RANDOM DELAYS FORMING IN
THE DENSE TRAIN FLOW
BORIS DAVYDOV1, VLADIMIR CHEBOTAREV2 & KSENIYA KABLUKOVA2
1Far Eastern State Transport University, Khabarovsk, Russia
2Computing Center, Far Eastern Branch, Russian Academy of Sciences, Khabarovsk, Russia

ABSTRACT
Random impacts and interaction of trains in a dense flow cause small deviations of arrival times. A model for the forming of train delays is proposed. Formulas for the delay distributions depending on distribution functions of random stops and initial departure times are obtained. Influence of the driver’s efforts on distribution resulting is also considered. We also analyze the properties of arrival intervals and deviation scatterings obtained from initial distributions with real-world parameters. Statistical data collected on a suburban railway are used to compare the model distributions with the real scatterings of train delays. Obtained results confirm a good fit with the proposed delays model.

Keywords: train traffic, online adjustments, stochastic model, optimal rescheduling, train delays.

1 INTRODUCTION
Stochastic model of train traffic allows determining the probabilistic characteristics of intertrain intervals and deviations of the arrival times at the terminal station. The calculation is based on the fact that the travel time of each train along the section is the sum of the intervals that correspond to the elementary operations such as run and stop. When the probabilistic approach is used, it is usually assumed that the execution time of each operation is scattered under the influence of factors that cannot be accurately predicted. These factors are variations in traction characteristics of the locomotive, the influence of weather conditions and also peculiarities of the driver’s work. In addition, the train is delayed by external events such as conflicts with other trains and unplanned restrictions on infrastructure. Each of these events leads to additional scattering of the corresponding operating time. Process of multistage formation the output probability distribution is described by the stochastic model in [1] which we will call the generalized stochastic model. This model is graphically interpreted in Fig. 1. The graph shows that means of the departure and arrival times, denoted by \(T_{dep}\) and \(T_{arr}\) respectively, are shifted relative to the planned moments \(T^X\). Operational times at the open tracks and at the stations \(A, B, C\) are scattered due to random influences. Departure moments are distributed within intervals \(\Delta^{dep}X\). Resulting scattering zone is a “fan” of possible train trajectories.

The work [2] shown the accumulation of delays can be described by a probabilistic operational graph. Each node of the graph corresponds to the planned stop. Parameters of the actual state are determined by a set of previous operations. Some of them reflect the running operations at the segments adjoining the station while others represent the passenger transfer. Random moment of departure from the station \(B\) is determined by set of arrival times together with the strategy of changing the order of trains having the delays.

The train sequentially passes a series of control points when moving along the route. It is necessary to explore in more detail the process of distribution alteration when moving from point to point which depends on mode of the corresponding operation on open tracks and at the station. In Section 3, we discuss how the distribution of train arrival intervals is formed in the condition of random departure moment scattering and occurrence of unplanned delays along the route. This allows you to make general requirements for the train schedule.
Section 4 proposes a model for the formation of delays in the presence of random deviations that arise during the departure and movement of trains. The results of the theoretical analysis are confirmed by statistical data obtained on the Russian railways. This consideration is given in Section 5. The paper concludes with a discussion the results obtained.

2 LITERATURE REVIEW

The well-known papers [3]–[5] provide an extensive survey of recovery models and algorithms for effective timetable creating and real-time railway rescheduling. Macroscopic model contains least details and have a more aggregated representation of some resources (e.g. a station), while microscopic representation includes a lot more detail. The micro-model is losing its effectiveness in the presence of random disturbances.

From the viewpoint of queuing theory all the points of conflict on open tracks and at stations may be interpreted as server systems. This is considered as the mesoscopic approach to create the train traffic model [6]. Complete time of a train passage is a sum of time intervals which corresponds to the elementary operations. Random components which are due to peculiar specific conditions of train traffic and style of the driver’s work are taken into account when creating a common distribution of output delays.

Analyses of previous studies showing an increasing interest in stochastic approaches and models that help create adequate forecasts and make optimal decisions for traffic recovery. Prediction of the uncontrolled oscillations of speed and stops gives the ability to simulate their likely consequences. One of the first papers considering this problem [7] proposed analytical method for delay computing by a convolution of the initial (inlet area) deviations and failures arising again. The analysis was limited because it used the assumption of uniform distribution of secondary delays. In reality, this assumption is not confirmed.

Most stochastic models to analyze propagation of delays focused on single track routes or on simple junctions. Classical models had been studied intensively, most importantly in [8], [9]. The papers proposed the approximation of delay distributions to reduce the
computational effort. Each running interval is considered as a random variable with the exponential distribution. The papers do not consider the probability distribution of the arrival headways and the mechanism of delay propagation through the chain of trains.

The paper [10] represents trains traffic in a form occupying an intermediate position between the macro-model and model with very detailed description. Approximation method for the exact representation of delay distribution is proposed. Cumulative distribution is calculated from the sequence of activities which is determined by a stochastic event graph.

The stochastic analytical model which uses the infinite server as a model of the train interaction proposed in [11]. The paper investigated delays to a fast train caught behind slower ones by capturing both scheduled and unscheduled movements. The running time distributions for each train service are obtained by solving a system of linear differential equations. This model gives a good insight into delay propagation on a simple line but becomes too complicated to handle when dealing with large scale real-world networks.

The stochastic model for delay propagation and forecasts of arrival and departure events proposed in [12]. The model takes the general train waiting policies on stations in account and considers discrete distributions of travel times. The problem is formulated with respect to a directed event graph. Authors assume the two-train model as a basis of the analysis which generates hints to dispatcher and to passengers. Studies of this model allows you to define the order of train departure and calculate distribution of the arrival moments. The discrete stochastic model was also used in [13] which enhances the formulas of the above mentioned paper to calculate probability distributions for train connections. Another model described in [14] is also designed to assess the knock-on delays. New scheme proposed for the exact determination of parameters the theoretical distribution model based on the maximum likelihood method.

The approach based on the periodic event scheduling problem was developed in [15]. The author included exogenous perturbation into the model and controls the spread of the resulting delay. The paper used an analytical approach such as the max-plus algebra (MPA). With MPA can be calculated some key characteristics, for example stability of the train schedule.

A number of studies considered the model that comprehends two train runs with stopping at a station for the passenger transferring [16]. The model illustrated as a Petri net graph. The approach proposed in [2] uses a probabilistic operational graph which considers transfer operation and conflict situations associated. The authors argue a mesoscopic modelling of traffic is the purposeful approach to compute the delay propagation. The paper uses two-train model and the FIFO-rule for conflict resolution. Using of both conditional and unconditional convolution and of “excess beyond” operations is specifically made. Proposed model is logically incomplete. The authors consider only station and crossing as the check points where conflicts can occur. To complete the model, you must use a resolution of the sequence-of-trains conflicts which appear at open tracks.

Various models which describe the functions of the frequency and length of non-scheduled delays are discussed in prior works. Thus, in the papers [10], [17]–[19] the exponential probability distribution is used as a stochastic model of disturbances. Obviously, accumulation of the individual random variables leads to exponential behavior of train arrival delays. This problem is explored in [20] and requires further study especially for the mixed flow of passenger and freight trains.

3 PROBABILISTIC DISTRIBUTION OF THE ARRIVAL HEADWAYS

We propose the following model of formation both arrival headways and arrival deviations (delays). Denote the primary delay of the first train by \( \tau \). Its distribution function is denoted
by \( G(t) = \mathbb{P}(\tau < t) \). If the random variable \( \tau \) is absolutely continuous then we denote its density function by \( g(t) \).

The primary non-scheduled stop of the first train causes a chain of secondary delays (see Fig. 2). The secondary delay of the \( k \)-th train is denoted by \( \tau_1, 2 \leq k \leq n \). We assume that \( \tau_1 \equiv \tau \). We denote by \( G_k(t) \) the distribution function of \( \tau_k, 2 \leq k \leq n \), i.e. \( G_k(t) = \mathbb{P}(\tau_k < t) \).

Secondary delays \( \tau_k, 2 \leq k \leq n \), depend on the initial headways too. In according to our model these headways, i.e. departure time interval between trains with numbers \( k \) and \( (k-1) \), \( 2 \leq k \leq n \), are of the form \( \mu_k + t_0 \), where \( t_0 \) is a constant which denote a minimal safety time interval between two consecutive trains, and \( \mu_2, \mu_3, \ldots, \mu_n \) are independent identically distributed random variables with a distribution function \( \Psi(t) = \mathbb{P}(\mu_i < t) \). The probability density function of \( \mu_k \) (in the case of its existence), is denoted by \( \psi(t) \).

Note, we impose a restriction of mutual independence of \( \mu_2, \mu_3, \ldots, \mu_n \), which would seem not always consistent with reality. However, such a restriction allows us to obtain relatively compact formulas, the conclusions of which, surprisingly enough, are already quite comparable with real statistics.

The arrival headways between trains with numbers \( k \) and \( (k-1) \) are denoted by \( v_k, 2 \leq k \leq n \). We denote the distribution function of \( v_k \) by \( W_k(t) = \mathbb{P}(v_k < t), 2 \leq k \leq n \).

In our previous work [1], we have investigated the headway deformation process caused by the delay of random duration. We considered the following two situations: when trains depart from the initial station on time (a) and when the departure times are shifted by random influences (b). Typically, the primary delay \( \tau \) has a two-parameter exponential distribution with a density \( g(t) = I(t > b) \lambda e^{-(t-b)}, \) where \( \lambda > 0 \) and \( b \geq 0 \).

![Figure 2: The process of delay propagation.](image-url)
In the case when \( \mu_k, 2 \leq k \leq n \), are equal to some positive constant \( T \), and \( \tau \) has a distribution with a density \( g(t) = I(t > b) \lambda e^{-(t-b)} \), one can deduce from [21, Theorem 2] the following equality,

\[
W_n(t) = I\left(0 \leq b \leq (k-2)T\right) \left[I(0 < t - t_0 \leq T) ae^{-\frac{t-t_0}{(k-1)T+b}} + I(t-t_0 > T)\right] + \\
+ I((k-2)T < b < (k-1)T) \left[I(0 < t - t_0 \leq (k-1)T - b) ae^{-\frac{t-t_0}{(k-1)T+b}} + I(t-t_0 > (k-1)T - b)\right] + I(b \geq (k-1)T) I(t > t_0).
\]

(1)

The graphs of distribution functions for the first three arrival headways are depicted in Fig. 3. It is assumed here the primary delay has the exponential distribution with the parameter \( \lambda = 0.3 \). This parameter corresponds to the value that is actually observed in the suburban train traffic of the Russian Railways. Fig. 3(a) shows that the primary delay \( \tau \) practically does not affect the fourth train and all subsequent ones.

In [21] we also announced, how distribution functions of arrival headways depend on the densities \( g(t) \) and \( \psi(t) \). The theorem [21, Corollary 4] approves the following. Let \( \mu_j, 2 \leq j \leq n \), be independent identically distributed random variables with a probability density function \( \psi(x) \). Besides of that the value of \( \tau \) is independent of \( \mu_j \) and has probability density function \( g(x) \). Then

\[
W_2(t) = I\left(t > t_0\right) \left[\int_{u}^{\infty} \left(\int_{z+1}^{\infty} g(x) dx\right) \psi(z) dz\right],
\]

\[
W_k(t) = I\left(t > t_0\right) \left[\int_{u}^{\infty} \psi(z) dz + \int_{u}^{\infty} \left(\int_{z+1}^{\infty} g(x) dx\right) \psi(z) dz\right] \psi^{(k-2)}(u) du, \quad 3 \leq k \leq n.
\]

(2)

(3)

Here \( \psi^{(j)}(u) \) is the \( j \)-fold convolution of the density \( \psi(u) \).

Figure 3: Behavior of the functions \( W_i(t) \). (a) Without the departure scattering; and (b) When there is initial gamma distribution.
As a result of the processing of statistical data, one can assume that $\mu_1, \mu_2, \ldots, \mu_n$ have the same gamma distribution with density $\psi(t) = I(t > 0)\frac{e^{-\beta t}t^\alpha}{\Gamma(\alpha)\beta^\alpha}$, where $\alpha$ and $\beta$ are positive parameters, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$.

The graphs of distribution functions of the arrival headways are constructed in Fig. 3(b) according to eqns (2), (3). Here the parameters of $\Gamma$-distribution the departure headways are equal to $\alpha = 14$, $\beta = 0.5$. Comparing the graphs for different $k$ and also as a result of calculations, we can conclude that stopping the train 1 with a high probability can affect the movement of only the next two trains.

The proposed model allows us to solve the following problem which is important from the point of view of the applications. Let us consider a case, when departure moments have no any scattering, and only a primary delay $\tau$ can be a cause of the incident. In this case there exists some positive constant $T$ such that $\mu_j = T$, $j = 2, \ldots, n$. Evidently, with an increase in $T$ the number of trains that are affected by given $\tau$ decreases.

Let $p$ is given probability, $0 < p < 1$, and denote the number of knock-on delays by $N$. The problem is the following: to find $T(p, m)$ such that $P(N \geq m) \leq p$ for all $T \geq T(p, m)$. In [21] we have shown that in the case when $\tau$ is exponentially distributed with a parameter $\lambda$, the following equality holds, $P(N \geq m) = e^{-\lambda mT}$. Thus the desired condition is the following:

$$T(m, p, \lambda) = T(m, p, \lambda) \geq \frac{1}{m\lambda} \ln \frac{1}{p}.$$ 

Let mentioned probability $p = 0.1$. The behaviour of $T(m, p, \lambda)$ as a function of the continuous parameter $m$ with $\lambda = 0.26$ and $\lambda = 0.15$ is shown in Fig. 4.

These graphs allow us to estimate how to optimally increase the departure headway in order to reduce the number of knock-on delays. However, the issue of the primary delay effect on the distribution of arrival delays on the terminal station remains unclear. This problem will be discussed in Section 4.

Figure 4: Behavior of the function $T(m, p, \lambda)$ with $\lambda = 0.15$ and $\lambda = 0.25$. 
4 THE MODEL OF FORMATION THE ARRIVAL TIME DEVIATIONS

The proposed model of train traffic assumes that the primary non-scheduled stop causes a chain of arrival delays. According to our model, the departure time interval between two consecutive trains with numbers $j$ and $(j - 1)$ is equal to $\mu_j + t_o$, where $\mu_j$ is a random variable, $j = 2, 3, \ldots, n$, $t_o$ is the minimal permissible duration of the time interval between trains. One can show that $T^{(m)}$, the departure time of the $m$-th train, satisfies the following equality:

$$T^{(m)} = \sum_{j=2}^{m} \mu_j + (m-1)t_o, \quad m = 2, 3, \ldots, n.$$  

Primary stop $\tau$ causes the knock-on delays of successive trains (see Fig. 5).

According to [21, Corollary 13], the distribution functions of arrival delays $(k = 2, \ldots, n)$ depend on both the primary delay distribution and distribution of departure headways in the following way:

$$G_k(t) = I(t > 0) \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} g(z) dz \right) \psi^{(k-1)}(y) dy,$$

where $\psi^{(j)}(y)$ is a $j$-fold convolution of the density $\psi(\cdot)$.

Statistical analysis of real train traffic data on the Russian railways shows that the primary delay $\tau$ has a two-parameter exponential distribution with a density $g(t) = I(t > b) \lambda e^{-\lambda(t-b)}$, where $\lambda > 0$, $b \geq 0$. Note that usually $0.2 \leq \lambda \leq 0.4$. As before we assume, $\mu_j$ has gamma distribution with density $\psi(t) = I(t > 0) \frac{e^{-\alpha b}^{\alpha-1}}{\Gamma(\alpha) \beta^\alpha}$, $\alpha > 0$, $\beta > 0$. In this case, according to [21, Corollary 15], it follows from (4) that $G_k(t)$ has the form

$$G_k(t) = I(t \geq 0) \left(1 - e^{-\alpha t}\right),$$

and $a$ satisfies the equality

$$a = (\lambda \beta + 1)^{-(k-1)\alpha}.$$  

Figure 5: The process of random train delay formation.
Graphs of the functions $G_i(t)$ according to the eqn (5) with real parameters of the Russian railway traffic are depicted in Fig. 6.

Let $P(\tau_i = 0) = p$, where $p$ is given probabilistic level, $0 < p < 1$, i.e. the delay of train $2$ is equal to zero with probability $p$. Let us find out how the probability $P(\tau_i = 0)$ varies with increasing $k$.

By virtue of (5), $P(\tau_i = 0) = G_i(0 +) = 1 - (\lambda \beta + 1)^{(t-\lambda)}$. Since $p = 1 - (\lambda \beta + 1)^{-\lambda}$, we obtain $P(\tau_i = 0) = 1 - (1 - p)^{t-\lambda}$, $3 \leq k \leq n$. Fig. 7 illustrates the obtained dependency $f(k,p) = 1 - (1 - p)^{t-\lambda}$ with fixed $p = 0.1$, $p = 0.3$ and increasing parameter $k$. Thus $P(\tau_i = 0) \to 1$ if $k \to \infty$.

5 CONFIRMATION THE MODEL ADEQUACY UNDER THE REAL STATISTICS

Let us consider the following random variable: the deviation of the real moment of arrival at a certain station from the scheduled one. Denote it by $\xi$. Statistical analysis of data on this random variable, received from the Russian railways, has led to the conclusion that in many cases they obey the modified exponential law with the distribution function (5), where $a$, $\lambda$,

![Figure 6: Graphs of output delays $G_i(t)$ for fixed $\alpha = 3$, $\beta = 7/\alpha \approx 2.3$, $\lambda = 0.25$.](image1)

![Figure 7: Graphs of the probability $P(\tau_i = 0)$ in cases: $P(\tau_i = 0) = p = 0.1$ and 0.3.](image2)
are some parameters, $0 \leq a \leq 1$, $\lambda > 0$. Using data on the suburban trains of the direction “Moscow-Tver” for the period: January 11–15, February 1–6, 2016, we obtained a sample from the distribution of $\xi$ of the size $n = 50$ with the sample mean 1.44 and sample variance 2.7. We tested the hypothesis that $\xi$ obeys distribution (5) with $\lambda = 0.35$ and $a = 0.64$. To this end we applied the Kolmogorov goodness-of-fit test with the significance level $\alpha = 0.05$ and obtained the fit between the hypothesis and the sample data (see Fig. 8).

It should be noted that in considered example the deviation $\xi$ is nonnegative. But in reality, it can frequently be both positive and negative. Positive values occur in the event of a delay. Negative values are possible due to the fact that sometimes early arrivals take place.

Although the hypothetical distribution function from Fig. 8 is constructed for deviations without any details about the train number $k$, it is correlated with the graph of the function $G_2(t)$ from Fig. 6 in some degree. This allows us to assume that the distribution of the deviation $\xi$ is mainly determined by the distribution of the delay $\tau_k$.

Another one important practical result of the considered model is that it enables us to estimate the standard deviation (SD) of the actual arrival delays at the destination station. As an example, we calculated this parameter for the suburban railway line. The data analyzed were collected at the Tver station in the period of 2016 January and February. Statistical data show the following real parameters for exponential distribution of $\tau$ and gamma distribution of $\mu$: $\lambda = 0.25$, $b = 0$, $a = 0.6$, $\beta = 11.7$, $k = 2$. Using eqns (5), (6) we have:

$$SD^2 = \int_{-\infty}^{\infty} (t-a_2)^2 dG_2(t) = \int_{-\infty}^{\infty} t^2 dG_2(t) - a_2^2 = \int_{-\infty}^{\infty} t^2 g_2(t) dt - a_2^2 \approx 10.997. \quad (7)$$

Here $a_2 = \int_{-\infty}^{\infty} t dG_2(t) = \int_{0}^{\infty} t g_2(t) dt \approx 1.763$. Thus, theoretical $SD$ is equal to 3.316. This corresponds with the real statistics which shows the SD amount is 3.32 min for the mentioned station.

Figure 8: Empirical distribution function and the hypothetical distribution function of the form (5) with $\lambda = 0.35$ and $a = 0.64$. 
6 CONCLUSIONS AND FUTURE WORK

The developed technique allows to solve the problem of forecasting the late arrivals of trains if there are known specific operating conditions of the railway section. The databases should have probabilistic parameters of primary stops and headways for each specific scenario. These data are formed by accumulating statistical data on functioning the section in previous periods.

The results of calculations allow us to conclude that the knock-on delay affects only two succeeding trains. This conclusion is valid for the current intensity of suburban flow on Russian railways. In this paper, we obtained relations that allow us to estimate the required headway for the given probability of secondary delay.

The study shows the late arrivals of trains to intermediate stations are well described by gamma distribution. Such a convenient approximation makes it possible to simplify the calculation of convolution the densities that correspond to various operations. It is determined that the form of distribution is significantly deformed when there is a set of destabilizing factors, in particular, the process of passenger boarding/alighting.

A further development of this study is an introduction of a random speed variation in the proposed stochastic model. This allows you determining the distribution of output delays more adequately.

ACKNOWLEDGEMENT

The research is partially funded by the JSC Russian Railways (grant 2016 for development the scientific school).

REFERENCES


