P R A C E N A U K O W E P O L I T E C H N I K I W A R S Z A W S K I E J
z. 96 Transport 2013

Dariusz Caban, Tomasz Walkowiak

Politechnika Wrocławska, Instytut Informatyki, Automatyki i Robotyki

CRITICAL STATES RELIABILITY ANALYSIS OF DISCRETE TRANSPORTATION SYSTEMS

The manuscript delivered: May 2013

Summary: The paper presents a resource constrained model of a discrete transportation system that can be used to simulate its operation in presence of faults. The simulation results are used to assess the conditional probability of system failure after a short time after it reaches a specific set of reliability states, i.e. the set of critical states. These are determined as the states when the system is still operational but the probability of failure in near future is unacceptably high. The critical states are shown to be a practical indicator that the system is dangerously degraded.
Keywords: reliability prediction, transportation systems maintenance, simulation

1. INTRODUCTION

The critical states of operation are often used in reliability analysis [2] to describe degraded systems that are still operational, but there is insufficient margin of operability to ensure reliable functioning. This is a special case of multistate reliability analysis, which generalizes the traditional up/down analysis [3, 4].

The critical states analysis is used to predict the impeding serious malfunctions of a system. As such, it can be used to schedule the various maintenance tasks, crucial in case of critical transportation infrastructures. As discussed in [6], vulnerability is of paramount importance in these systems. Critical states are the means to predict these vulnerabilities in a degraded system. As remarked in [1], it is important to consider the resilience of critical infrastructures, i.e. their reliability in abnormal situations.

In the paper, we consider the risks connected with managing a discrete transportation system (DTS), i.e. a system in which vehicles translocate discrete quantities of goods. The proposed reliability analysis is based on the assessment of system operation using Monte Carlo simulation, using a custom designed simulator [8]. The simulator is used to establish a secure level of system resources (vehicles and drivers) and to determine the set of critical states of operation.

2. MODEL OF THE ANALYZED SYSTEM

A discrete transportation system (DTS) relies on a fixed number of vehicles that carry goods by fixed amounts. The flow of goods is discrete, as it corresponds to the movement of the vehicles between the various destinations (or nodes in the travel routes). The system model consists of locations from which goods are collected and to which they are carried and of vehicles travelling between the nodes. The vehicles are manned by drivers. The system is modeled by the various interacting resources used to achieve the transportation goals.

2.1. DISCRETE TRANSPORT SYSTEM RESOURCES

The system makes use of the following resources: the set of nodes, the set of routes between the nodes, the set of vehicles, the set of transportation assignments, the set of travel schedules, the set of vehicle operators (drivers) K, assigned to vehicles when they transport goods between the nodes and the set of maintenance teams that service the vehicles after a break down.

Nodes, routes and assignments

There is a single central node and a number of local ones. The central node is the only destination of assignments generated at the local nodes. The central node generates goods assignments destined to all the local nodes.

Routes represent the direct connections between the nodes. They are characterized by the distance that the vehicles must travel. Taking into account the average travelling speed of vehicles, this determines the latency connected with moving from one node to another. The latency is distorted by the travel delays caused by the traffic congestion. These delays are modeled using a random distribution.

The assignments are generated independently of each other, using random distributions. Poisson distribution is usually used to model this. Each local node has an attribute which determines its characteristic rate of assignments generation. The central node is described by an array of assignments rates, one for each local node. There is a fixed time in which each assignment must be completed. Depending on the nature of the DTS system, this time is fixed by local regulations or is part of the service agreement between the assignees and the transport service provider.

Vehicles

It is assumed that all vehicles of the same type have similar properties. They are described with the same functional and reliability related parameters: capacity (expressed as the number of standard containers), average cruising speed (determining the route latency), failure rate, renewal time. All the vehicles are based in the central node and travel from it to realize the assignments.

At any moment in time, a vehicle may be in one of the following states: it might be en route between nodes (a specific distance from the starting node, carrying specified amount of goods), it might be waiting for goods in a node, it might be stopped due to unavailability of a driver or due to regulatory rest period of its driver.

A vehicle may be realizing multiple assignments at the same time. It is fully loaded with goods if the pending assignments allow it. If there are insufficient assignments for nodes towards which the vehicle is destined, then it may be partially loaded or even travelling empty. It collects goods en route if there are pending assignments in the visited nodes.

Timetables

Vehicles are travelling in accordance to fixed timetables (travel schedules). Each timetable determines the time to leave the central node and a sequence of nodes that must be visited by the vehicle as well as the times of these visits. It describes the daily work of the vehicle associated with the timetable, independent of the actual needs as determined by the assignments.

Vehicles are loaded to their capacity (if there are sufficient assignments) at the central node on starting a travel schedule. On reaching each consecutive node in the timetable, the goods destined to it are unloaded and the goods waiting there are loaded in their place. The time used for unloading and loading is randomly chosen. If there are other vehicles in the node, then they are queued and the period of loading/unloading is extended commensurately. The timetables do not specify the time to leave a node (except the timetable start time).

When the vehicle returns to the central node (at the end of a schedule) it is completely unloaded. It can then be associated with another timetable or it may be placed in the pool of available vehicles, waiting to be associated with a job.

The timetables are not directly associated with vehicles or drivers. Instead, any available vehicle and operator is allocated to each schedule. If there are no vehicles or drivers available, then the timetable cannot be realized.

Vehicle drivers

Whenever a vehicle is assigned to a job (to a timetable), a driver must also be associated with it. Any unallocated driver can be associated with any vehicle. Only one driver at a time is associated with a vehicle (since we do not consider long distance routes with standby drivers).

The working time of vehicle operators is regulated by local and EU law. The daily working hours are limited (to 8 hours), there are also compulsory rest breaks while driving. Thus, at any time the driver can be either: driving, resting (between work shifts), pausing (i.e. having a compulsory break while driving) or waiting to be assigned a job.

It is assumed that the drivers work in 8 hour shifts. The state of each driver changes to “resting” whenever his daily working time limit is exceeded and he arrives at the central node. He stays in this state until the beginning of his shift next day. Then, his state changes to available. If there is a pending driving schedule (timetable) and an available vehicle, then his state changes to "driving".

While driving, the driver has to heed the limits on the maximum length of time that he can work without a break. Normally, the timetables assure that the required breaks are fulfilled while the vehicle is loaded in the visited nodes. If a route is unnaturally long or there are travel delays on the way, then the driver is required to take a break en route. The parameters describing the driver model include the daily working hours limit, maximum uninterrupted driving period, minimum break duration.

The allocation of drivers to the jobs is governed by some simple rules: vehicles cannot carry goods between nodes if there is no operator available, driver is chosen from among those, whose daily working time limit allows them to complete the job with at most 10% overtime (i.e. estimated journey time is less than 110% of the leftover time limit).

Maintenance teams

The model does not distinguish any specific parameters of the maintenance teams, just their number. If a vehicle breaks down, it will be repaired by one of the maintenance teams. The distribution of the repair time is associated with the vehicle, not with the team.

2.2. OPERATIONAL FAULTS

The operation of the system is prone to be disrupted due to the occurrence of incidents affecting the availability of the various resources. There are three main categories of operational faults:

- vehicle breakdowns,
- driver absentees at work (due to illness or other incidental leave),
- traffic congestion (resulting in random delays in traversing the routes by the vehicles).

All the faults are temporary. Broken down vehicles are repaired, drivers get well after a period of absence, traffic gets back to normal. In consequence the system is fully repairable and there is no long term degradation.

The vehicles are assumed to break down occasionally, in accordance to their reliability parameters (failure rates). They then stop operation and wait for a maintenance team. On being repaired (after a random repair time), the vehicles continue the work they were realizing before breakdown. Each maintenance team repairs only one vehicle at a time. If all the maintenance teams are currently occupied, then the vehicle repair is delayed until one becomes available.

Drivers get ill or otherwise temporarily unavailable. After a prescribed leave of absence they come back to work. Driver illness is modeled as a stochastic process with three categories of illness [8]:

- short sickness (1 to 3 days),
- typical illness (7 to 10 days),
- protracted disability (10 to 300 days).

In the model the three types of disabilities are treated independently. In each case the actual period of absence is randomly chosen from the appropriate range.

Traffic congestion is not explicitly modeled as a fault. The delays caused by the traffic jams are considered as a random component of the route cruise times.

2.3. RELIABILITY STATES OF THE DTS

The reliability state of the system is characterized by the number of operational resources, at a specific point in time. Thus, it is a vector of the number of drivers n_{Ki}, that are not ill, and the number of vehicles n_{Vi} that are operational:

$$s_i = [n_{Ki}, n_{Vi}] \tag{1}$$

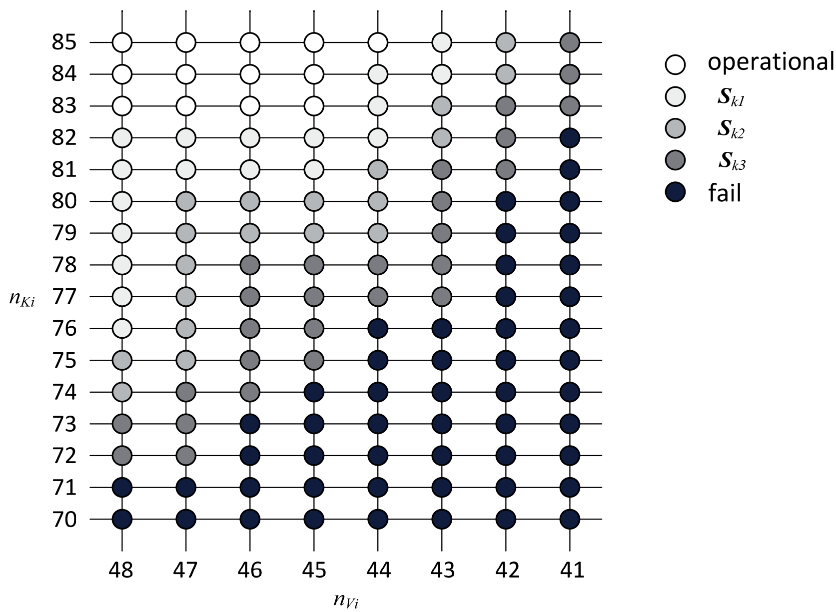

Fig. 1. The map of reliability states of the system with classification information indicating the operational states, the fail states and variants of the critical states

Initially, all the drivers and all the vehicles are available. This initial state of operation $s_o = [n_{Ko}, n_{Vo}]$ corresponds to the most likely situation when all the vehicles are operational and all the drivers are at work. The choice of the initial number of resources is important for the proper planning of system operation. It entails determining the minimal resources ensuring the required system throughput, as well as providing some redundancy to achieve some level of fault tolerance. As discussed in other publications [8], the problem can be solved using simulation techniques.

The same simulation tools may also be used to analyze a DTS system with fixed resources. This is discussed in Part 3 of the paper. In this case, the aim is to categorize all the reliability states s_i into two disjoint sets: the operational states $\boldsymbol{S}_R$ and the fail states $\boldsymbol{S}_F$. It should be noted that this categorization is relative. It is based on the assessed ability of the system to transport all the goods on time. It is assumed that impaired operation cannot cause rejection of transportation assignments. Thus, all the assignments are serviced by the system, though delays in delivery can occur. The likelihood of these delays, depending on the current system reliability state, is the basis of states classification.

During system operation the actual number of available vehicles and drivers varies in time due to the operational faults. This is characterized by the reliability state, as defined in (1), at any given moment of time. For any set of these states, the conditional probability that, in a short time horizon τ, the level of timely deliveries drops below a critical value A_{crit} is determined. When the probability is small the system is regarded as operational. If it is large, the system is said to be in a fail state.

The critical states are defined as a border case between the operational and fail states [2], i.e. they are a subset of the operational states with just a single or a few drivers/vehicles more than in the fail state. The states are used to predict impeding system failure. Fig. 1 describes three different examples of fixing the critical states of the system. Hollow circles denote operational states, black filled circles correspond to fail states. Various levels of gray filling describe the three possible sets of critical states $\boldsymbol{S}_{k1}$, $\boldsymbol{S}_{k2}$ and $\boldsymbol{S}_{k3}$ ($\boldsymbol{S}_{k1}$ is a superset of $\boldsymbol{S}_{k2}$, which is a superset of $\boldsymbol{S}_{k3}$).

3. SYSTEM ANALYSIS

3.1. MEASURE OF SYSTEM PERFORMANCE AND RELIABILITY

The classification of the reliability states of the system, discussed in Part 2, is based on the assessment of the performance of the system. For this purpose, it is necessary to define an adequate reliability/performance measure of the system. Various approaches can be used, a simple one that we propose is based on the guaranteed time of delivery.

Each assignment has a guaranteed time of delivery Δt_g. The real time of delivery Δt is a random variable, which depends on the current volume of assignments, travel delays, reliability state, etc. If an assignment is completed before the deadline, i.e. $\Delta t \leq \Delta t_g$, there is no penalized delay. There is no reward for the early delivery, either. On the other hand, if the assignment is completed after its deadline, $\Delta t > \Delta t_g$, then there is a late delivery penalty incurred. The short term measure of the quality of performance is obtained by counting the assignments that are delivered on time (before the deadline).

The ratio of on-time deliveries a_r is defined as the proportion of assignments that are delivered on time to the total number of assignments in the system during a fixed time period. A 24 hour reporting period is assumed for determining this ratio. The time instances (t_0, t_1, … , t_n) fix the boundaries of the consecutive days, for which the ratio is considered. $N_d(t_i)$ denotes the number of assignments completed in period (t_i, t_{i+1}), while $N_{pd}(t_i)$ denotes the part of them completed on time. The average ratio of on-time deliveries is defined as:

$$a_{ri} = \frac{N_{pd}(t_i)}{N_d(t_i)+1} \tag{2}$$

The average ratio A_r, calculated over the various reporting periods and the various random variable realizations, is used to characterize the overall system performance:

$$A_r = \underset{i}{E}(a_{ri}) \tag{3}$$

The average ratio of on-time deliveries is not adequate for classification of the reliability states (described in Part 2). For this purpose, the conditional probability is assessed that if at time t_i the system is in state s_i, then during at least one period $t_j \in (t_i, t_i+\tau)$ the ratio a_{ri} will drop below the critical value A_{crit}. This probability is further called as the conditional probability of failure and depends on the state s_i and the considered time horizon τ.

3.2. SIMULATION BASED ANALYSIS

The analysis is performed using a simulator, custom designed for this purpose [7, 8]. It is based on the publicly available SSF simulation engine that provides all the required simulation primitives and frameworks, as well as a convenient modeling language DML [5] for inputting all the system model parameters. By repeating the simulator runs multiple times using the same model parameters, we obtain several independent realizations of the same process (the results differ, since the system model is not deterministic).

The average ratio of on-time deliveries is determined by taking all the daily ratios, observed in all the various realizations and computing their arithmetic average. Fig. 2 presents the results of such assessment for the case-study system (described in 3.3), where the calculations are repeated for various possible levels of initially allocated resources.

The conditional probability of failure is much harder to determine. In this case, it is necessary to examine the reliability states occurring during all the simulation runs. If a reliability state of interest occurs, then the daily ratio of on-time deliveries for the next τ days is compared against the A_{crit} value. The probability is assessed as the ratio of the occurrences where the critical level is not met, normalized by all the occurrences of the reliability state. Some reliability states occur very rarely, so it is necessary to have a huge number of simulation runs to obtain valid assessment. For this reason it is not viable to evaluate the conditional probability for all reliability states (to obtain a graph similar to the results presented in Fig. 2). We had to limit the calculations only to the sets of critical states.

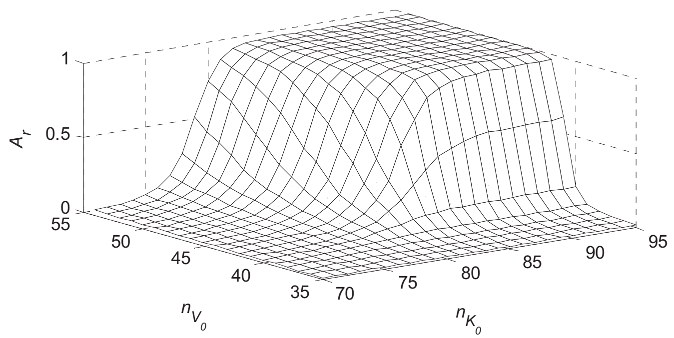

Fig. 2. The average ratio of on-time deliveries for various numbers of vehicles and drivers

3.3. CASE STUDY

All the simulation results that illustrate the presented considerations (Fig. 2 and 3) were computed using a real-life example of a transportation system. The system consists of a central node located in Wroclaw and 22 local nodes located throughout the region. The stream of assignments (generation of cargo) is assumed the same for all the destinations. It is modeled as a Poisson stream with the rate set to 4.16 per hour in each direction. On average this corresponds to 4400 containers to be transported every day.

The set of vehicles is uniform, each can carry 10 containers at a time. The velocity of vehicles is modeled by the Gaussian distribution (average of 50 km/h, standard deviation of 5 km/h). The average loading time is 5 minutes. The mean time to failure of each vehicle is assumed as 20,000 hours. The average repair time is 36 hours (left-truncated normal distribution with standard deviation of 18 hours).

The vehicles are operated by drivers working in 2 shifts (morning: 6 a.m. till 2 p.m., afternoon: from 1 p.m. until 9 p.m.). The rates of driver illness for defined three categories are equal to: 0.003, 0.001 and 0.00025 respectively. The system works with fixed timetables, organized so that the vehicles and drivers have a grace time of 20 minutes after completing one journey, before starting on the next one.

The initial numbers of vehicles and drivers are established on the basis of simulation results presented in Fig. 2. The numbers are fixed as 48 vehicles and 85 drivers, which ensure that the potential amount of cargo transported daily exceeds the average demand by 15% (overall). The figure is also used as the basis of initial classification of states to operational or fail sets, as shown in Fig. 1.

The alternative sets of critical states are fixed arbitrarily to cover the border cases between the sets of operational and fail states. Three case study examples are considered. In the first (set $\boldsymbol{S}_{k3}$), a minimal set of critical states is defined with a very narrow margin between operational and fail states. On the other hand, the set $\boldsymbol{S}_{k1}$ is defined with a wide margin of redundant vehicles and drivers. It should be noted that a small set of critical states yields a small probability of its occurrence. In the considered example, the probability is 0.437 for $\boldsymbol{S}_{k1}$, 0.045 for $\boldsymbol{S}_{k2}$ and 0.002 for $\boldsymbol{S}_{k3}$.

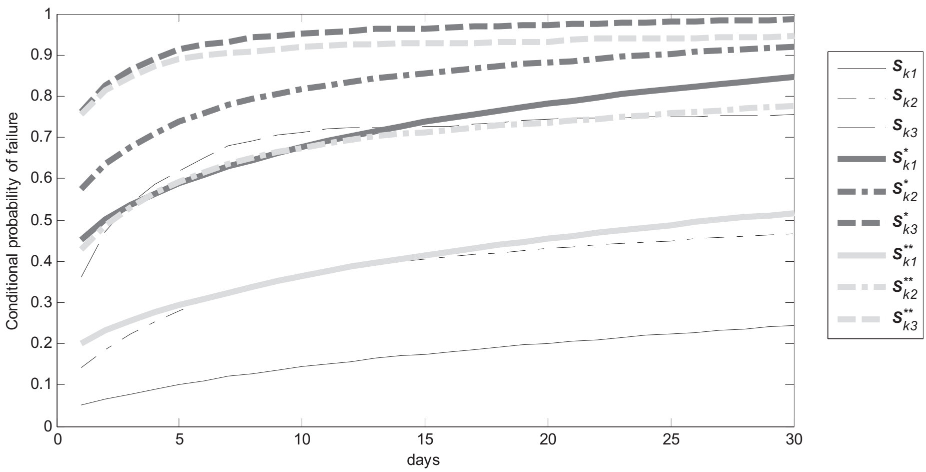

Fig. 3. The conditional probability of failure in different time horizons. Thin black plots correspond to the initial system. Thick grey plots describe various degraded systems

Fig. 3 presents the conditional probability of failure obtained for these sets in case of the various time horizons (black lines). The choice of the set of critical states is arbitrary – the larger the set of operational states that we deem critical, the worse its predictive value (as illustrated in Fig. 3). On the other hand, the critical state is also more likely to occur. As such, it is a management decision that will influence the balance between trustworthiness of the predictions and unjustified alarms.

The usefulness of the critical states becomes apparent in case of a degraded system. The effect of degradation is simulated by altering the intensities of vehicle breakdowns and the frequency of driver illness. This affects the probability of occurrence of the critical states. For the same sets of critical states, the probabilities increase to 0.861 for $\boldsymbol{S}_{k1}$, 0.237 for $\boldsymbol{S}_{k2}$ and 0.241 for $\boldsymbol{S}_{k3}$ when the frequency of illness is increased by 2 times. When the intensity of vehicle breakdowns is increased by 4, the probabilities are increased to 0.595, 0.126, 0.011 respectively. The conditional probabilities of failure also increase as shown by the grey plots in Fig. 3.

4. CONCLUSIONS

The presented results bear out the feasibility of using the concept of critical state operation in the considered class of DTS systems.

The proposed approach to determining the sets of operational, critical and fail states is demonstrated to yield useful information about system performance. While the operational and fail states are sharply defined, the critical states can be fixed to fulfill the specific diagnostic needs of the system. Clearly, in the presented case study, the set $\boldsymbol{S}_{k2}$ is the preferred compromise between the diagnostic capabilities and the risk of false alarms. Still, the other sets can be more adequate in the specific management policies.

Literature

1. Bukowski L.: Dependability Based Concept of Critical Infrastructures Resilience. Proceedings of the XLI Winter School of Reliability Szczyrk 2013.
2. Kołowrocki K., Soszyńska-Budny J.: Integrated Safety and Reliability Decision Support System. Journal of KONBiN, 4(20), 2011.
3. Lisnianski A., Frenkel I., Ding Y.: Multi-state System Reliability Analysis and Optimization for Engineers and Industrial Managers. Springer, 2010.
4. Natvig B.: Multistate Systems Reliability Theory with Applications. Wiley Series in Probability and Statistics, Wiley, New York 2010.
5. Nicol D., Liu J., Liljenstam M., Guanhua Y.: Simulation of Large Scale Networks Using SSF. Proceedings of the 2003 Winter Simulation Conference, 2003, s. 650-657.
6. Nowakowski T.: Podatność na zagrożenia a niezawodność i ryzyko – analiza pojęć. Proceedings of the XLI Winter School of Reliability Szczyrk 2013, s. 13-18.
7. Walkowiak T., Mazurkiewicz J.: Analysis of Critical Situations in Discrete Transport Systems. Proceedings of International Conference on Dependability of Computer Systems, IEEE Computer Society Press, Los Alamitos 2009, s. 364-371.
8. Walkowiak T., Mazurkiewicz J.: Human resource influence on dependability of discrete transportation systems. In: W. Zamojski (et al.) Dependable computer systems, Springer 2011.

ANALIZA NIEZAWODNOŚCIOWA STANÓW KRYTYCZNYCH DYSKRETNYCH SYSTEMÓW TRANSPORTOWYCH

Streszczenie: W artykule przedstawiono model dyskretnych systemów transportowych z ograniczonymi zasobami, wykorzystywany do analizy symulacyjnej ich działania w warunkach występowania uszkodzeń. Wyniki symulacji wykorzystywane są do oceny prawdopodobieństw warunkowych upadku systemu w krótkim horyzoncie czasowym po osiągnięciu określonego zbioru stanów sprawności. Stany krytyczne określane są w oparciu o to prawdopodobieństwo, jako stany, gdy system jest jeszcze sprawny, ale ryzyko upadku w bliskiej przyszłości jest zbyt wysokie. Wykazano, że stany krytyczne mogą być praktycznym wskaźnikiem nadmiernej degradacji systemu.

Słowa kluczowe: analiza niezawodnościowa, zarządzanie systemem transportu dyskretnego, symulacja