Non-Markovian Performability Evaluation of ERTMS/ETCS Level 3

Laura Carnevali[1], Francesco Flammini[2], Marco Paolieri[1(✉)], and Enrico Vicario[1]

[1] Department of Information Engineering, University of Florence, Florence, Italy
{laura.carnevali,marco.paolieri,enrico.vicario}@unifi.it
[2] Ansaldo STS, Naples, Italy
francesco.flammini@ansaldo-sts.com

Abstract. The European Rail Traffic Management System/European Train Control System (ERTMS/ETCS) is an innovative standard introduced to enhance reliability, safety, performance, and interoperability of trans-European railways. In Level 3, the standard replaces fixed-block safety mechanisms, in which only one train at a time is allowed to be in each railway block, with *moving blocks*: a train proceeds as long as it receives radio messages ensuring that the track ahead is clear of other trains. This mechanism increases line capacity, but relies crucially on the communication link: if messages are lost, the train must stop within a safe deadline even if the track ahead is clear. We develop upon results of the literature to propose an approach for the evaluation of *transient* availability of the communication channel and probability of train stops due to lost messages. We formulate a non-Markovian model of communication availability and system operation, and leverage solution techniques of the ORIS Tool to provide experimental results in the presence of multiple concurrent activities with non-exponential durations.

Keywords: European Rail Traffic Management System (ERTMS) · European Train Control System (ETCS) · Real-time systems design · Markov Regenerative Process (MRP) · Transient analysis · Stochastic state classes

1 Introduction

Quantitative evaluation of models with stochastic timers often provides crucial support in the engineering of dependability requirements. Both analytic and simulative approaches can serve the objective, with different limitations and advantages. In particular, when applicable, analytic approaches facilitate the exploration of the space of preliminary design, especially in the presence of rare events. The limits for applicability are determined by various factors, and notably by the class of the underlying stochastic process of the model [10].

If all activity durations are distributed according to (memoryless) exponential distributions, the underlying stochastic process is a continuous-time Markov

M. Beltrán et al. (Eds.): EPEW 2015, LNCS 9272, pp. 47–62, 2015.
DOI: 10.1007/978-3-319-23267-6_4

chain and evaluation can resort to consolidated and efficient analytic approaches [11,23,26,27]. However, the system under analysis is sometimes strongly characterized by activity durations that are deterministic (e.g., timeouts) or distributed according to general (i.e., nonexponential) distributions, imposing hard constraints on the minimum or maximum value. In this case, the underlying stochastic process is non-Markovian, but it can still satisfy the Markov property (conditional independence of future evolution from past history, given the current state) at selected time instants called *regeneration points*. In Markov Regenerative Processes (MRPs) [20], a new regeneration is eventually reached with probability 1, and the analysis can be formulated in terms of a local and a global kernel that characterize the behavior of the process between subsequent regeneration points. Solutions for the evaluation of kernels have been consolidated only for models satisfying the so-called *enabling restriction*, which requires that at most one generally distributed transition is enabled in each state [5,9,10,21]. Recent results based on the method of stochastic state classes have overcome the limit [18,32], enabling the numerical solution of models beyond the enabling restriction, and in particular MRPs reaching regenerations in a bounded number of state transitions [18]. The ORIS Tool provides an implementation of the approach [8], opening the way to the analysis of a large class of problems.

Level 3 is the most promising operation level of the European Rail Traffic Management System/European Train Control System (ERTMS/ETCS) [15,16] in terms of capacity gains and trackside installation savings, and it represents a challenging case study in the engineering of future train control systems. As in Communication Based Train Control (CBTC) for metro-railways, the ERTMS/ETCS Level 3 standard adopts a radio-based moving-block technology, where the maximum distance before a virtual stop is computed in real-time from locations and speeds of all trains, requiring continuous two-way communication between each train and the control center.

The case study has been widely addressed in the literature of quantitative evaluation of dependability [1,2,13,14,17,22,30,35]. Notably, in [30,35], probabilistic parameters were derived from the analysis of the standard specification and cast into a hierarchical modeling and evaluation approach, based on rare events simulation and analysis of non-Markovian models under enabling restriction, both supported by the TimeNET Tool [33,34]. Reliability analysis is addressed in [17] by leveraging the MODEST language [6] and the Möbius Tool [12]. Reliability aspects are also assessed in [14] by means of a compositional approach that integrates analysis of fault trees and evaluation of Bayesian networks. Dependability and safety metrics are evaluated in [2] focusing on the parameters that affect the probability of an emergency train stop. A multi-formalism model is used in [13] to evaluate the influence of basic design parameters on the probability of system-level failure modes, showing that system availability is lower than the threshold prescribed by the specification. In [22], the ERTMS/ETCS Level 2 railway signaling system (using radio communication but not moving blocks) is modeled as a system of systems and its dependability parameters are evaluated using statecharts, taking into account human factors, network failures, and imprecise failure

Acronym	*Meaning*
BTS	Base Transceiver Station
DET	Transition with deterministic duration
ERTMS	European Rail Traffic Management System
ETCS	European Train Control System
EVC	European Vital Computer
EXP	Exponentially distributed transition
GEN	Transition with general (i.e., nonexponential) distribution
IMM	Immediate transition
MA	Movement Authority
MRP	Markov Regenerative Process
PR	Position Report
RBC	Radio Block Center
sTPN	Stochastic Time Petri Net

Fig. 1. Summary of acronyms used in the rest of the paper.

and repair rates. Reverse engineering is used in [1] to perform static analysis of the software of a complex safety-critical subsystem of the ERTMS/ETCS, supporting both correctness verification of software and its refactoring.

In this paper, we develop upon the results of [30,35] to propose a model of communication availability including multiple concurrent activities with generally distributed durations. The model accounts for the concurrent nature of communication failures due to handovers between neighboring radio stations, and random burst noise or connection losses. We provide a safe approximation of the *transient* availability of the communication layer, and leverage this measure in a higher-level operational model of moving-block signaling, in which a train proceeds as long as it receives messages ensuring that the track ahead is clear of other trains. Through a first-passage analysis of this model, we compute the *transient* probability that the train has stopped due to lost messages, as opposed to previous work focusing on steady state analysis [30,35]. Since the "arrive and depart" mechanism of trains is inherently transient, the results provide a further step in the analysis of the effects of communication failures on moving-block signaling. The evaluation leverages the analysis of MRPs based on stochastic state classes [18] and its implementation within the ORIS Tool [8].

The rest of the paper is organized in four sections. In Sect. 2, we examine the ERTMS/ETCS case study, with specific focus on the Level 3 implementation. In Sect. 3, we recall syntax and semantics of a non-Markovian variant of stochastic Petri nets [32] and the salient traits of regenerative analysis [18]. In Sect. 4, we present a non-Markovian model of communication availability and derive a safe approximation that is used, in turn, to compute transient performability measures on the operational model based on moving blocks. Conclusions are drawn in Sect. 5.

2 The ERTMS/ETCS L3 Case-Study

The ERTMS/ETCS is a recent standard that has been developed to enhance performance, reliability and safety of trans-European railway networks. In fact, the standard has been an intercontinental success, so that ERTMS/ETCS compliant railways have been or are being engineered in several installations even outside Europe (e.g., China, United Arab Emirates). Though actual systems can be very complex, heterogeneous and highly distributed, the working principles of ERTMS/ETCS are rather straightforward. Trains are equipped with on-board automatic train control devices, which are embedded real-time computers known as European Vital Computers (EVCs). EVCs are connected with train-borne apparels (e.g., odometer, brakes) to allow automatic braking in case the speed is over the allowed limit. To compute the maximum allowed speed (known as the *braking curve* or *dynamic speed profile*), the EVC needs to receive the following information from the trackside subsystems: *i*) Movement Authority (MA), i.e., the maximum distance before a virtual stop signal; *ii*) Static speed profile, i.e., the maximum speed depending on track morphology; *iii*) Possible temporary speed restrictions or conditional/unconditional emergency stops. Such information can be provided to the EVCs using different communication means in the three levels of operation defined by the standard. At Level 2 and 3, the so-called Radio Block Centers (RBCs) are employed, enabling continuous radio-signalling using GSM-R (similar to the well-known mobile phone standard) and the safe Euroradio protocol. In turn, the RBC needs to know the Position Reports (PRs), that is the precise position of all the trains on the track. The EVC obtains this information by reading "telegrams" sent by the so-called *balises*, which are devices installed between the track lines and acting like milestones. PRs are sent by the EVC to the RBC either periodically, at each newly encountered balise, or upon specific RBC requests.

Most of the lines that are currently operational, starting from the first Rome-Naples Italian high-speed railway, are based on the ERTMS/ETCS Level 2, which employs fixed-block signaling. That means the MA is computed by summing track circuits and routes that are neither occupied by any train nor in out-of-service/exclusion conditions. Such an implementation needs an interface between the RBC and the underlying (possibly legacy) interlocking systems. The ERTMS/ETCS Level 2 is generally considered safer at the expense of longer headways due to the obviously less fine-grained spacing.

To increase line capacity, Level 3 introduces *moving-block signaling*: the MA of the chasing train is computed considering the minimum safe rear-end of the foregoing train, and not the status of track-circuits. In those conditions, it is rather intuitive that system safety is highly dependent on train integrity checks, hence the EVC has to provide this additional information to the RBC. Moving-block signaling has received higher attention in mass-transit (e.g., subways), due to the required high-frequency of trains (few minutes waiting times), and it is adopted in Communication Based Train Control (CBTC) for metro-railways. To justify the adoption of Level 3 for new high-capacity railways, where the braking distances and data latencies are essential factors to take into account,

it is very important to preliminarily evaluate the real expected performance by model-based analysis. Such an analysis can also assess which are the variables having a higher impact on system performance. Actually, since both performance (computing latencies, communication delays) and reliability (data transmission errors, connection faults) aspects need to be addressed, this kind of assessment comprises a classical problem of performability evaluation.

In [35,36], the maximum delay d (i.e., deadline) after which automatic braking is activated by the on-board system is derived for the condition of a train that runs at speed $v = 300\,\mathrm{km\,h^{-1}}$, as usual in most real installations, and it is expressed as a function of the following factors: *i*) the train headway s, i.e., the distance between the maximum safe front-end of the train and the minimum safe rear-end of its predecessor, with those positions corrected taking into account the estimated odometric measurement errors; *ii*) the braking distance s_{brake}, assumed to be approximately 3 km including the aforementioned position errors; *iii*) the packet age p_{age}, i.e., the maximum staleness of a received packet (p_{age} is assumed to be 12 s in the worst case). More specifically, in [35,36] it is shown that $d = (s - s_{braking})/v - p_{age}$. Based on these assumptions, it is evinced that headways cannot be shorter than 4 km, that is the theoretical minimum. In such a scenario, model-based analysis is essential to evaluate the train stop probability as a function of the required headways, or, conversely, the minimum headways allowing acceptable system availability measures. Also, sensitivity to other parameters can be evaluated in order to support system design choices in industrial and operational settings.

3 Non-Markovian Modeling and Analysis

The system is modeled using a variant of non-Markovian stochastic Petri nets called *stochastic Time Petri Nets* (sTPN) [32], enriched with enabling functions, flush functions, and priorities, augmenting the modeling convenience without impacting on the nature and complexity of the analysis, as in [25,28].

3.1 Stochastic Time Petri Nets

Syntax. An sTPN is a tuple $\langle P; T; A^-; A^+; A^\bullet; m_0; EFT^s; LFT^s; \mathcal{F}; \mathcal{C}; E; L; R\rangle$, where: P is a set of places; T is a set of transitions; $A^- \subseteq P \times T$, $A^+ \subseteq T \times P$, and $A^\bullet \subseteq P \times T$ are the sets of precondition, postcondition, and inhibitor arcs, respectively; $m_0 : P \to \mathbb{N}$ is the initial marking associating each place with a non-negative number of tokens; $EFT^s : T \to \mathbb{Q}_0^+$ and $LFT^s : T \to \mathbb{Q}_0^+ \cup \{\infty\}$ associate each transition with a *static Earliest Firing Time* and a static *Latest Firing Time*, respectively, such that $EFT^s(t) \leq LFT^s(t)\ \forall\ t \in T$; $\mathcal{F} : T \to F_t^s$ associates each transition with a static Cumulative Distribution Function (CDF) such that $x < EFT^s(t) \Rightarrow F_t^s(x) = 0$ and $x > LFT^s(t) \Rightarrow F_t^s(x) = 1$; $\mathcal{C} : T \to \mathbb{R}^+$ associates each transition with a weight; $E : T \to \{true, false\}^{\mathbb{N}^P}$ associates each transition t with an *enabling function* $E(t) : \mathbb{N}^P \to \{true, false\}$ that, in turn, associates each marking with a boolean value; $L : T \to \mathcal{P}(P)$

is a *flush function* associating each transition with a subset of P; $R : T \to \mathbb{N}$ associates each transition with a priority. A place p is called an *input*, an *output*, or an *inhibitor* place for a transition t if $\langle p, t\rangle \in A^-$, $\langle t, p\rangle \in A^+$, or $\langle p, t\rangle \in A^\bullet$, respectively. A transition t is called *immediate* (IMM) if $[EFT^s(t), LFT^s(t)] = [0, 0]$ and *timed* otherwise. A timed transition t is called *exponential* (EXP) if $F_t^s(x) = 1 - e^{\lambda x}$ over $[0, \infty]$ for some $\lambda \in \mathbb{R}_0^+$ and *general* (GEN) otherwise. A GEN transition t is called *deterministic* (DET) if $EFT^s(t) = LFT^s(t) > 0$ and *distributed* otherwise (i.e., if $EFT^s(t) \neq LFT^s(t)$). For each distributed transition t, we assume that F_t^s is absolutely continuous over $[EFT^s(t), LFT^s(t)]$ and, thus, that there exists a Probability Density Function (PDF) f_t^s such that $F_t^s(x) = \int_0^x f_t^s(y)dy$.

Semantics. The *state* of an sTPN is a pair $\langle m, \tau\rangle$, where $m : P \to \mathbb{N}$ is a marking that associates each place with a non-negative number of tokens and $\tau : T \to \mathbb{R}_0^+$ associates each transition with a (dynamic) real-valued time-to-fire. A transition is *enabled* by a marking if each of its input places contains at least one token, none of its inhibitor places contains any token, and its enabling function evaluates to true. An enabled transition t is *firable* if its time-to-fire is not higher than that of any other enabled transition and, in case t is IMM or DET, if its priority is not lower than that of any other enabled IMM/DET transition. When multiple transitions are firable, one of them is selected as the firing transition with probability $Prob\{t \text{ is selected}\} = \mathcal{C}(t)/\sum_{t_i \in T^f(s)} \mathcal{C}(t_i)$, where $T^f(s)$ is the set of transitions that are firable in s. When a transition t fires, the state $s = \langle m, \tau\rangle$ is replaced by a new state $s' = \langle m', \tau'\rangle$. Marking m' is derived from m by: *i)* removing a token from each input place of t and removing all tokens from the places in $L(t) \subseteq P$, which yields an intermediate marking m_{tmp}, *ii)* adding a token to each output place of t. Transitions that are enabled both by m_{tmp} and by m' are called *persistent*, while those that are enabled by m' but not by m_{tmp} or m are called *newly-enabled*. If the fired transition t is still enabled after its own firing, it is always regarded as newly enabled [4,31]. While the time-to-fire of persistent transitions is reduced by the time elapsed in s, that of newly-enabled transitions takes a random value sampled according to the static CDF.

3.2 Regenerative Transient Analysis Through Stochastic State Classes

The method of stochastic state classes [7,32] faces the analysis of the underlying stochastic process of models with multiple concurrent GEN transitions. To this end, the marking and the vector of times to fire of GEN transitions are characterized after each firing. This yields an embedded discrete time Markov chain encoded in a so-called *stochastic graph*, whose states are called *stochastic state classes*. Each class is made of a marking plus the joint support and PDF of the times-to-fire of enabled GEN transitions. To support transient evaluation, in [18] a fresh clock named τ_{age} is added to each class to account for the absolute elapsed time. The marginal PDF of τ_{age} permits to derive the PDF of the absolute time at which a class can be entered, enabling the evaluation

of continuous-time transient probabilities of reachable markings within a given time horizon, provided that the number of classes that can be reached within that time interval is either bounded or can be truncated under the assumption of some approximation threshold on the total unallocated probability.

In general, the approach of [18] supports transient analysis of models with underlying Generalized Semi-Markov Process (GSMP) with equal-speed timers [10,24]. Nevertheless, the complexity of the solution technique can be significantly reduced when applied to models with underlying Markov Regenerative Process (MRP) that always reaches a regeneration point within a bounded number of steps, i.e., a state where the future behavior is independent from the past. In fact, transient analysis can be restrained within the first regeneration epoch from each regenerative point, and finalized to the derivation of the local and global kernels that characterize the behavior of the MRP [5,9,10]. Transient probabilities of reachable markings at any time can then be derived by numerical integration of generalized Markov renewal equations [20].

4 Performability Evaluation of ERTMS/ETCS Level 3

We consider a model of communication availability that features a non-Markovian representation of failures due to handovers (Sect. 4.1), and we derive a safe estimation of the transient evolution of communication availability through a 3-step function (Sect. 4.2). Such approximation is cast within a non-Markovian model of communication beyond the limits of the enabling restriction (Sect. 4.3), evaluating the transient probability that a timeout occurs within time t (Sect. 4.4).

4.1 A Non-Markovian Model of Communication Availability

At the ERTMS/ETCS Level 3, the GSM-R communication channel appears the most relevant source of unreliability, due to almost unavoidable data transmission errors, connection losses, and Base Transceiver Station (BTS) handovers. In [35,36], stochastic parameters characterizing communication failures are derived from specification documents and guidelines, evaluating the probability of stops through a combined use of analytic evaluation under enabling restriction and rare-event simulation, both supported within the TimeNET Tool [33].

Here we present a model of communication failure that develops upon the results of [35,36], leveraging the same stochastic parameters while extending the model structure to encompass a non-Markovian representation of handovers. The model is shown in Fig. 2. As in [35,36]: the arrival and duration of "bursts" of noise are modeled by the EXP transitions `startBurst` and `endBurst` with rate $0.007\,33\,\mathrm{s}^{-1}$ and $3\,\mathrm{s}^{-1}$, respectively, derived by fitting the specification that the mean time between two bursts is at least 7 s, with each burst shorter than 1 s in 95% of the cases; the occurrence of a connection loss is represented by the EXP transition `loss` with rate $2.77 \times 10^{-8}\,\mathrm{s}^{-1}$, derived from the specification that the probability to have a connection loss per hour is less than 10^{-4}; the

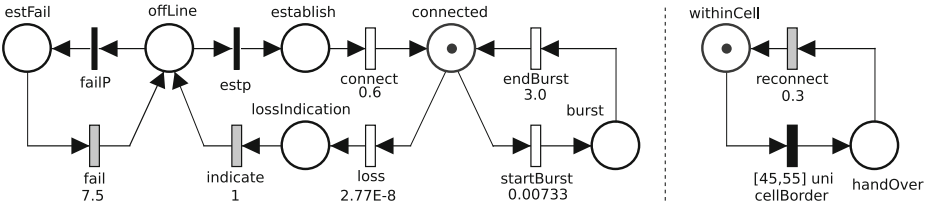

Fig. 2. The sTPN model of communication availability (times expressed in s), combining the sub-models of failures due to handovers (right) and transmission errors or connection losses (left). IMM, EXP, DET, and GEN transitions are represented by thin bars, thick empty bars, thick gray bars, and thick black bars, respectively.

time needed to detect a connection loss is required to be not greater than 1 s, thus it is accounted by the DET transition `indicate`; the reconnection attempt is required to be successful with a probability higher than 99.9%, which is represented as a switch between the IMM transitions `estP` and `failP`, having weight 0.999 and 0.001, respectively; in case of reconnection success, the establishment time must be less than 5 s in 95% of the cases, which is modeled by the EXP transition `connect` with rate $0.6\,\mathrm{s}^{-1}$; in case of reconnection failure, a reconnection is retried after 7.5 s, which is modeled by the DET transition `fail`.

In [35,36], connection loss due to handovers is modeled by an EXP transition conflicting with `startBurst` and `loss`, whose rate is derived as the inverse of the time spent by a train that runs at the maximum speed $v = 500\,\mathrm{km\,h}^{-1}$ to cover the 7 km distance between BTS, i.e, $1/0.0198 = 50.4\,\mathrm{s}^{-1}$. As opposed to [35,36], we model failures due to hand-overs and failures due to transmission errors or connection losses as concurrent events. Actually, this reflects the nature of the phenomenon, as handovers indeed occur in parallel to transmission errors and connection losses. According to this, in the model of Fig. 2, the sub-model that accounts for failures due to handovers (the right part) is concurrent to the sub-model that represents failures due to transmission errors and connection losses (the left part). The time between subsequent communication failures due to handovers is modeled by a GEN distribution with bounded support rather than with an EXP distribution over $[0, \infty)$. This captures the fact that the distance between subsequent BTS is nearly constant and the speed of trains ranges within a min-max interval. In the present experimentation, a uniform distribution supported over $[45, 55]$ s is associated with the GEN transition `cellBorder` in the model of Fig. 2. The mean value of such transition (i.e., 50 s) is a conservative approximation of the mean value of the namesake EXP transition in the model of [35,36] (i.e., 50.4 s). As in [35,36], the time to reconnection is modeled by the DET transition `reconnect`, whose duration equal to 0.3 s is the maximum time allowed by the requirements specification.

4.2 Evaluation of the Communication Availability Model

To reduce the stiffness of the problem due to failures that occur with different time-scales, we separate the analysis of independent events. According to this,

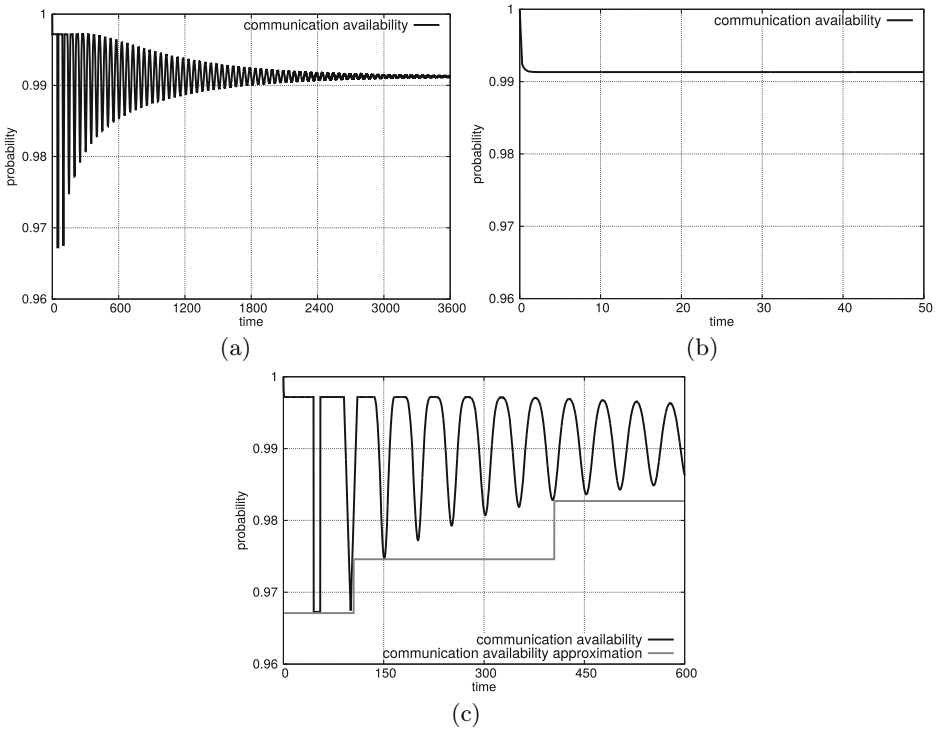

Fig. 3. Transient probability that the communication is available, derived through regenerative analysis of: (a) the model of Fig. 2 and (b) the communication availability sub-model of [35,36]. (c) A conservative approximation of the curve of Fig. 3-a for the time scale $[0, 600]$ s through a 3-step function. Times are expressed in s.

the two sub-models shown in Fig. 2 are separately analyzed using the ORIS Tool [8]. Overall, regenerative analysis of both sub-models is performed in nearly 15 min on a machine equipped with an Intel Xeon 2.67 GHz and 32 GB RAM, assuming time bound equal to 3600 s, approximation threshold equal to zero (i.e., exact analysis), and discretization step in the integration of renewal equations equal to 0.1 s. The analysis yields the transient probability that the communication is not available due to a transmission error or a connection loss, i.e., $p_c(t) = Prob\{\texttt{connected} == 1 \text{ at time } t\}$, and the transient probability that the train is not crossing the border between the communication areas of two neighboring BTS, i.e., $p_w(t) = Prob\{\texttt{withinCell} == 1 \text{ at time } t\}$. By multiplying the obtained numerical results, we derive the transient probability that the communication is available, i.e., $p(t) = p_c(t) \cdot p_w(t)$, whose plot is shown in Fig. 3-a. The plot shows an oscillating pattern, with ripples of decreasing heights, converging to a neighborhood of 0.9912 after a settling time of about 3000 s. This is mainly due to the floating trend of $p_w(t)$, which in fact has a settling time around 3000 s. Conversely, $p_c(t)$ exhibits an exponential trend with a much shorter settling time around 5 s.

Fig. 3-b shows the transient probability that the communication is available, derived through the analysis of communication availability sub-model of [35,36]. Also in this case, regenerative analysis is performed in nearly 20 min with time bound equal to 3600 s, approximation threshold equal to zero, and step equal to 0.1 s. While the curve shown in Fig. 3-b tends to approximately 0.9913, which is very close to the settling value of the curve shown in Fig. 3-a, the transient behavior of the two curves is significantly contrasting. As a notable difference, in the model under enabling restriction, the settling time is nearly 10 s and actually elapses by the time the first message is sent from the RBC to the following train. Conversely, in the model beyond enabling restriction, the settling time is much longer and the curve still exhibits ripples with height in the order of 10^{-4} after that time, until the time bound of 3600 s.

The transient probability $p(t)$ that the communication is available can be safely under-approximated by means of a monotone non-decreasing step function. Fig. 3-c shows the original curve of Fig. 3 for the time scale $[0, 600]$ s and an approximation by the following 3-step function:

$$f(t) = \begin{cases} 0.9671 & \text{if } 0 \leq t \leq 105, \\ 0.9746 & \text{if } 105 < t \leq 405, \\ 0.9827 & \text{if } 405 < t < \infty. \end{cases} \tag{1}$$

While a greater number of steps could provide a finer grained approximation, a 3-step function turns out to be sufficient for the purposes of the subsequent treatment. Note that the complexity of the subsequent analysis is substantially insensitive to the number of steps used in the approximation, and it only depends on the time at which the last jump of the step function is positioned, i.e., 405 s in the present experimentation. In fact, beyond that time instant, the estimate on communication availability is constant and does not carry memory over time, reducing by 1 the number of GEN transitions that are concurrently enabled.

4.3 A Non-Markovian Model of ERTMS/ETCS Level 3

Following the results of [35,36], the proposed approach resorts to the hierarchical composition of models, by relying on the assumption that the availability of communication is independent of the exchange of PR between track-side equipments and on-board devices. In so doing, the method also takes advantage of some approximations of model variables that are guaranteed to be stochastically ordered. As opposed to [35,36], the approach leverages a solution technique that goes beyond the limits of the enabling restriction. In the methodological perspective, this largely relaxes modeling restrictions, as the requirement that the underlying stochastic process is a Markov regenerative process poses less constraints on the model expressivity than the limitation on the number of concurrent GEN timers. In the applicative perspective, this permits to refine the models presented in [35,36] and the way they are composed through a more accurate representation of communication failures due to hand-overs. As a major result, solution

can be attained through a fully analytic treatment without facing complexities and limits of simulation in the presence of rare events.

The overall model of communication failure is shown in Fig. 4. As discussed in [35,36], at ERTMS/ETCS Level 3, train integrity checks are performed in 5 s in order to maximize track throughput, and the results are sent together with the PR (transition `genMsg`). RBC processing time for PR is assumed to be 0.5 s at most, while message transmission delays up-link and down-link are required to be: between 0.4 s and 0.5 s on average, less than 0.5 s in 95% of the cases, less than 1.2 s in 99% of the cases, and less than 2.4 s in 99.99% of the cases. RBC processing and up-link transmission delays are accounted by the GEN transition `transmitUp`, while down-link transmission delays are represented by the GEN transition `transmitDown`; in the present experimentation, `transmitUp` and `transmitDown` have a uniform distribution over their respective support. Whenever the deadline on the time between subsequent messages is violated (transition `timeout`), the train starts to brake until it comes to a complete stop: in this case, the resetup/restart delay (not considered in the model of Fig. 4) is assumed to be 15 min long, with all MA dropped during this time.

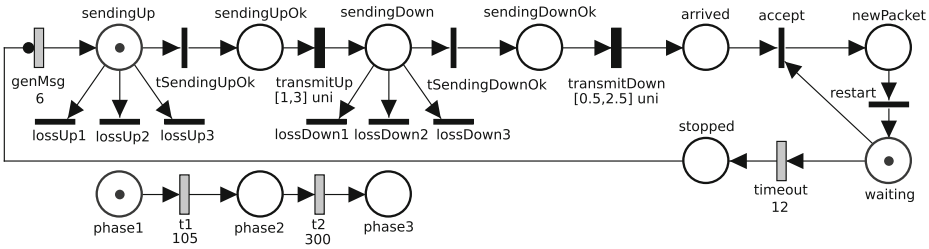

Fig. 4. The sTPN model of the ERTMS/ETCS Level 3 case study beyond the limits of the enabling restriction (times expressed in s). IMM, DET, and GEN transitions are represented by thin bars, thick gray bars, and thick black bars, respectively.

To manage the complexity of the analysis, in [35,36], the sub-model of communication availability is analyzed in isolation and its results are used to derive the rates of a condensed birth-death process made of 2-states, which is then recast in the overall model of communication failure. Yet, such model is not amenable to analysis with methods operating within the limits of the enabling restriction, and the evaluation is thus performed through rare-event simulation [19] supported by the TimeNET Tool, deriving the probability that the train is stopped for different values of the packet age and the head-to-head distance between trains. Conversely, in the proposed approach, the approximation of communication availability is recast within the overall system by means of a phased sub-model with phases of deterministic duration, so that, in each phase, failures of the communication up-link and down-link are accounted by a switch between IMM transitions, whose weights reflect the value of the approximating step function in the corresponding time interval. More specifically, in the model

of the overall system shown in Fig. 4: the IMM transitions `lossUp1`, `lossUp2`, and `lossUp3` represent failures of the communication up-link, while the IMM transition `tSendingUpOk` accounts for its availability; `lossUp1`, `lossUp2`, and `lossUp3` have an enabling condition that evaluates to true only during the corresponding phase (i.e., `phase1` == 1, `phase2` == 1, and `phase3` == 1, respectively); their weights are set equal to 0.03402, 0.02607, and 0.01761, respectively, while the weight of `tSendingUpOk` is maintained equal to 1, so that the probability that the communication is available turns out to be equal to 0.9671, 0.9746, and 0.9827 in phase 1, phase 2, and phase 3, respectively, as defined in Eq. 1. Similarly, the IMM transitions `lossDown1`, `lossDown2` and `lossDown3` model failures of the communication down-link, while the IMM transition `tSendingDownOk` represents its availability; they have the same enabling condition and weight of `lossUp1`, `lossUp2`, `lossUp3`, and `tSendingDownOk`, respectively.

Transition `restart` is associated with an enabling function that evaluates to true if `sendingUp` contains a token, and it has higher priority than `lossUp1`, `lossUp2`, `lossUp3`, and `tSendingUpOk`. This guarantees that, whenever the timeout fires, the last received packet has an age equal to (12+timeout) s. Moreover, since we evaluate the transient probability that a timeout occurs within time t (i.e., the transient probability of the first token arrival in place `stopped`), an inhibitor arc is added from place `stopped` to transition `genMsg`, and transition `timeout` is associated with a flush function that removes any token in any place, except for place `stopped`, upon its firing.

4.4 Evaluation of the ERTMS/ETCS Level 3 Model

Regenerative analysis of the model of Fig. 4 is performed in nearly 1 min with time bound 3600 s, time step 1 s, and approximation threshold equal to zero. The analysis is repeated for different values of the DET transitions `genMsg` (i.e., 6 s, 8 s, and 10 s) and `timeout` (i.e., 12 s, 15 s, and 18 s), and the obtained results are shown in Fig. 5. Such values of `genMsg` are selected based on the requirement that the time between two subsequent PR is $\geq$ 5 s. The values of `timeout` are thereby chosen with the purpose of showing the variability of the studied reward.

For an assigned value of `genMsg`, the probability that the train is stopped within time t decreases as the timeout increases. In fact, in Figs. 5-a, 5-b, and 5-c, the black curve dominates the gray curve which, in turn, dominates the light gray curve. Nevertheless, the gap between the curves may significantly vary among cases. For instance, in Fig. 5-a (i.e., `genMsg` = 6 s), the stop probability is nearly equal to 0.2018, 0.1514, and 0.0114 at time 600 s, and nearly equal to 0.5486, 0.4396, and 0.0305 at time 3600 s, for *timeout* equal to 12 s, 15 s, and 18 s, respectively. Conversely, in Fig. 5-b (i.e., `genMsg` = 8 s), the stop probability is approximately equal to 0.6654 at time 600 s and reaches nearly 0.9930 at time 3600 s for *timeout* equal to 12 s, while it has substantially the same trend for *timeout* equal to 15 s and 18 s (with a difference in the order of 10^{-2} at time 3600 s). Overall, this motivates the opportunity of a sensitivity analysis to assess the considered reward depending on the system parameters.

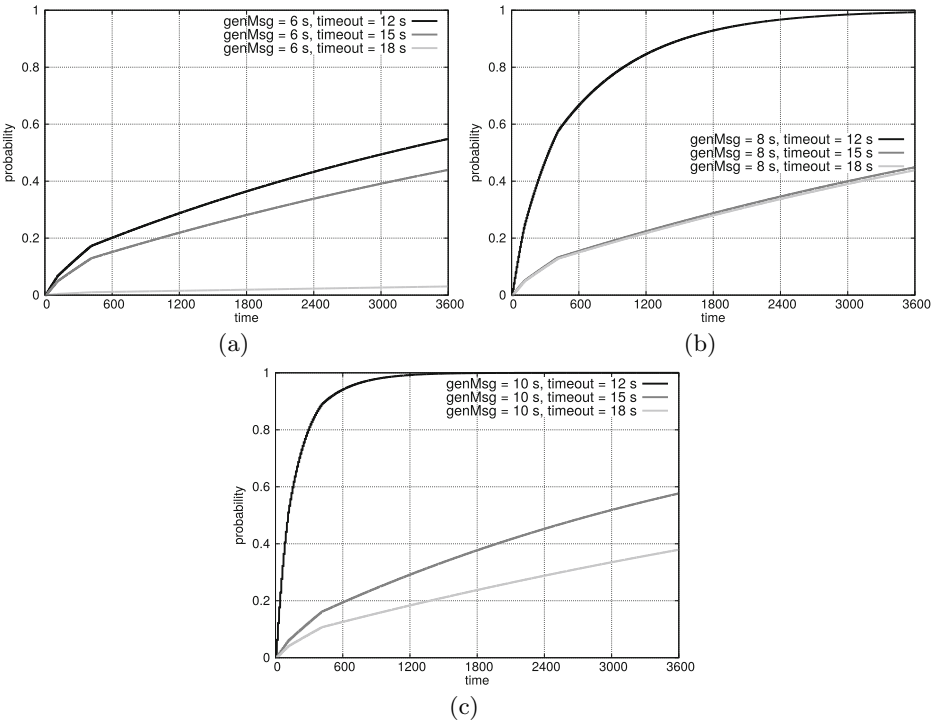

Fig. 5. Transient probability that a timeout occurs within time t (expressed in s).

5 Conclusions

We address performability modeling and evaluation of the ERTMS/ETCS Level 3 case study, supporting exploration of the space of model parameters through regenerative transient analysis [8,18,32]. As in [35,36], the approach relies on the hierarchical composition of sub-models to manage the complexity and stiffness of the problem. Yet, in this paper, we evaluate a model beyond the limits of the enabling restriction. While the approach of [35,36] is concerned with the evaluation of the steady state probability that a timeout occurs, we focus on the transient behavior and derive the probability that a timeout expires within an assigned time. This comprises a measure that is not of less interest than the one studied in [35,36]. In fact, whether the steady state value of the timeout probability is greater than the required threshold or not, it is valuable to study its trend over time and the factors that mainly affect it.

Overall, the approach provides insight in the problem characterization, showing that working beyond the limits of a Markovian setting poses major complexities, but it also provides an advantage in composing results. While this paper specifically addressed the ERTMS/ETCS Level 3 case-study, the model used for system performability evaluation is general enough to be easily adapted to suit similar systems featuring radio-signalling and moving block, like the modern

Communication Based Train Control (CBTC) applications for metro railways. The approach is also open to sensitivity analysis and integration with simulative and model-driven approaches [3], possibly in conjunction with other tools such as TimeNET [33], Möbius [12], and SHARPE [29].

References

1. Abbaneo, C., Flammini, F., Lazzaro, A., Marmo, P., Mazzocca, N., Sanseviero, A.: UML based reverse engineering for the verification of railway control logics. In: Int. Conf. on Dependability of Computer Systems, pp. 3–10. IEEE (2006)
2. Babczyński, T., Magott, J.: Dependability and safety analysis of ETCS communication for ERTMS level 3 using performance statecharts and analytic estimation. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) Proceedings of the Ninth International Conference on DepCoS-RELCOMEX. AISC, vol. 286, pp. 37–46. Springer, Heidelberg (2014)
3. Bernardi, S., Flammini, F., Marrone, S., Mazzocca, N., Merseguer, J., Nardone, R., Vittorini, V.: Enabling the usage of UML in the verification of railway systems: The DAM-rail approach. Reliability Eng. & System Safety **120**, 112–126 (2013)
4. Berthomieu, B., Diaz, M.: Modeling and verification of time dependent systems using time Petri nets. IEEE Trans. on Software Eng. **17**(3), 259–273 (1991)
5. Bobbio, A., Telek, M.: Markov regenerative SPN with non-overlapping activity cycles. In: Comp. Perf. and Dependability Symposium, pp. 124–133. IEEE (1995)
6. Bohnenkamp, H.C., D'Argenio, P.R., Hermanns, H., Katoen, J.-P.: MODEST: A compositional modeling formalism for hard and softly timed systems. IEEE Trans. on Software Eng. **32**(10), 812–830 (2006)
7. Carnevali, L., Grassi, L., Vicario, E.: State-density functions over DBM domains in the analysis of non-Markovian models. IEEE Trans. on Software Eng. **35**(2), 178–194 (2009)
8. Carnevali, L., Ridi, L., Vicario, E.: A framework for simulation and symbolic state space analysis of non-Markovian models. In: Flammini, F., Bologna, S., Vittorini, V. (eds.) SAFECOMP 2011. LNCS, vol. 6894, pp. 409–422. Springer, Heidelberg (2011)
9. Choi, H., Kulkarni, V.G., Trivedi, K.S.: Markov regenerative stochastic Petri nets. Performance Evaluation **20**(1–3), 337–357 (1994)
10. Ciardo, G., German, R., Lindemann, C.: A characterization of the stochastic process underlying a stochastic Petri net. IEEE Trans. on Software Engineering **20**(7), 506–515 (1994)
11. Ciardo, G., Trivedi, K.: SPNP: stochastic Petri net package. In: Int. Workshop on Petri Nets and Performance Models, pp. 142–151. IEEE (1989)
12. Courtney, T., Gaonkar, S., Keefe, K., Rozier, E., Sanders, W.H.: Möbius 2.3: an extensible tool for dependability, security, and performance evaluation of large and complex system models. In: IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN), pp. 353–358 (2009)
13. Flammini, F., Marrone, S., Iacono, M., Mazzocca, N., Vittorini, V.: A multi-formalism modular approach to ERTMS/ETCS failure modeling. International Journal of Reliability, Quality and Safety Engineering **21**(1) (2014)

14. Flammini, F., Marrone, S., Mazzocca, N., Vittorini, V.: Modelling structural reliability aspects of ERTMS/ETCS by fault trees and Bayesian networks. In: Proc. of the European Safety & Reliability Conference, ESREL, vol. 6 (2006)
15. EEIG Ertms User Group. ERTMS/ETCS RAMS System Requirements Specification, UIC, Brussels (1999)
16. EEIG Ertms User Group. ERTMS/ETCS Systems Requirements Specification, UIC, Brussels (1999)
17. Hermanns, H., Jansen, D.N., Usenko, Y.S.: From StoCharts to MoDeST: a comparative reliability analysis of train radio communications. In: Int. Workshop on Software and performance, pp. 13–23. ACM (2005)
18. Horváth, A., Paolieri, M., Ridi, L., Vicario, E.: Transient analysis of non-Markovian models using stochastic state classes. Perf. Eval. **69**(7–8), 315–335 (2012)
19. Kelling, C.: A framework for rare event simulation of stochastic Petri nets using RESTART. In: Conf. on Winter Simulation, pp. 317–324. IEEE (1996)
20. Kulkarni, V.G.: Modeling and analysis of stochastic systems. CRC Press (1996)
21. Lindemann, C., Thümmler, A.: Transient analysis of Deterministic and Stochastic Petri Nets with concurrent deterministic transitions. Performance Evaluation **36**, 35–54 (1999)
22. Qiu, S., Sallak, M., Schon, W.: Modeling of ERTMS level 2 as an SoS and evaluation of its dependability parameters using statecharts. IEEE Systems Journal **8**(4), 1169–1181 (2014)
23. Miner, A.S., Parker, D.: Symbolic representations and analysis of large probabilistic systems. In: Baier, C., Haverkort, B.R., Hermanns, H., Katoen, J.-P., Siegle, M. (eds.) Validation of Stochastic Systems. LNCS, vol. 2925, pp. 296–338. Springer, Heidelberg (2004)
24. Glynn, P.W.: A GSMP formalism for discrete-event systems. Proceedings of the IEEE **77**, 14–23 (1989)
25. Sanders, W.H., Meyer, J.F.: Stochastic activity networks: formal definitions and concepts. In: Brinksma, E., Hermanns, H., Katoen, J.-P. (eds.) EEF School 2000 and FMPA 2000. LNCS, vol. 2090, pp. 315–343. Springer, Heidelberg (2001)
26. Stewart, W.J.: Introduction to the numerical solution of Markov chains, vol. 41. Princeton University Press, Princeton (1994)
27. Telek, M., Rácz, S.: Numerical analysis of large Markov reward models. Performance Evaluation **36**, 95–114 (1999)
28. Trivedi, K.S.: Probability and statistics with reliability, queuing, and computer science applications. John Wiley and Sons, New York (2001)
29. Trivedi, K.S., Sahner, R.A.: SHARPE at the age of twenty two. ACM SIGMETRICS Perf. Eval. Review **36**(4), 52–57 (2009)
30. Trowitzsch, J., Zimmermann, A.: Using UML state machines and Petri nets for the quantitative investigation of ETCS. In: Int. Conf. on Performance evaluation methodologies and tools, pp. 34. ACM (2006)
31. Vicario, E.: Static analysis and dynamic steering of time dependent systems using time Petri nets. IEEE Trans. on SW Eng. **27**(1), 728–748 (2001)
32. Vicario, E., Sassoli, L., Carnevali, L.: Using stochastic state classes in quantitative evaluation of dense-time reactive systems. IEEE Trans. on Software Eng. **35**(5), 703–719 (2009)

33. Zimmermann, A.: Dependability evaluation of complex systems with TimeNET. In: Int. Workshop on Dynamic Aspects in Dependability Models for Fault-Tolerant Systems, (DYADEM-FTS 2010) (2010)
34. Zimmermann, A., Freiheit, J., German, R., Hommel, G.: Petri net modelling and performability evaluation with TimeNET 3.0. In: Haverkort, B.R., Bohnenkamp, H.C., Smith, C.U. (eds.) TOOLS 2000. LNCS, vol. 1786, pp. 188–202. Springer, Heidelberg (2000)
35. Zimmermann, A., Hommel, G.: A train control system case study in model-based real time system design. In: Int. Parallel and Distributed Processing Symposium, pp. 118–126. IEEE (2003)
36. Zimmermann, A., Hommel, G.: Towards modeling and evaluation of ETCS real-time communication and operation. Journal of Sys. and Soft. **77**(1), 47–54 (2005)