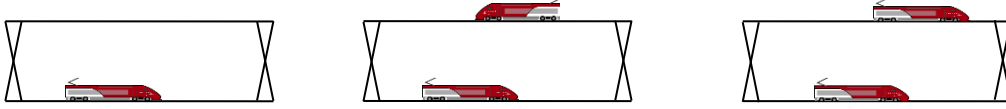
$P_{l,W}$ represents the proportion of time that the tracks are in following state:
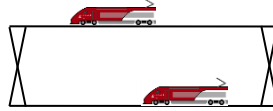
$P_{l,0}$ represents the proportion of time that the tracks are in any of the following two states:

Figure 3. Track states associated with $P_{l,E}$, $P_{l,W}$ and $P_{l,0}$.

Let $\Pr_s$ denote the proportion of fast trains that finally switch track segment (e.g. eastbound trains traveling on the upper track and westbound trains traveling on the lower track). Recall that, when a fast train arrives, it attempts to switch if the nearest slow train arrives less than $(T_s - T_f)\sigma$ time units ago. Thus, the probability a fast train attempts to switch is equal to the probability that at least one slow train arrived during the previous time interval of length $(T_s - T_f)\sigma$. Thus the probability a fast train attempts to switch is:

$$\Pr(\text{attempts to switch}) = 1 - \exp(-\lambda_s (T_s - T_f)\sigma) \qquad (1)$$

Then the probability a fast train finally switches is:

Pr_S = Pr (attempts to switch) * Pr (the reverse track is empty | attempts to switch) (2)

According to the PASTA (Poisson Arrivals See Time Averages) Theorem (Wolff, 1982), P_D is the probability a new arriving train sees its designated track occupied by trains traveling in the same direction, P_R is the probability a new arriving train sees its designated track occupied by trains traveling in reverse direction, and P_0 is the probability a new arriving train sees its designated track empty. Also, independently, P_D is the probability a new arriving train sees the reverse direction track occupied by trains traveling in the reverse direction, P_R is the probability a new arriving train sees the reverse direction track occupied by trains traveling in the same direction as itself, and P_0 is the probability a new arriving train sees the reverse direction track empty. When a fast train arrives, if given no information on the status of the designated track, the fast train sees the reverse direction track empty with probability P_0. Since the status of both the upper and lower tracks are not independent (e.g., given the upper track is empty, the lower track cannot be occupied by a train traveling westbound), given the information that some slow trains arrived not long before the fast train, the probability that this fast train sees the reverse direction track empty is not exactly P_0. However, in our analysis, we approximate this dependent probability by P_0. Later in the numerical experiments, we will see that this approximation turns out to be relatively accurate. So we have:

$$\text{Pr}_S \approx (1 - \exp(-\lambda_s (T_s - T_f)\sigma))P_0 \qquad (3)$$

Let B_D be the length of a busy period, starting from the idle state, for a track segment occupied by trains in the designated direction and B_D^f and B_D^s be the length of this busy period when initiated by a fast and slow train, respectively. Then, conditioning on what type of train (fast or slow) first initiated this busy period, the

expected length of a busy period of a track segment occupied by train in the designated direction is:

$$E[B_D] = (\lambda_s / (\lambda_s + \lambda_f)) E[B_D^s] + (\lambda_f / (\lambda_s + \lambda_f)) E[B_D^f] \tag{4}$$

To derive the expression for $E[B_D^f]$, we condition on when the second train of the busy period arrives and its type (fast or slow). If the second train arrives within T_f time units, the busy period regenerates; else if the second train arrives after T_f time units, this busy period will have a length of T_f.

$$
\begin{aligned}
E[B_D^f] = & \int_0^{T_f} (E[B_D^f] + t) \frac{\lambda_f}{\lambda_s + \lambda_f} (\lambda_s + \lambda_f) \exp(-(\lambda_s + \lambda_f)t) dt \\
& + \int_0^{T_f} (E[B_D^s] + t) \frac{\lambda_s}{\lambda_s + \lambda_f} (\lambda_s + \lambda_f) \exp(-(\lambda_s + \lambda_f)t) dt \\
& + T_f \exp(-(\lambda_s + \lambda_f)T_f)
\end{aligned}
\tag{5}
$$

We note in reality T_f may be different since we make several assumptions such as no bundling. Similarly, to get the expressions for $E[B_D^s]$, we condition on the details of the second train of the busy period. However, since the leading train is a slow train, if the second train is a fast train and arrives no later than $T_s - T_f$ time units, it will catch up with the slow train and then travel at the speed of the slow train. In this case, the arrival of this fast train does not extend the length of this busy period and thus can be ignored. Also note if the second train is a fast train that arrives no later than $(T_s - T_f)\sigma$ time units, it might switch to the reverse direction track and thus has no contribution to the busy period on the designated track. Hence, any fast train that arrives no later than $T_s - T_f$ time units after the slow train are excluded in the expression of $E[B_D^s]$ since they have no impact on the length of the busy period. So we have:

$$E[B_D^s] = \int_0^{T_s - T_f} (E[B_D^s] + t)\lambda_s \exp(-\lambda_s t)dt$$

$$+ \exp(-\lambda_s(T_s - T_f))[\int_0^{T_f} (E[B_D^s] + t + T_s - T_f)\frac{\lambda_s}{\lambda_s + \lambda_f}(\lambda_s + \lambda_f)\exp(-(\lambda_s + \lambda_f)t)dt$$

$$+ \int_0^{T_f} (E[B_D^f] + t + T_s - T_f)\frac{\lambda_f}{\lambda_s + \lambda_f}(\lambda_s + \lambda_f)\exp(-(\lambda_s + \lambda_f)t)dt$$

$$+ T_s \exp(-(\lambda_s + \lambda_f)T_f)] \tag{6}$$

The busy period of a track segment occupied by a train in the designated direction may begin either when a train arrives and finds its designated track idle or when the busy period of a track segment occupied by a train not in its designated direction finishes and at least one train traveling in the designated direction is waiting to use the track. Let $f_{D,i}$ be the rate that the busy period of a track segment with a train in the designated direction begins with the track being idle. And let $f_{D,r}$ be the rate that the busy period of a track segment with a train in the designated direction begins with the ending of reverse direction busy period. We have:

$$f_{D,i} = (\lambda_s + \lambda_f)P_0 \tag{7}$$

$$f_{D,r} = g_D f_R \tag{8}$$

g_D is defined to be the probability that when the reverse direction busy period ends, at least one train traveling in the designated direction of the track is waiting at the end of the segment. And f_R denotes the rate a busy period starts with a train in reverse direction (not in the direction of the designation of the track):

$$f_R = \lambda_f \Pr_S \approx \lambda_f(1 - \exp(-\lambda_s(T_s - T_f)\sigma))P_0 \tag{9}$$

According to the switchable dispatching rule, the busy period of a reverse direction is a constant which is equal to T_f. Thus P_R can be expressed as:

$$P_R = f_R T_f \tag{10}$$

Similarly, we can write:

$$P_D = f_{D,i}E[B_D] + f_{D,r}E[B_D^r] \tag{11}$$

16

$E[B_D^r]$ is defined to denote the expected length of a busy period with trains traveling in the designated direction, which begins after a fast train traveling in the reverse direction gets off the track. Among the trains waiting to travel on the designated track, if there are any slow trains, then the expected length of this busy period is equal to $E[B_D^s]$. The probability there are slow trains in the queue is $1 - \exp(-\lambda_s T_f)$. If all trains in the waiting queue are fast trains, B_D^r has an expected length of $E[B_D^f]$. The probability all trains waiting in the queue are fast trains is $(1 - \exp(-\lambda_f T_f)) \exp(-\lambda_s T_f)$. Now the expression for P_D becomes:

$$
\begin{aligned}
P_D &= (\lambda_s + \lambda_f) P_0 E[B_D] \\
&\quad + (1 - \exp(-\lambda_s T_f)) f_R E[B_D^s] \\
&\quad + (\exp(-\lambda_s T_f) - \exp(-(\lambda_s + \lambda_f) T_f)) f_R E[B_D^f]
\end{aligned}
\tag{12}
$$

In addition, we have the following fact:

$$
P_D + P_R + P_0 = 1
\tag{13}
$$

Now we have three unknown variables P_D, P_R and P_0, together with the three equations (9), (12) and (13), we are able to solve for the values of P_D, P_R and P_0. Next, we are going to express the delays of the trains in terms of these three values.

3.2.2. Delay of slow trains

The delay for slow trains is easy to calculate. The delay of slow trains will only occur when a slow train arrives at the segment and its designated track is occupied by a train traveling in the reverse direction. In this case, the delay of the slow trains will be the remaining length of the reverse direction busy period (denoted by B_R^e). It is well known that the remaining busy period found by Poisson arrivals has the following equilibrium distribution (Greenberg at el. 1988):

$$
\Pr(B_R^e \leq t) = \frac{\int_0^t \Pr(B_R > x) dx}{E[B_R]}
\tag{14}
$$

B_R is a constant with a value of T_f. Thus according to equation (14), B_R^e has a value uniformly distributed between 0 and T_f. The expected length of the remaining busy period is:

$$E[B_R^e] = T_f / 2$$

Now we condition on whether or not a slow train will be delayed, where U_s represents delay for a slow train:

$$\begin{aligned} E[U_s] &= (1 - P_R) \cdot 0 + P_R \cdot E[B_R^e] \\ &= P_R T_f / 2 \end{aligned} \tag{15}$$

3.2.3. Delay of fast trains

To obtain the delay for fast trains, we condition on the status of the designed track as the fast train arrives. Let U_f be the delay for a fast train:

$$\begin{aligned} E[U_f] = {}&P_0 \cdot 0 \\ &+ P_R \cdot E[U_f \mid \text{held for reverse train to finish traveling}] \\ &+ P_D \cdot E[U_f \mid \text{designated track is occupied by trains in the same direction}] \end{aligned} \tag{16}$$

Without loss of generality, in the following analysis, we assume $2T_f \leq T_s$. For the case where $2T_f > T_s$, the same logic of deriving the delay function applies (the difference will be that the drawing in Figure 6 will change, since the length of T_s does not comprise two T_f's anymore). Recall that the threshold value σ controls how frequently the fast train will switch, setting σ to 1 means the switchable policy will switch fast trains as long as they will be delayed, for even a small amount of time, on their designated track. Without loss of generality, in the following derivations, σ is set to 1. For other values of σ, the same derivation logic also applies.

If the designated track for a fast train is occupied by a train in the reverse direction, the fast train cannot switch to the other track, since the reverse direction

track has to be occupied by at least one slow train. Thus the fast train has to wait for the reverse train to finish traveling and the delay of this fast train may have two components. The first component is the remaining length of the reverse direction busy period, which has a uniform distribution from 0 to T_f because the fast train arrives according to a Poisson process. So when the fast train arrives, the other fast train, which is travelling in the opposite direction, will equally likely to be at any position on the designated track of the newly arrived train. Thus the remaining length of the busy period has a uniform distribution. The second component is present if there are slow trains in the queue ahead of the fast train waiting. In this case, the fast train has to follow the slow train traveling through the whole segment, resulting in an extra delay of $T_s - T_f$. If the fast train has to wait for t time units, where $0 \leq t \leq T_f$, then we condition on if any slow train arrives during the time interval of length $T_f - t$ before the fast train's arrival time. If change-of-sequence is allowed at the waiting area, the second component can be removed because fast trains can change sequence with slow trains while waiting and fast trains will always stay ahead of slow trains. Thus if change-of-sequence is allowed, the expected delay of fast trains under the switchable policy will be lowered. However, in this study, the change-of-sequence is not considered. We then have:

$E[U_f \mid \text{held for reverse train to finish traveling}]$

$$= \int_0^{T_f} \frac{1}{T_f} \exp(-\lambda_s (T_f - t)) t \, dt + \int_0^{T_f} \frac{1}{T_f} (1 - \exp(-\lambda_s (T_f - t)))(t + T_s - T_f) dt \quad (17)$$

Now the only part missing to derive is the last term in equation (16). That is, we need to compute the expected delay for a fast train when it arrives and sees trains traveling in the same direction on the designated track. The basic idea for deriving this delay is to condition on how many and when the slow trains arrived during a time interval before the arrival of the fast train. In order to illustrate the derivations, let's focus on the situation on the lower track which is the designated track for eastbound trains (shown in Figure 4).
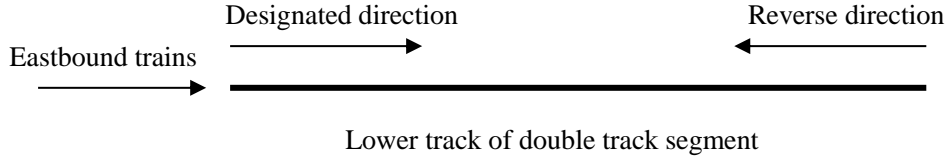
Figure 4. Traveling direction on lower track

When a new eastbound fast train arrives, if the lower track is occupied only by eastbound fast trains, then the newly arrived fast train will not experience any delays before its departure from the station. Therefore we only need to consider the position of any eastbound slow train on the lower track. Note that if at time t a fast train arrives, we only need to focus on the slow trains that arrived between time t and $t - T_s$. Slow trains that arrive before time $t - T_s$ will not have any impact on the delay of the newly arrived fast train, even though the slow train might still be on the track when the fast train arrives. Hence, we condition on how many slow trains that arrived during the T_s time units interval before the arrival of the new fast train. Figure 5 describes the positions of the slow trains that will impact the newly arriving fast train. We start with the case where there is only one slow train that has arrived during this interval; this occurs with probability $\lambda_s T_s \exp(-\lambda_s T_s)$.
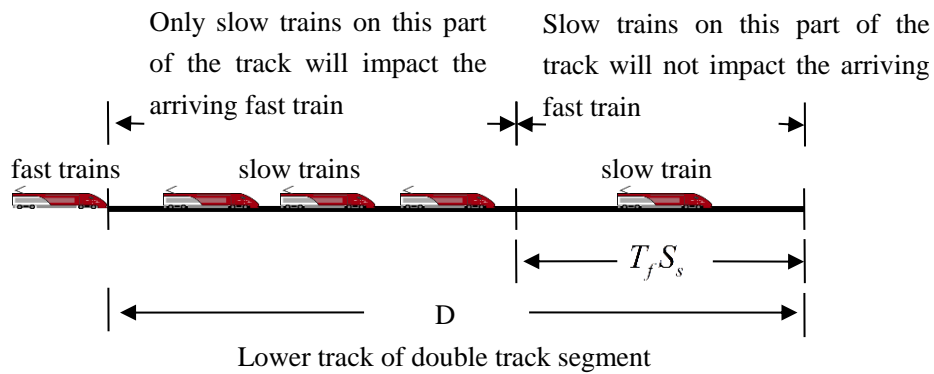
Figure 5. Positions of slow trains that will impact the arriving fast train.

Now consider a time axis from right to left as shown in Figure 6. The numbers on top of the line in Figure 6 are the readings for the time axis. T_s and T_f were introduced as values which equal to the minimum running time of slow and fast trains.

In the following analysis, T_s and T_f will also be used to denote time stamps $0+T_s$ and $0+T_f$. Now suppose a fast train arrives at time instance T_s and if a slow train had previously arrived at the time instance 0 and it was not held back, then the 0 marker on the time axis will be the position on the track of the slow train when the fast train arrives. Similarly, if a fast train arrives at time T_s and a slow train had previously arrived at time T_f and the slow train was not held back, then the T_f marker will be the slow train's position on the track when the fast train arrives; and if the slow train has to wait for 1 time unit, then the slow train will be at position T_f+1.

In Figure 6, the train goes from left to right and the time axis goes from right to left. This is because of the fact that the earlier the train arrives (closer to the 0 point of time axis), the further away it will be on the track (closer to the right end of the track). So by combining the plot of the track and the time axis together into one line, a single point on the line can both represent the location of the train and the time it starts travelling.
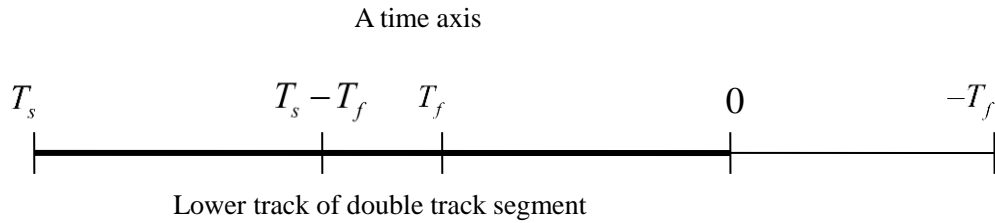
A time axis

T_s T_s-T_f T_f 0 $-T_f$

Lower track of double track segment

Figure 6. A time axis

If there is only one slow train in the interval of length T_s, the arrival time is uniformly distributed along the interval, with a density function of $1/T_s$. Without loss of generality, we assume that the fast train arrives at time T_s, and since we assume one slow train is in the network, the slow train must have arrived in the time interval (0, T_s). We next consider three different components of the delay by dividing the time (0, T_s) into three different cases. Case 1, Case 2 and Case 3 deal with the

situations where the slow train arrives in the intervals $(0, T_f)$, $(T_f, T_s - T_f)$ and $(T_s - T_f, T_s)$, respectively (see Figure 7).
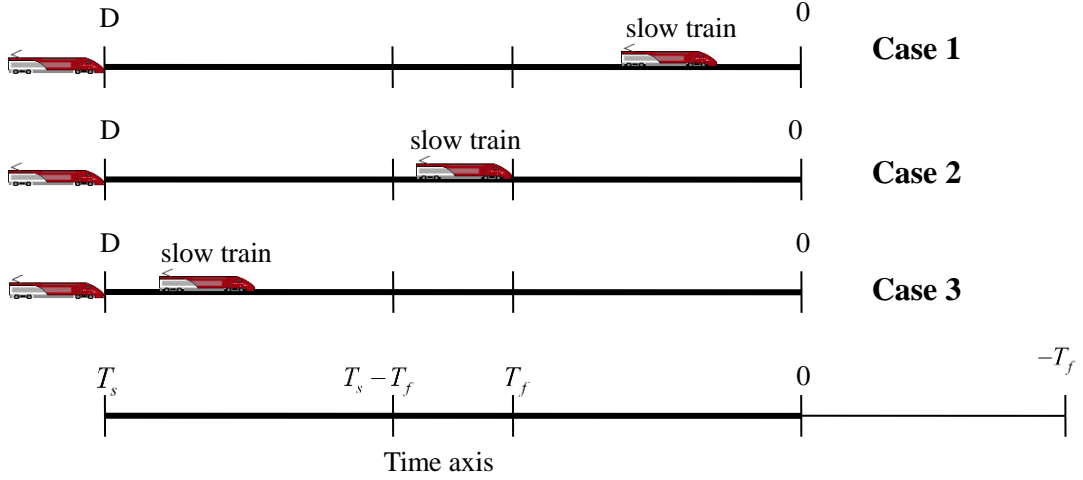

Figure 7. Illustration of Case 1, Case 2 and Case 3

Case 1

Now suppose the slow train arrives in the interval $(0, T_f)$ (see Figure 8). If the slow train does not have to wait for the track upon its arrival, this slow train will not delay the newly arrived fast train. Upon the arrival of the slow train, with probability P_R, the slow train will be held back for a certain amount of time. If the slow train is delayed long enough, it may delay the new fast train (see Figure 8). Note that, if the slow train arrives in the interval $(0, T_f)$ and σ is 1, the newly arrived fast train will not try to switch tracks. By conditioning on the time the slow train arrived and the length of time it was held back, we have the expected delay of the fast train for Case 1 as:

$$E[U_{f,c1}] = \int_0^{T_f} \frac{1}{T_s} P_R \int_0^{T_f} \frac{1}{T_f} \max[0, t + u - T_f] du dt$$

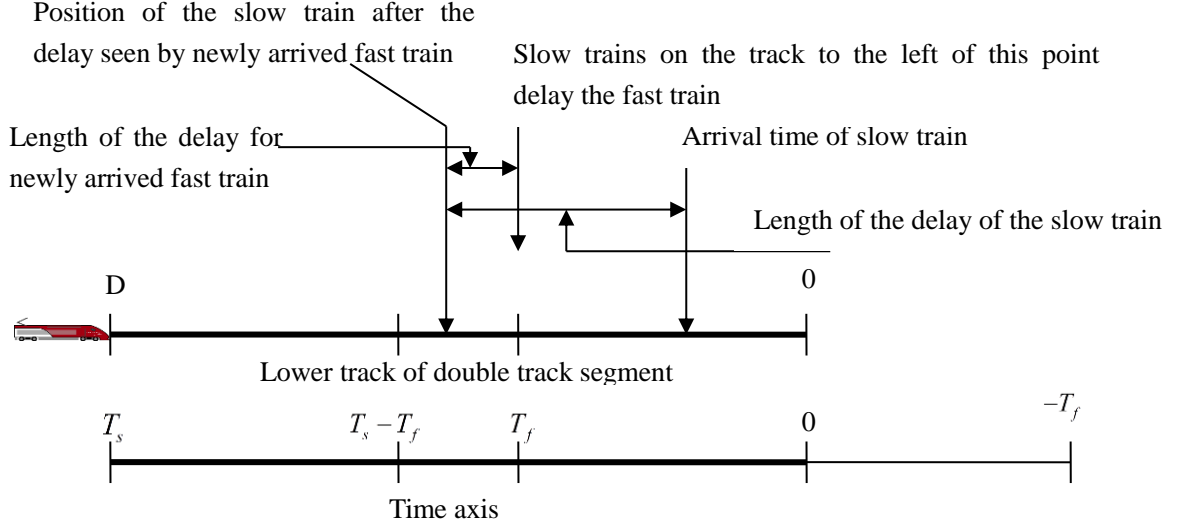$$= \int_0^{T_f} \frac{1}{T_s} P_R \int_{T_f - t}^{T_f} \frac{1}{T_f} (t + u - T_f) du dt$$

(18)

Position of the slow train after the
delay seen by newly arrived fast train

Slow trains on the track to the left of this point
delay the fast train

Length of the delay for
newly arrived fast train

Arrival time of slow train

Length of the delay of the slow train

D

0

Lower track of double track segment

T_s $T_s - T_f$ T_f 0 $-T_f$

Time axis

Figure 8. Case 1: slow train arrives in the interval (0, T_f)

Case 2

Now suppose that the slow train arrives in the interval $(T_f,\ T_s - T_f)$ (see Figure 9). For this case, since σ is equal to 1 and the length of the time interval between the arrivals of the fast train and the slow train is less than $T_s - T_f$, the fast train may switch tracks with probability P_0. If the fast train successfully switches, it will experience zero delay. If the fast train does not switch, it will experience a delay as shown in Figure 9, depending on whether or not the earlier arriving slow train experienced a delay. Again, by conditioning on the arrival time of the slow train and the delay of the slow train and multiplying the whole expression by $1 - P_0$, we have the expected delay for the newly arrived fast train for Case 2:

$$
\begin{aligned}
E[U_{f,c2}] = & \int_{T_f}^{T_s - T_f} \frac{1}{T_s}(1 - P_R)(1 - P_0)(t - T_f)\,dt \\
& + \int_{T_f}^{T_s - T_f} \frac{1}{T_s} P_R(1 - P_0) \int_0^{T_f} \frac{1}{T_f}(t + u - T_f)\,du\,dt
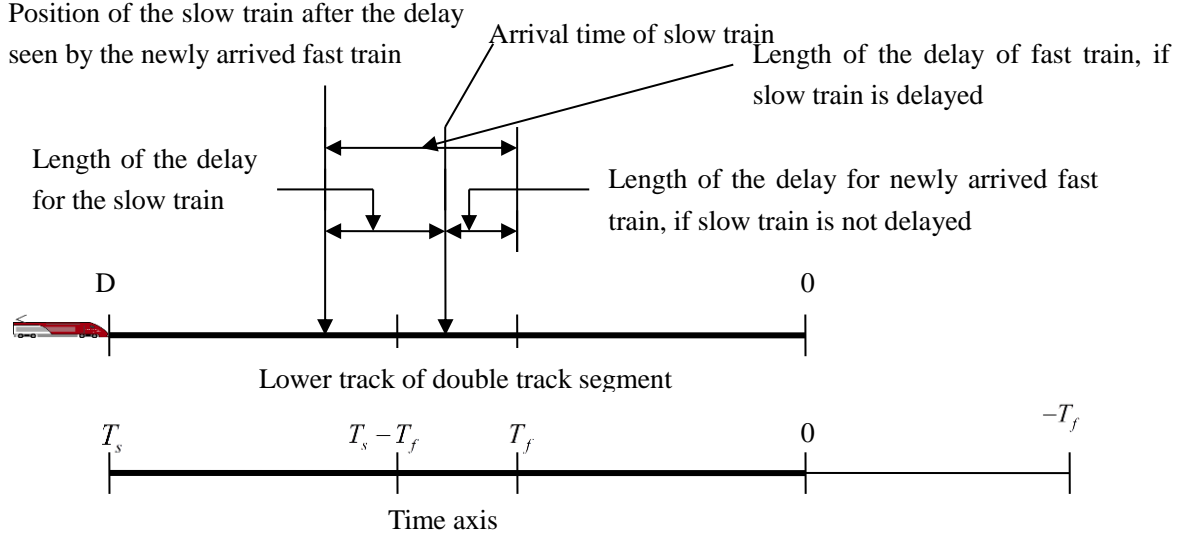\end{aligned}
\tag{19}
$$

Position of the slow train after the delay
seen by the newly arrived fast train

Arrival time of slow train

Length of the delay of fast train, if
slow train is delayed

Length of the delay
for the slow train

Length of the delay for newly arrived fast
train, if slow train is not delayed

D

0

Lower track of double track segment

T_s $T_s - T_f$ T_f 0 $-T_f$

Time axis

Figure 9. Case 2: slow train arrives in the interval $(T_f,\ T_s - T_f)$

Case 3

If the slow train arrived in the interval $(T_s - T_f,\ T_s)$, the delay for the newly arrived fast train will be similar to Case 2 (see Figure 10). The only point to note is that the possible length of delay for the slow train cannot be as long as T_f anymore. If the slow train is delayed for too long, as the fast train arrives, it will see the slow train being held back instead of moving along the track. And this delay will be captured in equation (17). Thus the maximum length for the delay for the slow train is limited by the difference of T_s and the slow train arrival time. We have:

$$E[U_{f,c3}] = \int_{T_s - T_f}^{T_s} \frac{1}{T_s}(1 - P_R)(1 - P_0)(t - T_f)dt$$

$$+ \int_{T_s - T_f}^{T_s} \frac{1}{T_s} P_R (1 - P_0) \int_0^{T_s - t} \frac{1}{T_f}(t + u - T_f)dudt \qquad (20)$$

24
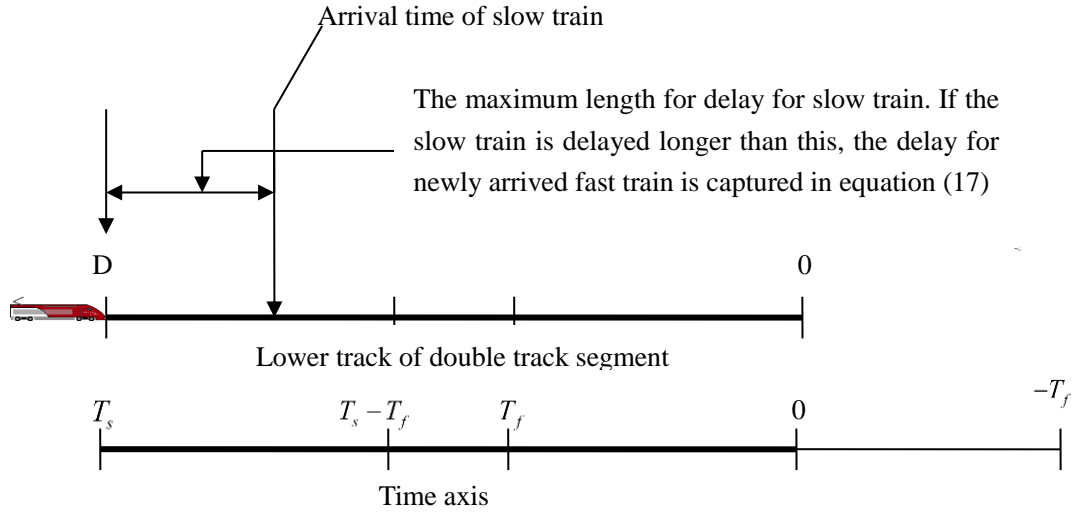
Figure 10. Case 3: slow train arrived in the interval $(T_s - T_f,\ T_s)$

Cases 1-3 only consider the scenario that has only one slow train arrival in the interval $(0,\ T_s)$. When there is more than one slow train, it is the position of the latest arriving slow train that determines the delay of the fast train. However, the probability of this latest slow train being delayed is dependent on the positions of the other slow trains on the track. Thus the delay probability is no longer P_R. To see why this is true, suppose the latest arriving slow train is delayed by t time units, then all the other slow trains must have arrived at most $T_f - t$ time units before the latest arriving slow train. If any slow train arrived more than $T_f - t$ time units before the latest arriving slow train, the latest slow train cannot be delayed by t time units. Thus the amount of time that the latest arriving slow train is delayed is not independent with the arrival times of the other slow trains. To remove this dependency, instead of conditioning on if the latest arriving slow train is delayed, we condition on if the earliest arriving slow train is delayed. The earliest arriving slow train in the interval will be delayed with probability P_R and the length of the delay will still have a uniform distribution from 0 to T_f. If the time between the arrivals of the earliest and the latest arriving slow train is larger than the length of the earliest train's delay, then the latest arriving slow

train will not be delayed; otherwise, the latest arriving slow train might be delayed (see Figure 11).
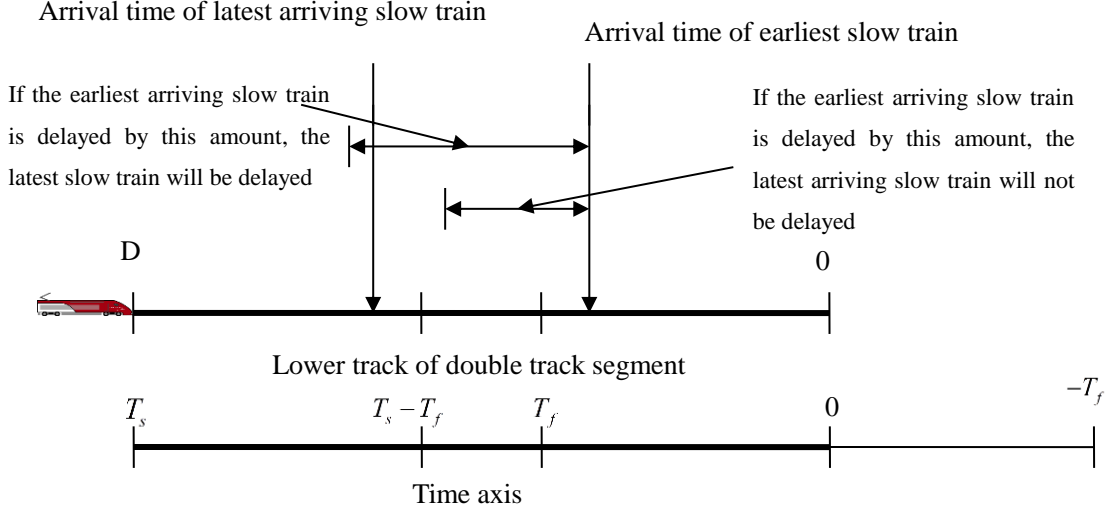

Figure 11. Condition on the latest and earliest slow train

With probability $(\lambda_s T_s)^j / j! \exp(-\lambda_s T_s)$, $j \geq 2$, there are a number of j slow trains in the track segment in the T_s interval before the arrival time of the fast train. And the PDF for the arrival times of the earliest (t_e) and the latest (t_l) slow trains, given there are j slow trains is:

$$f(t_e, t_l) = \frac{j(j-1)}{T_s^2}(\frac{t_l - t_e}{T_s})^{j-2}$$

Similar to the case of one slow train in the interval, we decompose the analysis into three cases according to the value of t_e. Case 1', Case 2' and Case 3' represent the situations where the earliest slow train arrives in the intervals $(0, T_f)$, $(T_f, T_s - T_f)$ and $(T_s - T_f, T_s)$, respectively (see Figure 12).
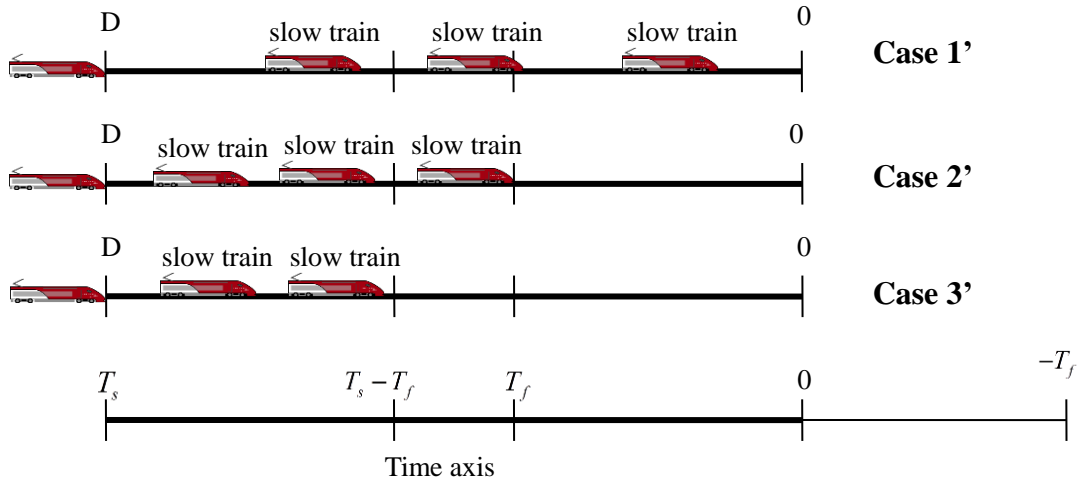
Figure 12. Illustration of Case 1', Case 2' and Case 3'

Case 1'

The earliest slow train arrives in the interval (0, T_f). If the slow train is not delayed, only if it arrives after time T_f will the newly arriving fast train be delayed. If the latest slow train also arrives in the interval (0, T_f), unless the earliest arriving train is delayed long enough, the fast train will not be delayed (see Figure 13).
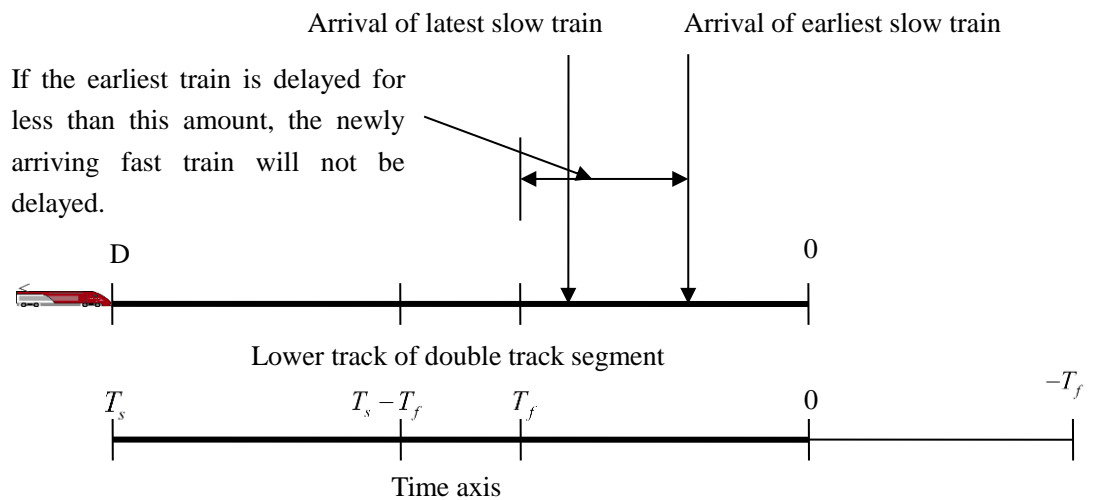

Figure 13. Case 1'

We summarize the results for the expected delay for the newly arriving fast train for case 1' below:

$$E[U_{f,c1'}] = \int_0^{T_f} \int_{T_f}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} (1-P_R)(1-P_0)(t_l - T_f) dt_l dt_e$$

$$+ \int_0^{T_f} \int_{t_e}^{T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R \int_{T_f - t_e}^{T_f} \frac{1}{T_f} (t_e + u - T_f) du dt_l dt_e$$

$$+ \int_0^{T_f} \int_{T_f}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_{t_l - t_e}^{T_f} \frac{1}{T_f} (t_e + u - T_f) du dt_l dt_e$$

$$+ \int_0^{T_f} \int_{T_f}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_0^{t_l - t_e} \frac{1}{T_f} (t_l - T_f) du dt_l dt_e$$

$$+ \int_0^{T_f} \int_{t_e+T_f}^{T_s} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} (1-P_0)(t_l - T_f) dt_l dt_e \qquad (21)$$

Case 2'

If the earliest slow train arrives in the interval $(T_f, \ T_s - T_f)$, we have:

$$E[U_{f,c2'}] = \int_{T_f}^{T_s - T_f} \int_{t_e}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} (1-P_R)(1-P_0)(t_l - T_f) dt_l dt_e$$

$$+ \int_{T_f}^{T_s - T_f} \int_{t_e}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_{t_l - t_e}^{T_f} \frac{1}{T_f} (t_e + u - T_f) du dt_l dt_e$$

$$+ \int_{T_f}^{T_s - T_f} \int_{t_e}^{t_e+T_f} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_0^{t_l - t_e} \frac{1}{T_f} (t_l - T_f) du dt_l dt_e$$

$$+ \int_{T_f}^{T_s - T_f} \int_{t_e+T_f}^{T_s} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} (1-P_0)(t_l - T_f) dt_l dt_e \qquad (22)$$

Case 3'

If the earliest slow train arrives in the interval $(T_s - T_f, \ T_s)$, we have:

$$E[U_{f,c3'}] = \int_{T_s - T_f}^{T_s} \int_{t_e}^{T_s} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} (1-P_R)(1-P_0)(t_l - T_f) dt_l dt_e$$

$$+ \int_{T_s - T_f}^{T_s} \int_{t_e}^{T_s} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_{t_l - t_e}^{T_s - t_e} \frac{1}{T_f} (t_e + u - T_f) du dt_l dt_e$$

$$+ \int_{T_s - T_f}^{T_s} \int_{t_e}^{T_s} \frac{j(j-1)}{T_s^2} (\frac{t_l - t_e}{T_s})^{j-2} P_R (1-P_0) \int_0^{t_l - t_e} \frac{1}{T_f} (t_l - T_f) du dt_l dt_e \qquad (23)$$

Now we have all the components to write:

$P_D \cdot E[U_f \mid$ designated track is occupied by trains in the same direction$]$

$$= \lambda_s T_s \exp(-\lambda_s T_s)[E[U_{f,c1}] + E[U_{f,c2}] + E[U_{f,c3}]] \tag{24}$$

$$+ \sum_{j=2}^{\infty} \frac{(\lambda_s T_s)^j}{j!} \exp(-\lambda_s T_s)[E[U_{f,c1'}] + E[U_{f,c2'}] + E[U_{f,c3'}]]$$

With Equation (17) and (24), the expected delay of the fast train can be found using equation (16).

4. Numerical experiments

In order to test the accuracy of the approximated analytical results for the train delays, we compare our analytical results with simulation results. The simulation results are obtained by running a C++ based Arena simulation model (Kelton et al. 2009) for 500000 hours with 5 replications. Our base settings are:

- Length of track segment, 8 miles
- Speed of a fast train, 140 mph
- Speed of a slow train, 50 mph
- Arrival rate of fast trains, 4.8 trains per hour
- Arrival rate of slow trains, 4.8 trains per hour

We assume trains arrive in two directions at the rates listed above. The threshold constant σ is set to 1. The base case results are shown in Table 1.

Table 1 Comparison between simulation results and analytical results

Results for switchable policy	Simulation results	Analytical results
Value for P_0	.3248	.3223
Value for P_D	.6432	.6432
Value for P_R	.3200	.3450
Mean/expected delay for slow train (minutes)	0.0549	0.0591
Variance of delay for slow train (minutes)	0.00072	-
Mean/expected delay for fast train (minutes)	0.977	0.974
Variance of delay for fast train (minutes)	0.0017	-

From Table 1 we can see that our analytical approximation provides accurate results for the values of P_D, P_R and P_0. The analytical results of delays for slow and fast trains differ from the simulation results by 7.63% and 0.269% respectively. Even though the percentage difference for slow train delays might seem relatively large as compared to the percentage difference for fast train delays, the absolute values for the slow train delays are very small (e.g., 0.0549 and 0.0591 minute). Our analytical model estimates the expected delay for a fast train to be 0.977 minute for a switchable policy. Using the model of Huisman et al. (2001), expected delay for a fast train is 1.307 for a dedicated track policy. For this scenario, the switchable dispatching policy reduces the fast train delays over a dedicated policy by 25.2%. In railway control, the fast trains (passenger trains) usually have higher priority than slow trains (freight trains), since fast trains are normally under tight schedules whereas slow trains have more flexible schedules. Thus the delays of fast trains are more important than the delays for slow trains and when we are comparing the two policies, the delay for fast trains is the major performance measure. The small extra delay for slow trains caused by the switchable policy is not a significant disadvantage. Next we keep all the parameters the same and vary the arrival rate of fast trains.

In Figure 14, we can see that as the arrival rate of the fast train varies, our analytical approximation provides values for the mean delays for fast trains that are close to the simulation results. The percentage error of fast train delay is calculated as:

$$\text{Percentage error} = \frac{|\text{simulated delay - analytical delay}|}{\text{simulated delay}} \times 100\%$$

In Figure 15, as the fast train arrival rate increases, the percentage error of slow train delay shows an increasing trend. However, the absolute values of the slow train delays are so small that the absolute values of the delay discrepancy are not significant. Thus we can conclude that our analytical expressions give accurate and robust approximations to the delays of trains under a switchable dispatching policy.
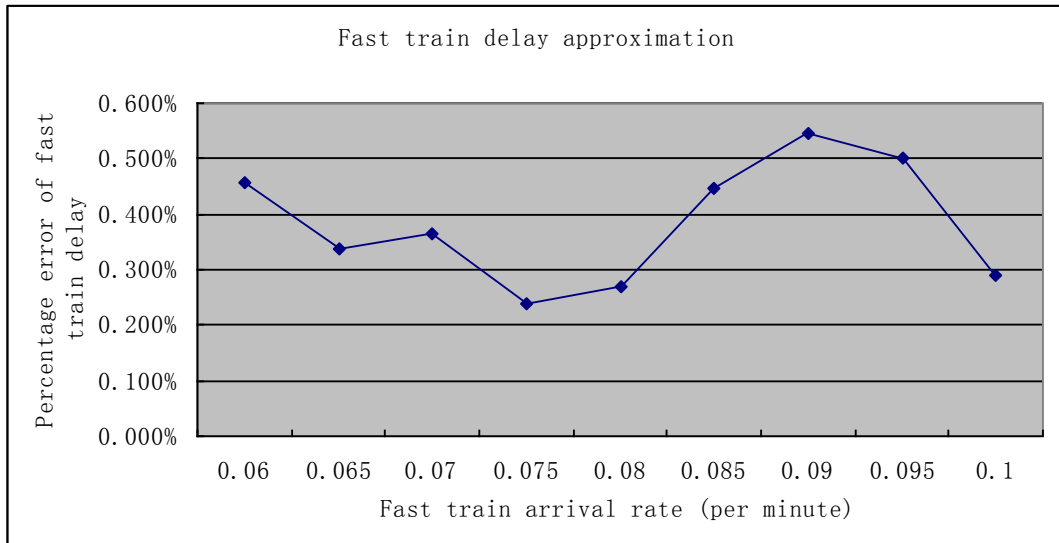
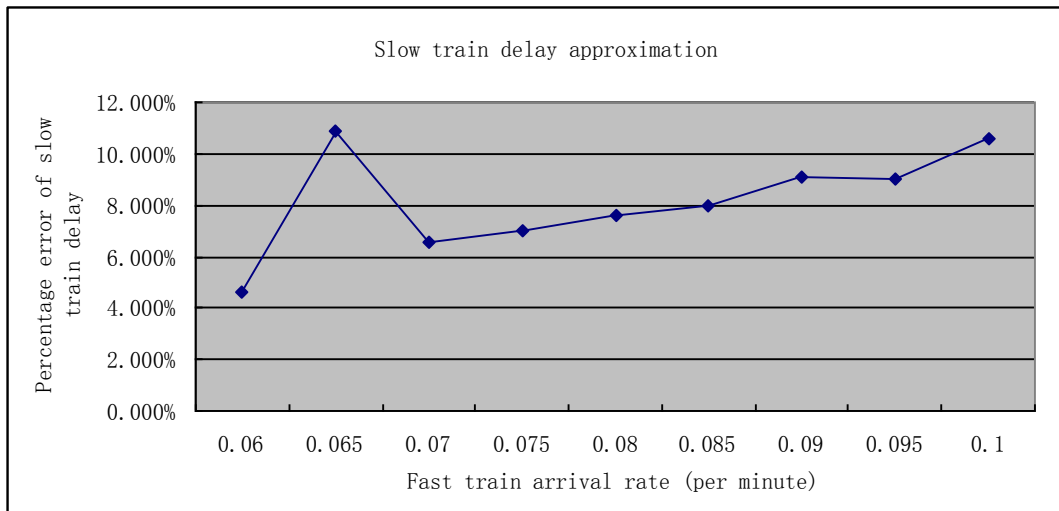**Figure 14. Percentage error for fast train delay**

Figure 15. Percentage error for slow train delay

Figure 16 shows the comparison of the performance measure (fast train delays) of the two dispatching policies under various train arrival rates. From the diagram we can see that the switchable policy has significant advantage over the dedicated policy and this advantage will become more significant as the track becomes less busy (e.g. when the arrival rates of slow and fast train are reduced). This is because as the track becomes less busy, when a fast train attempts to switch tracks, it will have a higher success rate.
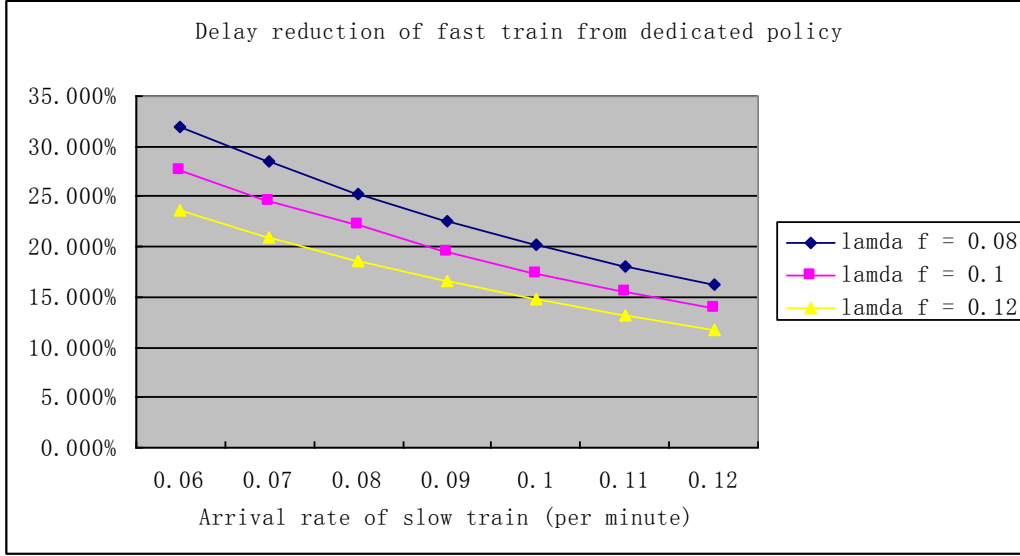
Figure 16. Varying arrival rates

Now keeping all the other parameters at their base settings, we next vary the speed of the slow and fast trains. Figure 17 shows the sensitivity analysis for train delays as a function of train speeds. From the graph, we see that the advantage of the switchable policy becomes larger as the difference in train speeds becomes larger. This is intuitive since the larger the speed gap is, the longer the delay is if a fast train catches up with a slow train; and the switchable policy reduces the chance a fast train catches up with a slow train.

The length of the segment is also an important factor for the switchable policy. The longer the segment is, the less chance that the track is empty. Thus the switching rate becomes smaller as the segment becomes larger. Intuitively, the advantage of the switchable policy will become more significant if the length of the segment is smaller. Figure 18 shows the relationship between the length of the segment and the advantage of the switchable policy. This plot also keeps all the other parameter settings at their base case.
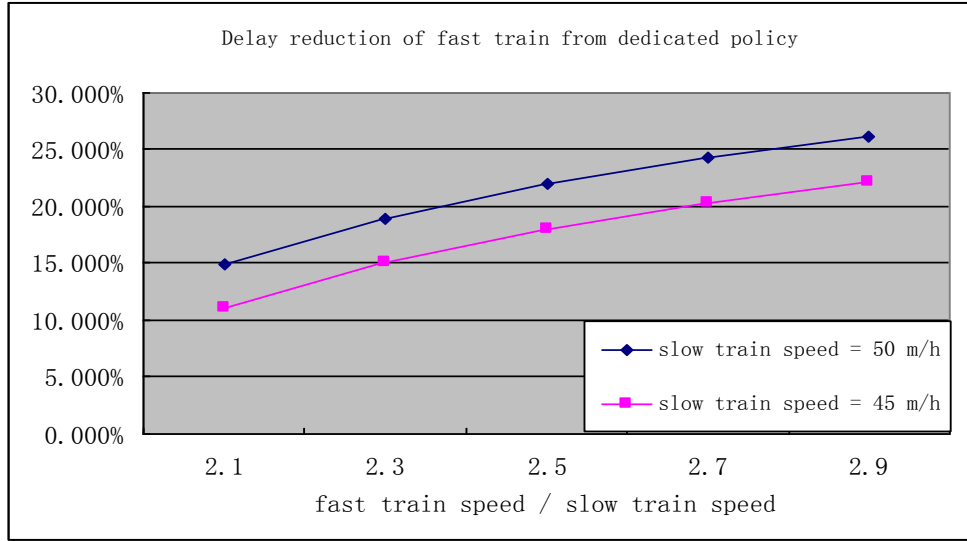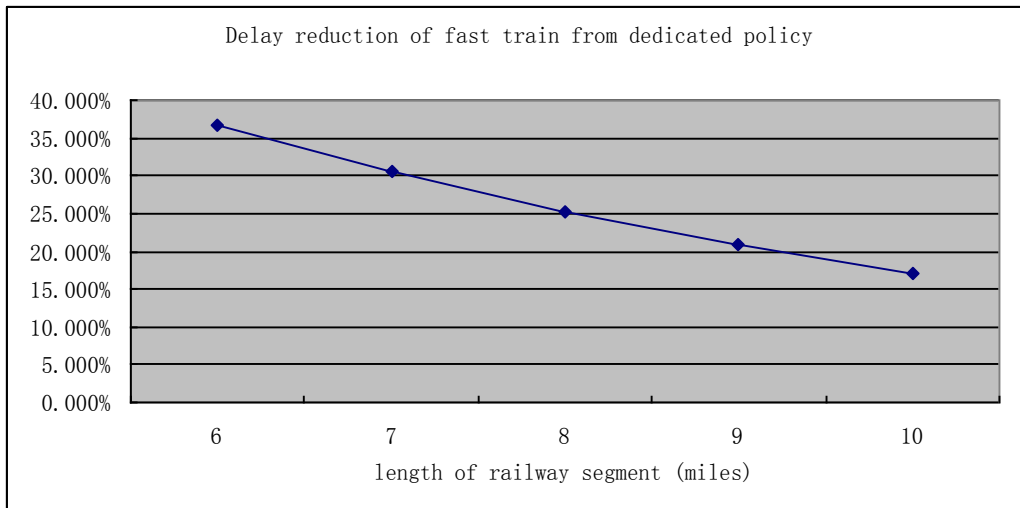
Figure 17. Varying train speeds

Figure 18. Varying segment lengths

So far, all the results for the switchable policy were obtained by setting σ to 1. However, the optimal value of σ is not necessarily 1. The optimal value of σ changes as the parameter values change. Figures 19, 20 and 21 show the effects of varying σ under the base case parameter settings. We can see that the optimal value of σ which produces the largest delay reduction for the fast trains is between 0.8 and 0.9 for this scenario. The delay of slow train constantly drops as σ decreases because the smaller σ is, the less frequent fast trains will switch. The delay of slow train is solely caused by the switched fast trains traveling in reverse direction. Hence

if fast trains switch less frequently, slow trains will experience less delay. Also note, as σ drops to 0, no fast trains will switch which makes the switchable policy the same as the dedicated policy, thus the delay reduction is 0. Figure 21 shows the switchable policy dominates the dedicated policy in terms of sum of delay times.
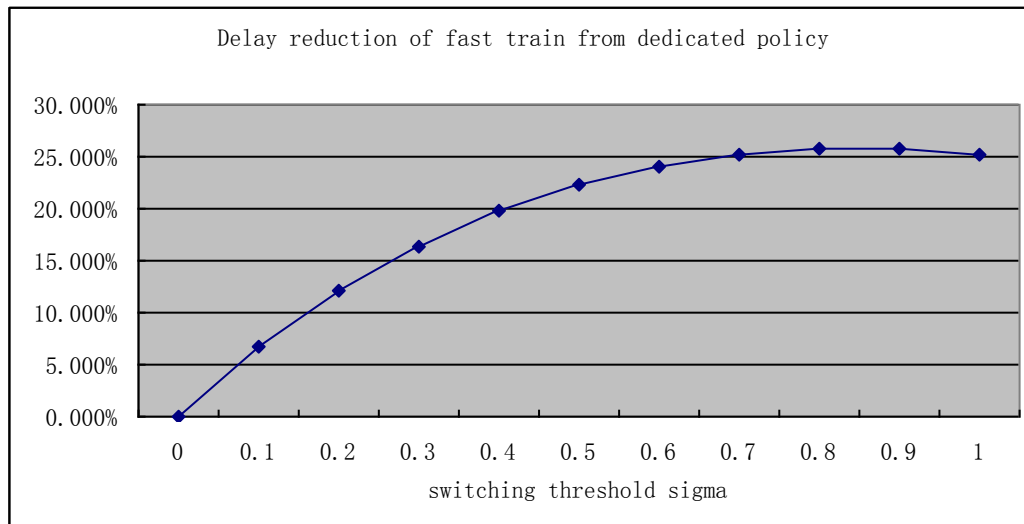

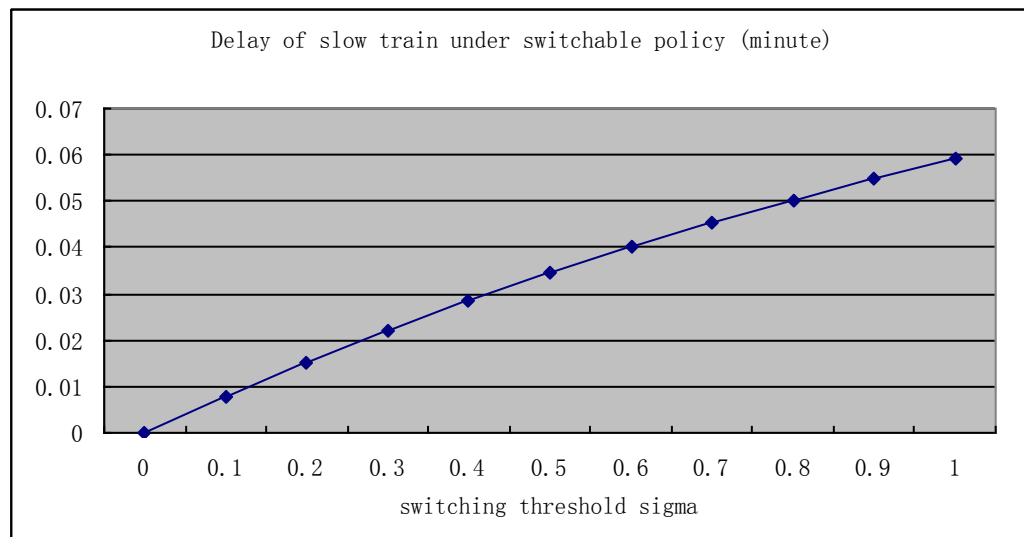**Figure 19. Effect on fast train delay reduction by varying** σ

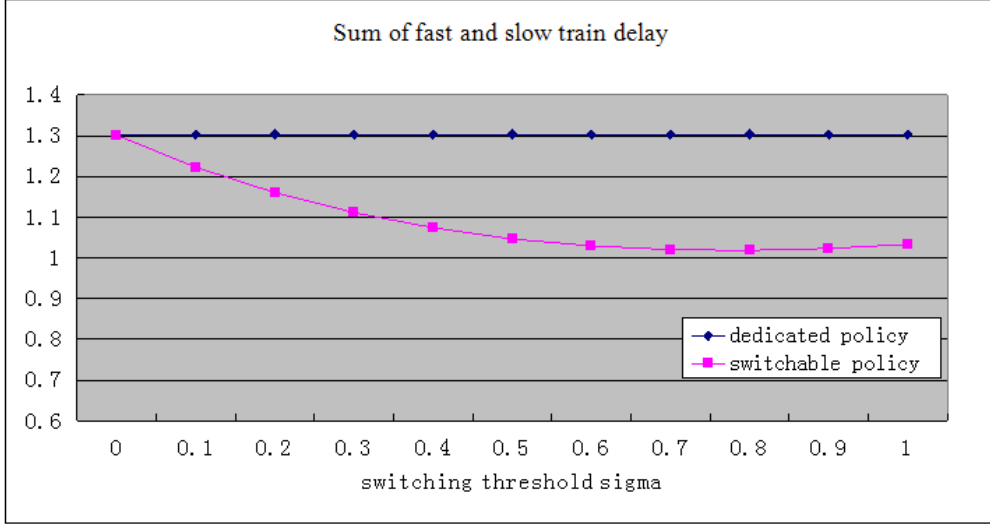**Figure 20. Effect on slow train delay by varying** σ

Figure 21. Effect on sum of fast and slow train delay by varying σ

5. Double-track with crossovers

For the above switchable policy, once a fast train switches the track, the fast train has to travel all the way on the other track to pass the slow trains ahead of it. Suppose there is one crossover linking two tracks in the middle of the segment, then the efficiency of a switchable policy can be further improved due to better utilization of the track.

Figure 22 shows a double-track railway segment with crossovers at the middle of the segment. Without loss of generality, the lower track is the designated track for the eastbound traveling trains and the upper track is the designated track for the westbound traveling trains. We denote the lower (upper) track segment west of the crossover by EB1 (WB1) and the lower track segment east of the crossover by EB2 (WB2). We extend the switchable policy in Section 2 to address the situation where a linking crossover is at the middle of the double-track segment and this new switchable policy is described below. The switchable policy in this case is also symmetrical for eastbound and westbound trains. Thus we focus on eastbound trains in the description below.
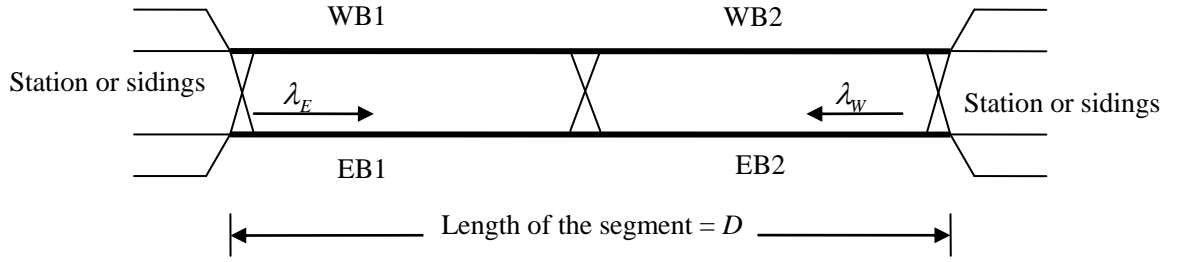
Figure 22. Double-track railroad segment

Switchable policy for double-track segment with crossovers at middle

1. All eastbound slow trains travel on the lower track. Upon arrival, if EB1 is not occupied by trains travelling in the opposite direction, the slow train starts travelling on EB1 immediately. If EB1 is occupied by trains travelling in the opposite direction, the slow train waits for the opposing moving trains to finish traveling on EB1 before proceeding. Upon reaching the end of EB1, if EB2 is not occupied by trains travelling in the opposite direction, the slow train starts travelling on EB2 immediately. If EB2 is occupied by trains travelling in the opposite direction, the slow train waits for the opposing moving trains to finish traveling on EB2 and switch to WB1 before proceeding.

2. When an eastbound fast train arrives at EB1, if no slow train has arrived at EB1 within $(T_s - T_f)/2$ time units from now, the fast train starts travelling on EB1. If there are opposing moving trains travelling on EB1, the fast train waits for the opposing moving trains to finish travelling on EB1 before proceeding. When a fast train arrives at EB1, if a slow train has arrived at EB1 within $(T_s - T_f)/2$ time units from now, then the fast train will try to use track WB1.

   The fast train will use WB1 only when it is empty and EB2 is not occupied by trains traveling westbound (if EB2 is occupied by a train travelling westbound, this train will switch to WB1 using the crossing, and thus deadlock might occur). If WB1 is occupied by trains travelling in either direction or EB2 is occupied by trains traveling westbound, the fast train will use EB1.

3. When an eastbound fast train reaches the end of track EB1, if no slow train has arrived at EB2 within $(T_s - T_f)\sigma / 2$ time units from now, the fast train starts travelling on EB2. If there are opposing moving trains travelling on EB2, the fast train waits for the opposing moving trains to finish travelling on EB2 and switch to WB1 before proceeding. When an eastbound fast train reaches the end of track EB1, if a slow train has arrived EB1 within $(T_s - T_f)\sigma / 2$ time units from now, then the fast train will try to use track WB2. The fast train will use WB2 only if it is empty. If WB2 is occupied by trains travelling in either direction, the fast train will use EB2.

4. When an eastbound fast train reaches the end of track WB1, if no slow train has arrived at EB2 within $(T_s - T_f)\sigma / 2$ time units from now, the fast train switches back to its designated track and starts travelling on EB2. When an eastbound fast train reaches the end of track WB1, if a slow train has arrived EB2 within $(T_s - T_f)\sigma / 2$ time units from now, then the fast train will try to continue to use track WB2. The fast train will use WB2 only if it is empty. If WB2 is occupied by trains travelling in either direction, the fast train will use EB2.

For this new switchable policy (also illustrated in Figure 23), when the fast train arrives at the first half of the track segment, this fast train will try to switch as long as it can catch up with a slow train on the first half of the track (e.g., one slow train has arrived at the first half of the segment within $(T_s - T_f) / 2$ time units). However when the fast train arrives at the second half of the track segment, this fast train will try to switch if at least one slow train has arrived at the second half of the segment within $(T_s - T_f)\sigma / 2$ time units. The threshold factor σ is only effective for the second half of the track segment.

After adding crossovers at the middle of the track segment, deriving the delays for trains analytically becomes intractable. Thus, in order to evaluate the performance of the switchable policy against a dedicated policy for the track segment with

crossovers, we can only rely on simulations to obtain the train delays for the switchable policy. We use the previous base case parameter settings for the following numerical experiments. Figure 24 shows the comparison of the performance measure (fast train delays) of the switchable policy and the dedicated policy for both the networks with and without crossovers. Figure 25 shows the relationship between the length of the segment and the advantage of the switchable policy for networks with crossovers and without crossovers. Figure 26 explores how the advantage of the switchable policy over the dedicated policy changes as we change the speed of slow and fast trains when there exists crossovers. We can see that the switchable policy reduces the fast train delay further if crossovers are introduced into the network. The intuition behind this are: (1) the switching rate of the fast train will increase because now fast train can switch as long as half of the reverse direction track is empty; (2) The switched fast train does not have to travel all the way on the reverse direction track. Adding crossovers at the middle of the track segment not only further reduces the fast train delay but also the slow train delay. Figure 27 illustrates the effect of the crossovers on the slow train delays.
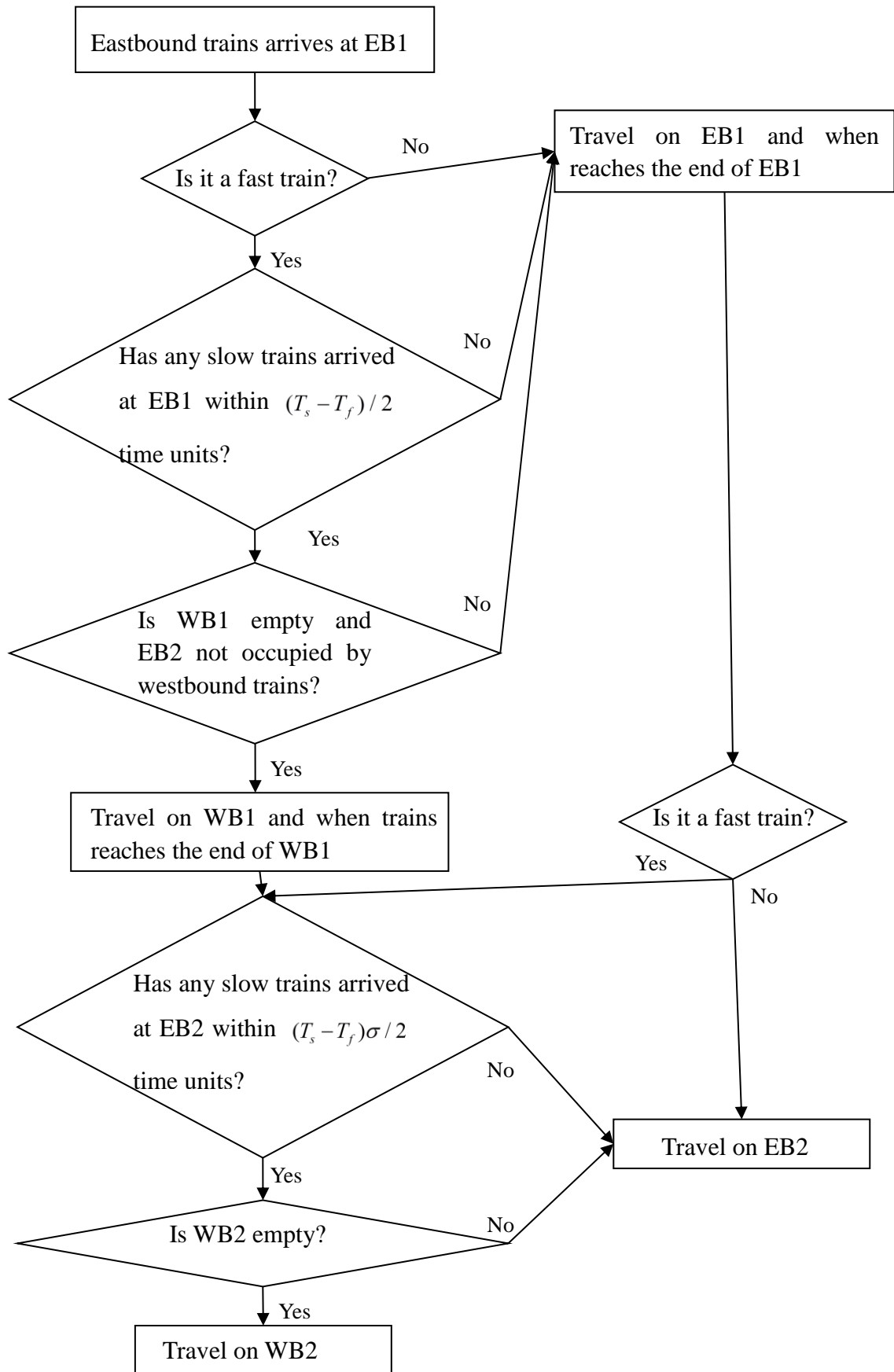
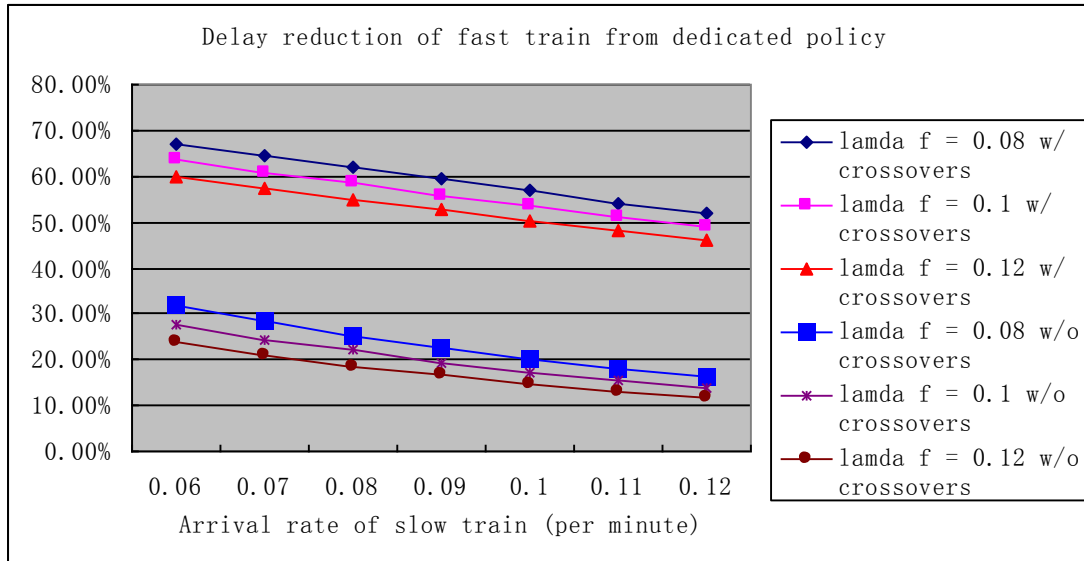**Figure 23. Switchable policy for double-track segment with crossovers**
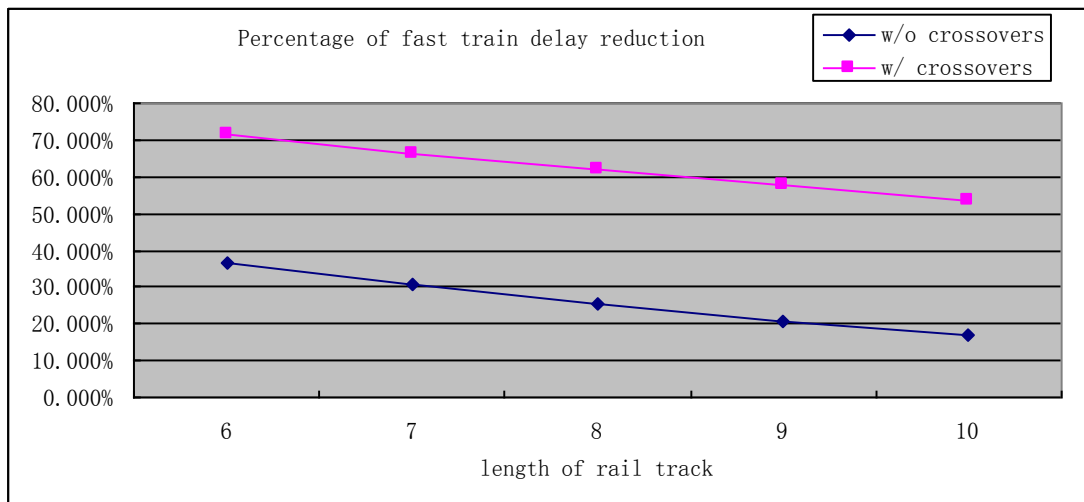
Figure 24. Varying arrival rates

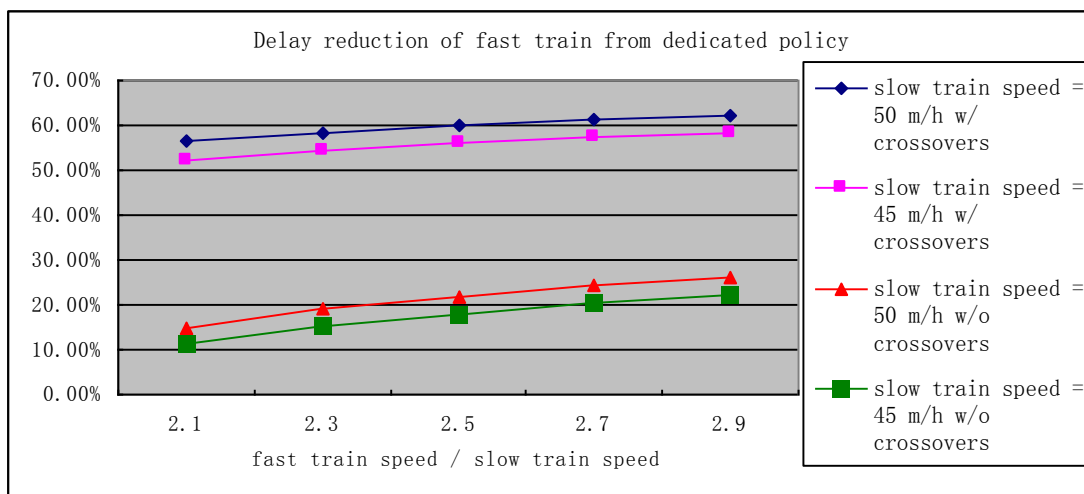**Figure 25. Varying track length**

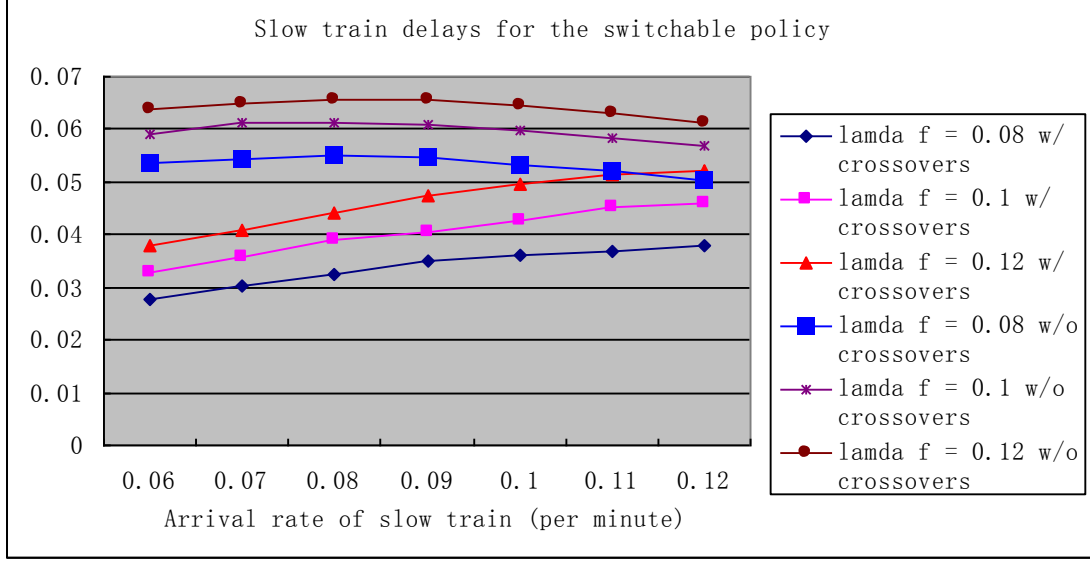

Figure 26. Varying train speeds

Figure 27. Slow train delays for the switchable policy

6. Conclusion

Nowadays, passenger trains are traveling faster and faster. However sometimes passenger trains have to share some portions of the railway with freight trains. The most natural and popular dispatching rule for double-track segments is to dedicate one track for trains traveling in one direction. The major drawback of this simple rule is that a fast train can be caught behind a slow train and experience significant delay. In this paper, we propose a switchable dispatching policy for a double-track segment. The new dispatching rule enables a fast train to pass a slow train by using the track traveled by trains in the opposite direction if the track is empty. We use queueing theory techniques to derive the delay functions of this policy. Our numerical experiments show that our fast train delay functions produce accurate results as compared to simulation results. The numerical experiments also show that a switchable policy can reduce the fast train delay by as high as 30% compared to a dedicated policy. We then extend our switchable policy to work on a double-track segment with crossovers at the middle of the track segment. Adding crossovers at the middle of the track segment further improves the efficiency of the switchable policy. The switchable policy reduces the fast train delay of a dedicated policy by as high as 65%.

When there is multiple train speeds (more than two), just checking the position of the slower proceeding train when the faster train arrives is not sufficient to determine whether or not to switch since there might still be another slower train ahead of it. Thus, a switching policy needs to take into consideration the total potential delay of the fast train. Mu (2011) developed a switching policy based on if the potential delay of the faster train is greater than a threshold value. On a simulation study this extended switchable policy was shown to dominate the dedicated policy for cases of multiple train speeds. There are cases where different trains have different priorities. One possible modification to the current switchable policy to care for different priorities is to give different switching threshold values to different train priorities (e.g., a high priority train will have a lower switching threshold value which makes the high priority train switch more frequently). Future research can modify the developed analytical model to address these extensions.

References

Chen, B., Harker, P.T., 1990. Two moments estimation of the delay on single-track rail lines with scheduled traffic. Transportation Science 24 (4), 261-275.

Dessouky, M.M., Leachman, R.C., 1995. A simulation modeling methodology for analyzing large complex rail networks. Simulation 65 (2), 131-142.

Dorfman, M.J., Medanic, J., 2004. Scheduling trains on a railway network using a discrete event model of railway traffic. Transportation Research Part B 38 (1), 81-98

Frank, O., 1966. Two-way traffic on a single line of railway. Operations Research 14 (5), 801-811.

Goverde, R. M.P., 2005. Punctuality of railway operations and timetable stability analysis, PhD Thesis, Delft University of Technology. TRAIL Thesis Series no. T2005/10.

Greenberg, B.S., Leachman, R.C., Wolff, R.W., 1988. Predicting dispatching delays on a low speed, single track railroad. Transportation Science 22 (1), 31-38.

Harker, P.T., Hong, S., 1990. Two moments estimation of the delay on a partially double-track rail line with schedule traffic. Journal of Transportation Research forum 31, 38-49.

Higgins, A., Kozan, E., 1998. Modeling train delays in urban networks. Transportation Science 32 (4), 346-357.

Hooghiemstra, J.S., Middelkoop, A.D., Teunisse, M.J.G., 1998. Improving the Dutch railway services by network-wide timetable simulation. Computers in Railways, vol. VI. WIT press, 299-308.

Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. Transportation Research Part B 35 (3), 271-292.

Huisman, T., Boucherie, R.J., Van Dijk, N.M., 2002. A solvable queueing network model for railway networks and its validation and application for the Netherlands. European Journal of Operations Research 142 (1), 30-51.

Kelton, W., Sandowski, R., Swets, N., 2009. Simulation with Arena, McGraw-Hill Science/Engineering/Math.

Kuckelberg, A., 2011. Microscopic Disposition track-bound transportation under real time conditions, PhD Thesis, Rheinisch-Westf`alischen Technischen Hochschule Aachen.

Li, F., Gao, Z., Li, K., Yang, L., 2008. Efficient scheduling of railway traffic based on global

information of train. Transportation Research Part B 42 (10), 1008-1030.

Lu, Q., Dessouky M., Leachman, R.C., 2004. Modeling train movements through complex rail networks. ACM Transactions on Modeling and Computer Simulation 14 (1), 48-75.

Medanic, J., Dorfman, M.J., 2002. Efficient Scheduling of traffic on a railway line. Journal of Optimization Theory and Algorithms 1152 (3), 587-602.

Mu, S., 2011. Minimizing Congestions on Railway Networks, PhD Thesis, University of Southern California.

Mu, S., Dessouky, M.M., 2011. Control rules for dispatching trains on general networks with multiple train speeds, Working Paper-2012-01, Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA.

Ozekici, S., Sengor, S., 1994. On a rail transportation model with scheduled services. Transportation Science 28 (3), 246-255.

Ren, J.F., Takahashi, Y., Hasegawa, T., 1996. Analysis of impact of network delay on multiversion conservative timestamp algorithms in DDBS. Performance Evaluation 26, 21-50.

Schwanhäußer, W., 1974. The design of the buffer times in the structure of the railway timetable, PhD Thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen.

Weymann, H.G.F., 2011. Quality of heuristics in the disposition of railway operations, PhD Thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen.

Wolff, R.W., 1982. Poisson Arrivals See Time Averages. Operations Research 30 (2), 223-231.

Yalçınkaya, Ö., Bayhan, G., 2009. Modeling and optimization of average travel time for a metro line by simulation and response surface methodology. European Journal of Operational Research 196 (1), 225-233.

Yuan, J., 2006. Stochastic modeling of train delays and delay propagation in stations, PhD Thesis, Delft University of Technology. TRAIL Thesis Series no.T2006/6.

Yuan, J., Hanson, I.A., 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. Transportation Research Part B 41 (2), 202-217.