The paper discusses the discrete transportation system’s (DTS) availability problem in case of the critical situation, which happens during its ordinary work. The formal model of the transportation system suitable for modelling based on the system behaviour observation is proposed. Monte Carlo approach is in use for a simulation. The system’s availability is calculated as global metric to find if the system is able to realise the loaded set of tasks. The critical situation discussed there is caused by reliability, as well as by functional and human reasons. No restriction on the system’s topology and on a kind of distribution describing the system’s functional and reliability parameters is the main advantage of the approach. The proposed solution seems to be essential for the owner and administrator of the transportation systems.

Keywords: reliability, discrete transportation system, Monte-Carlo simulation, critical situation sets

1. Introduction

At the beginning of the discussion about the critical situations – observable during exploitation processes of the discrete transport system – we have to say that they are not always predictable for system’s owners and managers, but can turn to be very costly for a company and sometimes even damaging. The aim of this paper is to show a method of analyzing the critical situations of a discrete transportation system (DTS), i.e., a transportation system in which goods are transported by a fleet of vehicles of limited capacity. The vehicles operate according to schedules, carrying goods between destinations. The goods are fixed in size (the volume of goods is always a discrete number).

The performance of discrete transportation system depends on different factors [9]. Some of them are deterministic ones, like: distances between destinations, number of drivers and number of trucks. Others have random features, like: amount of goods to be transported, the reliability characteristics of trucks, transportation time (due to traffic jams) and the absence of drivers’ sickness. Each part of the system is characterised by absolutely unique set of features and can cause the critical situation of the whole system if it starts to work in unusual way or the fault or error of it has been noticed. It is hard for an administrator, manager or an owner to understand the system behaviour and to combine the large scale of variant states of it in single – easily observable and controlled global metric as a pointer to make the proper decision in short time period. To overcome this problem we propose a functional approach. The transportation system is analysed from the functional point of view, focusing on business service realized by a system [16]. The analysis is following a classical [4]: modelling and simulation approach. It allows calculating different system measures which could be a base for decisions related to administration of the transportation systems. The metric are calculated using Monte Carlo techniques [7]. No restriction on the system structure and on a kind of distribution is the main advantage of the method. Such approach allows forgetting about the classical reliability analysis based on Markov or Semi-Markov processes [2] – idealised and hard for reconciliation with practice. The results of the system observation – understand as the set of data collected during the simulation process are the basis to define the critical situations and they allow providing the probably proper solution to lift-up the systems in effective way if the critical situation occurs. This is the only sensible way, because the critical situations are the real and not removable part of the system life.

The organization of this chapter is as follows. We start with the description of the Polish Post regional centre of mail distribution (Section 2), which is a base for a developed by authors the discrete transportation system model (Section 3). In our opinion, it seems to be the most adequate to the level of detail to provide the well-established description of the critical situation (Section 4). The quality of the analysed system is measured by the availability defined as an ability to realize the transportation task at
a required time (described in Section 5). Next (Section 6), we give an example of using presented model for the analysis of the Dolny Slask Polish Post regional transportation system in case of critical situation.

2. Real Sophisticated Transportation System – an Idea

The analysed transportation system is a simplified case of the Polish Post. The business service provided by the Polish Post is the delivery of mails. The system consists of a set of nodes placed in different geographical locations. Two kinds of nodes could be distinguished: central nodes (CN) and ordinary nodes (ON). There are bidirectional routes between nodes. Mails are distributed among ordinary nodes by trucks, whereas between central nodes by trucks, railway or by plain. The mail distribution could be understood by tracing the delivery of some mail from point A to point B. At first mail is transported to the nearest ordinary node A. Different mails are collected in ordinary nodes, packed in larger units called containers and then transported by trucks scheduled according to some time-table to the nearest central node. In central node containers are repacked and delivered to appropriate (according to delivery address of each mail) central node. In the Polish Post there are 14 central nodes and more than 300 ordinary nodes. There are more than one million postal items going through one central node within 24 hours. It gives a very large system to be modelled and simulated. Therefore, we have decided to model only a part of the Polish Post transportation system – one central node with a set of ordinary nodes.

Essential in any system’s modelling and simulation is to define the level of details of the modelled system. It could be done if a kind of measures calculated by the simulator is known. Since the business service given by the post system the postal delivery is on time. Therefore, we propose to calculate the time of transporting mails by the system. Since the number of mails presented in the modelled system is very large and all mails are transported in larger amounts – containers, we have decided to use containers as the smallest observable element of the system. Therefore, the main observable value calculated by the simulator will be the transportation time of a container from the source to the destination node. The income of mails to the system, or rather containers of mails as it was discussed above, is modelled by a stochastic process. Each container has a source and destination address. The central node is the destination address for all containers generated in the ordinary nodes. Where containers addressed to any ordinary node are generated in the central node.

The generation of containers is described by some random process. In case of a central node, there are separate processes for each ordinary node. Whereas, for ordinary nodes there is one process, since commodities are transported from ordinary nodes to the central node or in the opposite direction. The containers are transported by vehicles, which require a driver to control it. Each vehicle has a given capacity – maximum number of containers it can haul. Central node is a base place for all vehicles. They start from the central node and the central node is the destination of their travel. The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity. Vehicles operate according to the time-table. The time-table consists of a set of routes (sequence of nodes starting and ending in the central node, times of leaving each node in the route and the recommended size of a vehicle). The number of used vehicles, number of drivers and the capacity of vehicles do not depend on temporary situation described by number of transportation tasks or by the task amount, for example.

It means that it is possible to realize the route by completely empty vehicle or the vehicle cannot load the available amount of commodity (the vehicle is too small). Time-table is a fixed element of the system in observable time horizon, but it is possible to use different time-tables for different seasons or months of the year. Summarizing, the movement of the containers in the system, a container is generated with destination address in some of node (source) at some random time. Next, the container waits in the node addressed to given node are unloaded and empty space in the vehicle is filled by containers addressed to a central node. The process of vehicle operation could be stopped at any moment due to a failure (described by a random process). After the failure, the vehicle waits for a maintenance crew (if there are no available

273
due to repairing other vehicles), is being repaired (random time) and after it continues its journey.
The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity.

3. System Model

We decided to formally model a part of the Polish Post transportation system, one regional section, which consists of one central node and a set of ordinary nodes.

A realization of the system service needs a defined set of technical resources. Moreover, the operating of vehicles transporting mails between system nodes is done according to some rules – some management system. Therefore, we can model discrete transportation system as a 4-tuple:

$$DTS = \langle Client, Driver, TI, MS \rangle,$$

where


3.1. Technical Infrastructure

The technical infrastructure of a discrete transport system is understood as a system of resources - vehicles, reloading machines, service men – and transport network utensils – roads with their traffic characteristics, terminal stations. The transport infrastructure of the discrete transport system can be presented as a directed graph. The terminal station is a node of the transport network system where the vehicles are uploading or unloading with goods caused by consumers’ tasks. The station may be equipped with a storehouse - with limited capacity for example - and supported by reloading machine, cranes, forklift tracks etc. Roads can be modelled as directed arcs connected to nodes. Engineering parameters of the road are integrated and reflected into one representative measure of transport resource available at road segment. Of course such parameter – the average speed for instance – depends on cargo, transport means type, direction of traffic, day time or month time, etc.

The technical infrastructure is allocated by the management system MS for transportation tasks realization on the base of available road nets and functional services. Transport resources are described by their functional (e.g., load capacity of a truck), technical (e.g., fuels expendable per kilometer) and reliability parameters (e.g., mean time between failures or mean time renewal), which may have deterministic or probabilistic nature. Employees (e.g., drivers, workers, machine operators) create a specific class of the system resources.

The transportation task is understood as a pickup of a fixed cargo from the start node and a delivery of it to final node according to assumed time-table. Of course, the transportation task may be defined in more complicated way, e.g., a cargo may be collected in a few nodes and reloaded in several ones. Transport schedule can be defined in different ways, for example a cargo ought to be delivered to the node before the end of fixed time-period, because a train cannot wait for a truck with the cargo.

A transportation task is defined as a sequence of actions and jobs performed by a transportation system in a purpose to obtain desirable results in accordance with initially predefined time schedule.

The transport services are used for execution of user’s (client’s) tasks. The services are built based on the technical infrastructure and the technological services which are involved into transportation task realization process according to decisions of a management system. The task realization process may include many sequences of services, functions and operations which are using assignment network (system) resources. In the computer science this process of assignments and realization steps is called choreography and this term will be modified into transport choreography.

The management system MS organizes a work of the DTS system that means available system resources are assigned to executed tasks and maintenance policies are used to minimize system loses caused by some disturbances in the system operation. Dispatcher decisions are taken based on needs of assumed transportation tasks and according (if it is possible) to assumed proper time-tables. When some disruptions (failures, faults) occur the dispatcher chooses adequate system reactions.

The dispatcher is helped by computer tools that improve an allocation process of system resources to tasks of transport, creation time-tables of system traffic (planned and reserved for emergency conditions), and a design of maintenance policies prepared to use both for in normal and extraordinary situations of the transportation system work. The system dispatcher is supported by special computer tools and decisions of the dispatcher are made on the base of computer hints, dispatcher experiences and his management intelligence. It is possible to define many classes of dispatchers of the discrete transport
system. A passive dispatcher realizes transportation tasks agree to previously defined conditions and schedules. The passive dispatcher uses earlier prepared lists of assumed DTS disruptions and lists of planned adequate system reactions in case of disruptions. A task oriented dispatcher is focused on execution of a selected task or a defined set of tasks and such strategies as FIFO, LIFO, FILO, etc., are applied. A dynamic dispatcher is monitoring on-line a system and takes decisions adequate to system situation; of course, the dynamic dispatcher cannot work as a fantastic virtuoso manager. If more detailed supporting data (collected from different components of the system) have been prepared a priori, then evaluated dependability properties (performance and reliability parameters) of the considered DTS system will be closer to reality.

3.2. Client’s Model

The service realised by the clients of the transportation system are sending mails from some source node to some destination one. Client’s model consists of a set of clients; each client is allocated in one of nodes of the transportation system. A client allocated in an ordinary node generates containers (since, we have decided to monitor containers not separate mails during simulation) according to the Poisson process with destination address set to ordinary nodes. In the central node, there is a set of clients, one for each ordinary node. Each client generates containers by a separate Poisson process and is described by intensity of container generation.

3.3. Driver’s Model

The human infrastructure is composed by the set of drivers. So the description of this part of system infrastructure requires the analysis of the drivers’ state and the algorithms, which model the rules of their work. Each driver could be in one of following states (sd): rest (not at work), unavailable (illness, vacation, etc.), available (at work – ready to start driving), break (during driving), driving.

The number of driver working hours is limited by the labour law. For analysed Post Transportation System the daily limit for each driver equals to 8 hours and a single driver operates with one truck. It gives a simple algorithm:

\[
\text{if } w_h > \text{limit} \text{, then state = “rest” & } w_h = 0,
\]

where
\[
w_h – \text{working hours, limit } = 8 \text{ hours.}
\]

Drivers in Polish Post work in two shifts, morning or afternoon one. So twice a day a driver’s state and shift type is analysed:

- at 6am for each driver:
  - if shift == morning & sd == rest then sd = available,
- at 1pm for each driver:
  - if shift == afternoon & sd == rest then sd = available.

The next problem ought to be modelled is the driver’s illness state. We propose the following approach:

- for every driver at 4 am:
  - if sd == rest and rand() < d_i then during x days (according to a given distribution) the driver is in sd = unavailable,

where:
\[
d_i – \text{driver’s illness parameter}.
\]

Moreover we propose to categorise the driver’s illnesses as follows: short sick: 1 to 3 days, typical illness: 7 to 10 days, long-term illness: 10 to 300 days [21]. We prepare the daily record of the driver. The algorithm to fix the driver to the vehicle is the last part of the driver’s model:

- if no driver – the vehicle does not start,
- driver can be chosen if: sd == available and \( w_h + \text{estimated time of journey} < \text{limit * 1.1}, \)
- the driver is chosen randomly or by least value approach: \( \text{abs(limit – w_h – estimated time of journey).} \)
3.4. Management System

The decisions (send a truck to a given destination node) are taken in moments when a container arrives to the central node. The truck is sent to a trip if:

- the number of containers waiting in for delivery in the central node of the same destination address as that just arrived is larger than a given number,
- there is at least one available vehicle,
- the simulated time is between 6 am and 22 pm minus the average time of going to and returning from the destination node.

The truck is sent to a node defined by destination address of just arrived container. If there is more than one vehicle available in the central node, the vehicle with size that fits the best to the number of available containers is selected, i.e., the largest vehicle that could be fully loaded. If there are several trucks with the same capacity available the selection is done randomly. The restriction for the time of truck scheduling (the last point in the above algorithm) are set to model the fact that drivers are working on two 8 hours shifts.

4. Critical Situation Discussion

The dependability of the system can be defined as the ability to execute the functions (tasks, jobs) correctly, in the anticipated time, in the assumed work conditions, and in the presence of threats, technological resources failures, information resources and human faults (mainly malfunctions). Contemporary systems, such as transportation systems, are very often considered as networks of services. The system’s dependability can be described by such attributes as availability (readiness for correct service), reliability (continuity of correct service), safety (absence of catastrophic consequences on the users and the environment), security (availability of the system only for authorized users), confidentiality (absence of unauthorized disclosure of information), integrity (absence of improper system state alterations) and maintainability (ability to undergo repairs and modifications). The system realises some tasks and it is assumed that the main system goal, taken into consideration during design and operation, is to fulfil the user’s requirements. The system’s functionalities (services) and the technical resources are engaged for task realisation. Each task needs a fixed list of services, which are processed based on the system’s technological infrastructure. The different services may be realised using the same technical resources and the same services may be realised involving different sets of the technical resources. It is easy to understand that the different values of performance and reliability parameters should be taken into account. The final statement is essential when tasks are realised in the real system surrounded by unfriendly environment that may be a source of threats and even intentional attacks. Moreover, the real systems are built on the base of unreliable technical infrastructures and components. The modern systems are equipped with suitable measures and probes, which minimise the negative effects of these inefficiencies (a check-diagnostic complex, fault recovery, information renewal, time and hardware redundancy, reconfiguration or graceful degradation, restart, etc).

The special service resources (service persons, different redundancy devices, etc.) supported by maintenance policies (procedures of the service resources using in purpose to minimise negative consequences of faults that are prepared before or created ad hoc by the system manager) are build in every real system [3, 12, 13, 16]. The maintenance policy is based on two main concepts: detection of unfriendly events and system responses (system reactions) to them.

In general, the system responses incorporate the following procedures:

- detection and identification of incidents,
- isolation of damaged resources in order to limit proliferation of incident consequences,
- renewal of damaged processes and resources.

Different events of discrete transportation system operations are considered:

- “normal” functional events described by such time parameters as the start or / and the end of the task, a moment of a system resources allocation, a time of occurrence of a new task, an (prognoses or real) task execution time, etc.,
- unfriendly incidents that are disturbed efficient system execution; for example failures of transport means, accidents on roads, delay time of reloading cars, faults of workers or the system dispatcher, etc.

It is easy to notice that the first class of system events is strictly connected with correct system task realization and the second one groups all events disrupting the efficient operation of the system and which may start the system defense reactions. In this way the first class of events will be called “efficient
functional events” and the second one “dependable incidents” or “unfriendly events”. The dependable incident is an event that might lead to some disruptions in the system behavior. The incident may cause some damage to the system resources; transport means (hardware), management actions, employees or information and, in consequence, it may disrupt the executed transportation processes. If a fault appears during the task execution then the system on the base of decision of its management system (its dispatcher) starts renewal processes. Time of technological renewal activities are added to the nominal time of the task so a real time of the task duration will be longer. The real duration time of the executed tasks depends on the nature of the system faults. Failures of hardware may need both renewals of technological resources and information resources. Consequences of human errors or computer software faults are limited to renewals of information processes. Sometimes an incident – road accident or a series of truck failures which are occurred in a short time interval – may have a more serious impact on the system behavior; it may escalate to a security incident, a crisis or a catastrophe. The failures of the transport resources – physical failures of vehicles or roads or reloading devices need to use adequate such the DTS means as service teams, garages, spare elements or substituted routes. Very often “technical” system renewal processes are considered with assuming of the limited resources, for example, one service team for five vehicles.

Other sources of the DTS disruptions we can find in organization and management:
- overloading of the technical infrastructure: roads, reloading machines, etc.,
- traffic accidents or jams - considered as human errors,
- dispatcher faults – he or she is not able to keep up the dynamic changes of the situation in the working DTS system.

In these cases exploitation system renewal processes are initiated by the system dispatcher. The processes very often consume more time and money than a renewal of a “simple (physical)” broken technical resource, e.g., a repair of a failed truck or a lift. Of course, the correctly tuned system ought to be characterised by the set of fault tolerant features. It means – the system is able to realise the loaded tasks even if the different faults occur because of reliability or functional insufficiencies. The problem pointed above needs a multi-criteria solution. In other words, it is possible to find a kind of Pareto set (Fig. 1) to guarantee the system as functional ready for the defined tasks. The critical situation occurs if actual operating point of the system is located outside of the Pareto set. The main goal if the critical situation is noticed is to drive the system to the Pareto set as soon as possible. The proper management reaction is the first option to rescue the system’s situation.

![Figure 1. General idea of the Pareto set]
5. Availability Measure

One can define the system’s availability in different ways, but always the value of availability can be easily transformed into economic or functional parameters perfectly understood by owner of the system. The availability is mostly understood as a probability that a system is up and is defined as a ratio of the expected value of the uptime of a system to the observation time. It is a simple definition still it requires defining what does it mean that transportation system is working.

In paper [20] we have proposed the definition of up time as a time when the number of delayed containers does not exceed a given threshold. Let introduce the following notation:

- $T$ – a time measured from the moment when the container was introduced to the system to the moment when the container was transferred to the destination (random value).
- $T_g$ – a guaranteed time of delivery, if exceeded the container is delayed.
- $N_{delayed}(t)$ – a stochastic process describing the number of delayed containers at time $t$, i.e., the number of containers for which $T > T_g$.

Therefore, the functional availability $A_k(t)$ can be defined as a probability that the number of delayed containers at time $t$ does not exceed $k$, the value $k$ is the level of acceptable delay:

$$FA_k(t) = \Pr\{N_{delayed}(t) \leq k\}.$$  \hspace{1cm} (2)

Next, we want to analyse the system’s performance against the risk of periodic shortage of drivers and/or trucks or enlarge in number of containers to be transported. The system is forced to a state of reduced operability for a fixed period of time.

For the purpose of such analysis we have to redefine system’s availability defined by (2). Let’s consider a 24 hour time period for determining the availability. Then, the sequence of time instances $(\tau_0, \tau_1, \ldots, \tau_n)$ will fix the boundaries of the consecutive days, for which the metric is assessed. $N_d(\tau_i, \tau_{i+1})$ denotes the number of containers delivered in the period $(\tau_i, \tau_{i+1})$. $N_{pd}(\tau_i, \tau_{i+1})$ denotes the number of delivered containers on time. Therefore, the system’s availability could be measured by an average ratio of on-time deliveries, defined as:

$$A_s(\tau_i) = E\left(\frac{N_{pd}(\tau_{i-1}, \tau_i)}{N_d(\tau_{i-1}, \tau_i) + 1}\right),$$ \hspace{1cm} (3)

where $E$ denotes the average value (and is determined from the simulator results by evaluating the mean value of multiple runs).

The denominator includes +1 modification to prevent the ratio to go to infinity in case of a full stoppage of the system (i.e., no containers delivered in the analysed period).

6. Critical Situation Analysis

6.1. Exemplar System

We propose for the case study analysis an exemplar DTS based on Polish Post regional centre in Wroclaw. We have modelled a system consisting of one central node (Wroclaw regional centre) and twenty two other nodes – cities where there are local post distribution points in Dolny Slask Province. The length of roads was set according to real road distances between cities used in the analyzed case study. The intensity of generation of containers for all destinations was set to 4.16 per hour in each direction giving in average 4400 containers to be transported each day. The vehicles speed was modelled by Gaussian distribution with 50 km/h of mean value and 5 km/h of standard deviation. The average loading time was equal to 5 minutes. There were two types of vehicles: with capacity of 10 and 15 containers. The MTF of each vehicle was set to 20000. The average repair time was set to 5h (Gaussian distribution). We also have tried to model the driver’s availability parameters. We have fulfilled this challenge by using the following probability of a given type of sickness – short sick: 0.003, typical illness: 0.001, long-term illness: 0.00025.
6.2. Critical Situation Feedback

As it has been mentioned in the Chapter 4 the system is designed to work at a given level of availability by including some redundancy to system resources (mainly trucks and drivers). It results in a situation that at night (when the system is non-operational due to the free time for drivers) there are almost no delayed containers. However, in case of unpredictable and rare situations like shortage of drivers (for example, due to some contagious diseases or a strike) or short enlargement in a number of containers to be transported, number of delayed containers starts to cumulate in the system. In other words, the system is not able to transport all containers.

The only solution to such critical situation is to enlarge system resources, i.e., hire more drivers and trucks for a short time. The question is how many additional resources should be added to the system to eliminate the critical situation and for how long time. The consequences of the critical situation are said to be eliminated when the daily ratio of on-time deliveries reaches a predefined level \( \alpha \) (say 0.995).

The designed by authors simulator of DTS [17, 20] allows us to help in taking some management decisions.

The Figure 2 presents the system’s availability (average ratio of on-time deliveries) after three days of operation for various numbers of vehicles and drivers and containers delayed in the system.

The system manager knowing the number of delayed containers presented in the system selects one of presented figures from Figure 2, next he selects the number of vehicles and drivers that allow eliminating the critical situation after three days (by selecting points for which the availability is larger than the predefined level).

Knowing the actual number of drivers and trucks she/he could estimate the number of resources to be additionally hired. If longer time for eliminating the critical situations is allowed – 7 days, for example, – the manager could use simulation results presented on Figure 3.

Figure 2. The system’s availability (average ratio of on-time deliveries) for various numbers of vehicles and drivers and containers after three days of operation
7. Conclusions

We have presented a formal model of discrete transportation system (DTS) including reliability, functional parameters as well as the human factor component. The DTS model is based on Polish Post regional transportation system and reflects all key elements of it with the set of the most important functional and reliability features of them. The critical situation is pointed out and described at the necessary level of details for the further analysis. The proposed availability metric is the source to create the Pareto set – guarantying the possible safety operating points for actual DTS.

The realised analysis based on the real data – taken from the Polish Post transportation system at Wroclaw area – allowed testing the effectiveness of the system management when the critical situation occurs. We introduced the exemplar DTS into critical situation to discuss how deep breakdown it has caused and to test, which management approach can drive the system operating point into Pareto set as soon as possible. The proposed approach allows performing deeper reliability and functional analysis of the DTS, for example:

- to determine what will cause a "local" change in the system,
- to make experiments in case of increasing number of containers per day incoming to system,
- to identify weak point of the system by comparing few its configuration,
- to better understand how the system behaves in ordinary and critical situations.

The solution presented can be used as practical tool for defining an organization of vehicle maintenance and transportation system’s logistics.

![Figure 3. The system’s availability (average ratio of on-time deliveries) for various numbers of vehicles and drivers and containers after seven days of operation](image-url)
References


